ESD-TR-88-329

Technical Report
849

Adaptive Preprocessing of Nonstationary Signals

M.D. Eggers
T.S. Khoun

9 May 1989

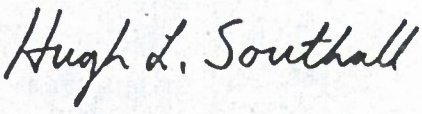
Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-85-C-0002.

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-85-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ADAPTIVE PREPROCESSING OF NONSTATIONARY SIGNALS

M.D. EGGERS
T.S. KHOUN
Group 93

TECHNICAL REPORT 849

9 MAY 1989

LEXINGTON MASSACHUSETTS

ABSTRACT

As the number and bandwidth of sensors increases, an acute demand for preprocessing sensor data obtained for machine-based decision making arises. Especially in a data fusion context, the data from numerous sensors must first be preprocessed to prevent saturation of the decision making mechanism — albeit man or machine.

Presented is a general preprocessing approach which provides a compact representation (feature vector) of sensor data. The approach, supported by a signal decomposition theorem, adaptively models in recursive fashion, the detrended sensor data as an autoregressive (AR) process of sufficiently high order. Provisions are included to accommodate nonstationary data by incorporating an information-theoretic transition detector to identify the segments of near-stationary data. Together, feature vectors (AR coefficients) are produced over near-stationary data segments which are scale invariant, translation invariant, normalized, and represent sufficient statistics. Furthermore, the merit of the preprocessor is quantitatively determined in a continuous manner from the resulting innovations (modeling error process).

Specific application results utilizing nonstationary radar data demonstrate the ability to simultaneously reduce data and maintain information content, without requiring a priori statistics and/or expert rules.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express appreciation to R. Levine who shared comments and, at times, lengthy sessions on preprocessing, and to Prof. Emanuel Parzen for his influence on shaping both my understanding and appreciation of time series analysis.

ADAPTIVE PREPROCESSING OF NONSTATIONARY SIGNALS

1. INTRODUCTION

1.1 MOTIVATION

As the number and bandwidth of sensors increases, an acute demand for preprocessing sensor data obtained for machine-based decision making arises. Especially in a data fusion context, where the information from numerous sensors is combined to yield high-level decisions, the data from individual sensors must be preprocessed before fusion to prevent saturation of the decision making mechanism, albeit man or machine.

Preprocessing can be viewed as constructing an alternative representation of the data, one which provides desired invariances and reduces redundancy. In short, the primary objective of preprocessing is to produce a compact representation (feature vector) of the sensor data which simultaneously

- Reduces data and
- Maintains information content.

1.2 GENERAL APPROACH

A general approach which satisfies the above preprocessing objectives is driven by Wold's decomposition theorem [1]. The fundamental theorem basically states that any stationary discrete-time process can be decomposed into a summation of a deterministic signal and an autoregressive (AR) process of sufficiently high order [2]. Consequently, a dynamic preprocessor following Wold's theorem first entails detrending the data for deterministic quantities and subsequently fitting an AR model to the remaining process as illustrated in Figure 1-1.

Unfortunately for most real situations, the sensor data are nonstationary. However, many of the nonstationary real signals are piecewise stationary. That is, although the signal statistics may vary significantly over the complete data record, localized regions can be identified where the statistics remain constant. Consequently, Wold's theorem can also be applied to nonstationary signals provided the near-stationary segments can be identified. Hence, by employing a statistical transition detector for identifying the near-stationary segments, a general preprocessor architecture for nonstationary signals can be constructed (see Figure 1-2).

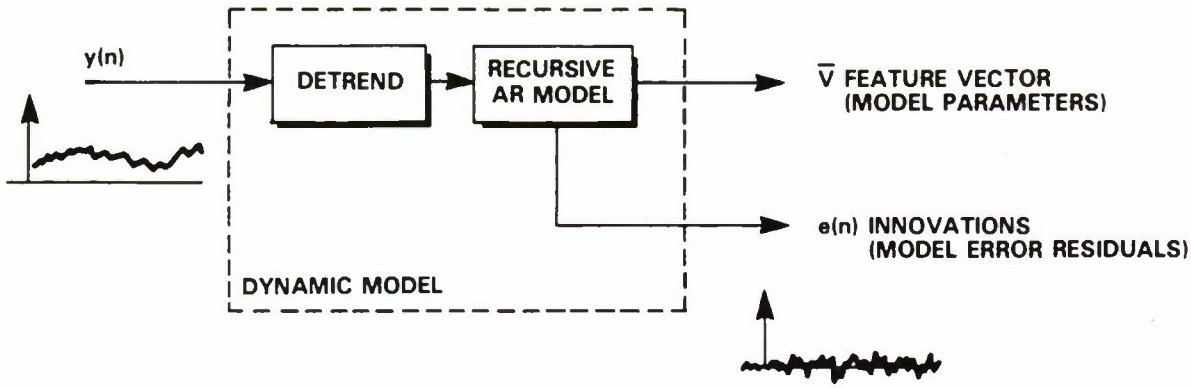


Figure 1-1. Preprocessing architecture for stationary signals.

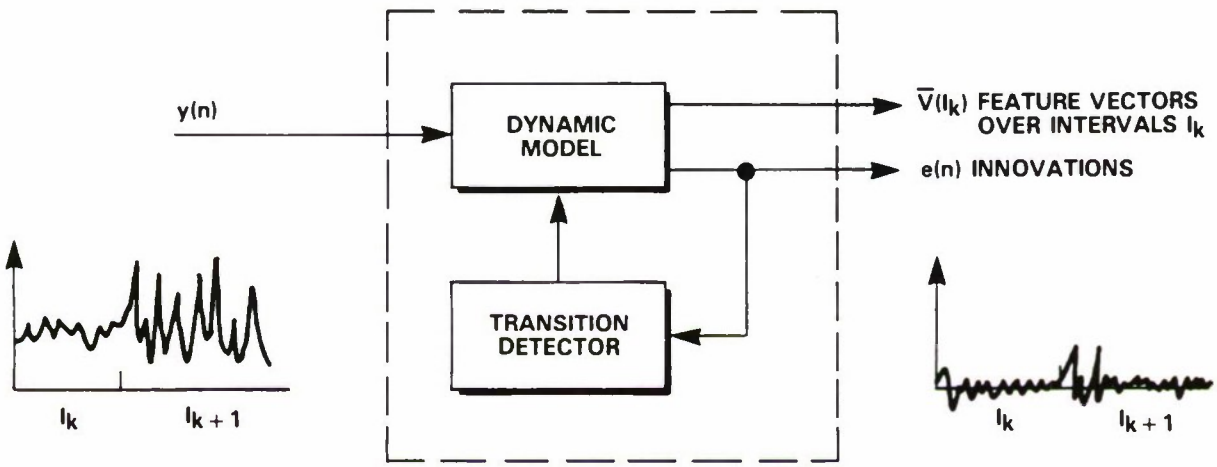


Figure 1-2. Preprocessing architecture for nonstationary signals. A feature vector $\bar{V}(I_k)$ is produced over each near-stationary segment I_k.

1.3 ORGANIZATION

The organization of this report follows the nonstationary preprocessor architecture (Figure 1-2) comprising a dynamic AR modeling module and transition detector.

In Section 2, an adaptive AR modeling algorithm is developed to satisfy objectives derived from demands posed by a real sensor environment, including unknown a priori statistical information and fast throughput. Hence, a recursive algorithm is sought for dynamically building an AR model for each of the near-stationary segments of data. To identify these segments, in Section 3 a transition detector is developed which couples well with the previously developed recursive AR modeling algorithm. Both detection and

estimation of transition times are achieved. The complete set of algorithms for preprocessing are displayed in Section 4.

The results obtained by applying the developed preprocessing algorithms on real data are presented in Section 5, and issues concerning detrending and model order selection are addressed. A summary of the results is followed by the conclusion section, providing direction for further research.

2. ADAPTIVE AR MODELING

2.1 INTRODUCTION

Motivated by Wold's theorem, consider modeling a stochastic process with an autoregressive (AR) model of order M .

Let

$$y(n) = \sum_{i=1}^{M} w_i^* \, y(n-i) + e(n)$$

$$= \bar{w}^H \, \bar{y}(n-1) + e(n) \tag{2.1}$$

where

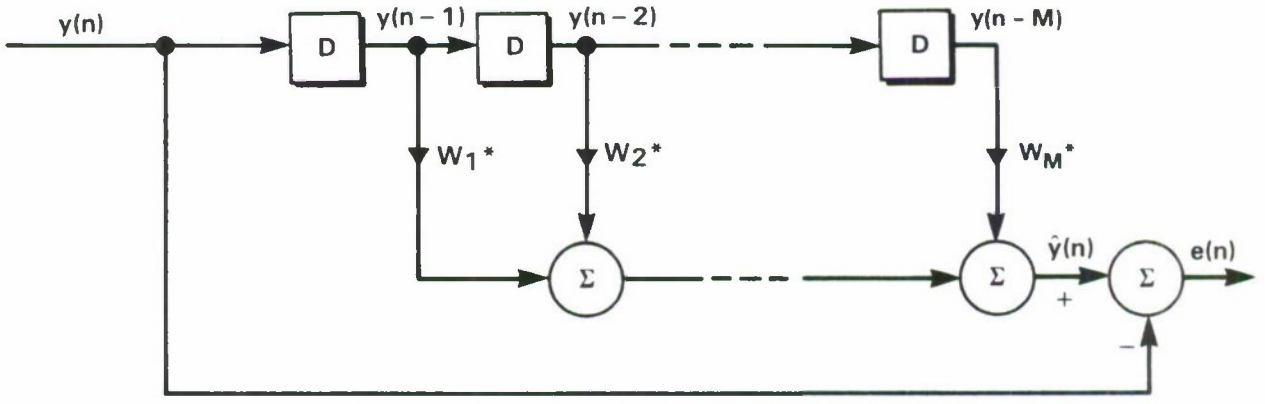
$$\bar{y}(n-1) \equiv \begin{pmatrix} y(n-1) \\ y(n-2) \\ \vdots \\ y(n-M) \end{pmatrix} \tag{2.2}$$

last M samples

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{pmatrix} \tag{2.3}$$

AR coefficients

and $e(n)$ represents the modeling error, often referred to as the "innovation process." The schematic representing the AR modeling of the random process $y(n)$ is shown in Figure 2.1. The diagram can be viewed as a digital transversal whitening filter, since the output $e(n)$ should be white noise of variance σ_e^2 if the colored process $y(n)$ is truly an AR (w, M) process.



109165-6

Figure 2-1. Digital transversal whitening filter representing the AR modeling process. The tap weights w are specified a priori.

The optimal set of coefficients $\bar{w}^{\circ}$, resulting in an uncorrelated innovation process, are obtained by solving a matrix equation [3,4] derived from the autocorrelation function of the process $y(n)$. Assuming $y(n)$ represents a stationary segment of data, then the autocorrelation function becomes (see Appendix A)

$$r(m) \equiv E\{y(n+m)\,y^*(n)\} \tag{2.4}$$

$$= \begin{cases} \sum_{i=1}^{M} w_i^{\circ *} r(m-i) + h(0)\sigma_e^2 & m = 0 \\ \sum_{i=1}^{M} w_i^{\circ *} r(m-i) & m > 0 \\ r(-m) & m < 0 \end{cases}$$

or equivalently for lags $m = 1,2,\ldots M$

$$\bar{R}\bar{w}^{\circ} = \bar{r} \tag{2.5}$$

where

$$\bar{R} \equiv E\left\{\bar{y}(n-1)\,\bar{y}^H(n-1)\right\} = \begin{pmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(-1) & r(0) & \cdots & r(M-2) \\ \vdots & & \cdots & \\ r(-M+1) & r(-M+2) & \cdots & r(0) \end{pmatrix} \tag{2.6}$$

$$\bar{r} = \begin{pmatrix} r(-1) \\ r(-2) \\ \vdots \\ r(-M) \end{pmatrix} \tag{2.7}$$

Thus, solving the matrix equation $\bar{R}\bar{w}^{\circ} = \bar{r}$ yields the AR coefficients $\bar{w}^{\circ}$. However, such batch processing procedure requires matrix inversion as well as a priori knowledge of the autocorrelation function and hence is not suitable for the current problem, requiring a recursive algorithm operating in the absence of a priori input process statistics and capable of adapting to changing statistical environments.

2.2 OPTIMAL LINEAR PREDICTION EQUIVALENCE

The desired adaptive recursive algorithm for computing the model coefficients is readily obtained by alternatively viewing the task at hand as an optimal linear prediction problem in the sense of Wiener [5].

Reformulating, consider predicting $y(n)$ given the past M samples $\bar{y}(n-1)$. For optimal linear prediction, the output $\hat{y}(n)$ is simply a linear combination of $\bar{y}(n-1)$ expressed as

$$\hat{y}(n) = \bar{w}^H \bar{y}(n-1) . \tag{2.8}$$

The objective is to choose the tap weights $\bar{w}$ (AR model coefficients) resulting in the best performance. Here, best performance is defined as minimum mean-squared-error (mse).

Expressing the error between the desired and the linear filter output by rearranging Equation (2.1)

$$e(n) = y(n) - \hat{y}(n) , \tag{2.9}$$

the performance measure (mse) becomes

$$J(\bar{w}) = E\{e(n)e^*(n)\} \tag{2.10}$$

representing the average innovation power or equivalently, the variance of the prediction error (AR modeling error). Now upon substitution

$$\begin{aligned} J(\bar{w}) &= E\left\{\left(y(n) - \bar{w}^H \bar{y}(n-1)\right)\left(y^*(n) - \bar{y}^H(n-1)\bar{w}\right)\right\} \\ &= r(0) - \bar{w}^H \bar{r} - \bar{r}^H \bar{w} + \bar{w}^H \bar{R} \bar{w} , \end{aligned} \tag{2.11}$$

the performance measure is seen to be quadratic in the weight vector, and hence yields a parabolic surface with a unique minimum.

Consequently, obtaining the optimal tap weights is achieved by minimizing $J(\bar{w})$ using the zero derivative criteria.

$$\frac{\partial J(\bar{w})}{\partial \bar{w}} = -2\bar{r} + 2\bar{R}\bar{w}^\circ = \bar{0} \tag{2.12}$$

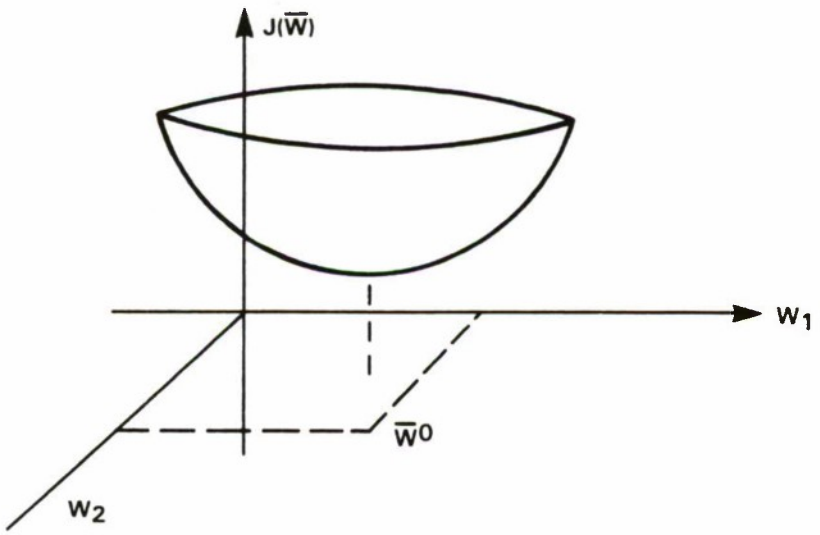
$$\text{or} \qquad \bar{R}\bar{w}^\circ = \bar{r} .$$

And thus the *equivalence is established between AR modeling and optimal linear prediction*, for the optimal predictor tap weights (impulse response of the linear prediction filter) are equivalent to the AR coefficients [Equation (2.5)].

2.3 RECURSIVE STEEPEST-DESCENT AR MODEL FORMATION

Returning to the objective of obtaining an adaptive recursive algorithm for the AR coefficients, notice that under the optimal prediction interpretation a performance measure (mse) was defined, guiding the selection of the coefficients. Supplied with such measure, a recursive algorithm can now be formulated following the method of steepest descent. The weight vector (AR model coefficients) at time $n + 1$ is given by

$$\bar{w}(n+1) = \bar{w}(n) - \eta \frac{\partial J(\bar{w})}{\partial \bar{w}}\Big|_{\bar{w}(n)} = \bar{w}(n) + 2\eta(\bar{r} - \bar{R}\bar{w}(n)). \tag{2.13}$$



109165-7

Figure 2-2. Performance measure surface as a function of the tap weights.

Furthermore, when the process $y(n)$ is stationary and the statistics $\bar{R}$ are known a priori, the algorithm converges to the optimal weights regardless of the initial conditions, provided the rate of adaption is bounded by

$$0 < \eta < \frac{1}{\lambda_{max}} \tag{2.14}$$

where λ_{max} is the largest eigenvalue of the correlation matrix $\bar{R}$ [2].

Although the steepest-descent algorithm [Equation (2.13)] is recursive and guaranteed to converge to the desired AR coefficients, the requirement of the signal correlation matrix ($\bar{R}$) prohibits its use under the previously imposed objectives. However, as an approximation to steepest descent, namely the least mean square (LMS) algorithm [6] can be utilized without knowing $\bar{R}$ a priori.

2.4 RECURSIVE LMS AR MODEL FORMULATION

The formulation of the LMS algorithm follows by replacing the statistical quantities $\bar{R}$ and $\bar{r}$ in Equation (2.13) by the instantaneous estimates. That is, substitute $\bar{y}(n-1)\bar{y}^H(n-1)$ and $\bar{y}(n-1)y^*(n)$ for $\bar{R}$ and $\bar{r}$, respectively, in Equation (2.13) to yield

$$\bar{w}(n+1) = \bar{w}(n) + 2\eta\left(\bar{y}(n-1)\,y^*(n) - \bar{y}(n-1)\,\bar{y}^H(n-1)\bar{w}(n)\right)$$

$$= \bar{w}(n) + 2\eta\bar{y}(n-1)\Big(y(n) - \bar{w}^H(n)\,\bar{y}(n-1)\Big)^*$$

$$= \bar{w}(n) + 2\eta\bar{y}(n-1)e^*(n). \tag{2.15}$$

The LMS algorithm is often initialized with $\bar{w}(0) = \bar{0}$ and is diagramed in Figure 2-3. In comparison with Figure 2-1, this *recursive* algorithm results in an *adaptive* transversal whitening filter which is *time varying* and *nonlinear*.

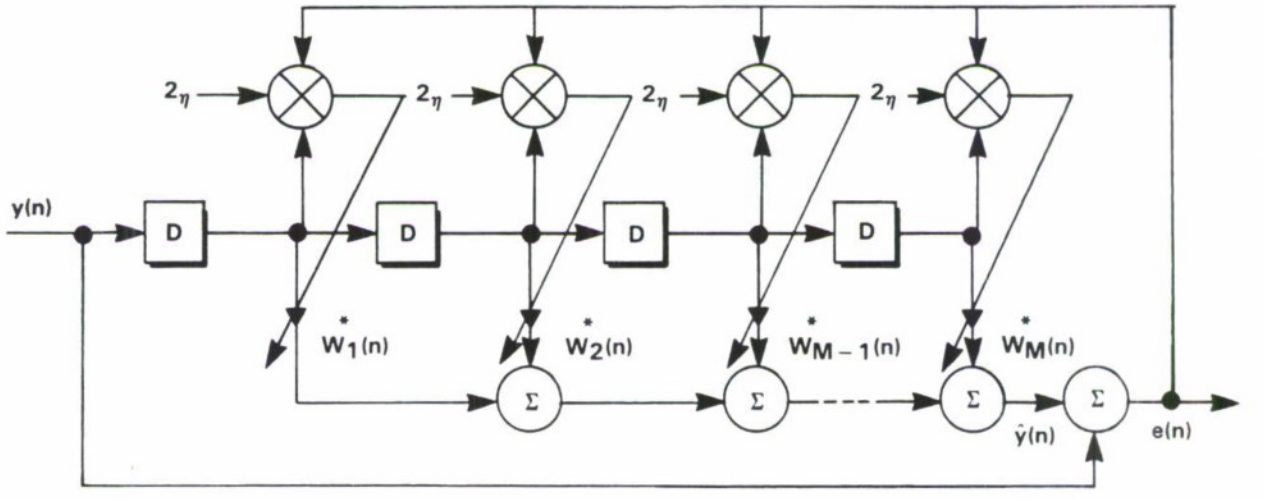


Figure 2-3. Adaptive digital transversal whitening filter representing an adaptive AR modeling process. Here the tap weights $\bar{w}(n)$ are determined adaptively.

Further examination of the algorithm reveals the weights at time $n + 1$ to be *explicitly* dependent upon the last $M + 1$ values of the random input process [through the product of $\bar{y}(n-1)$ and the innovations $e(n)$] and *implicitly* dependent upon all past values of the random input process [through the previous weight vector $\bar{w}(n)$]. Hence, the memory of the filter is characterized by the choice of the adaption (learning) parameter η; for small η the filter memory is long, wherein the dependence is primarily implicit amongst all past values of $y(n)$, and for large η the filter memory is short, wherein the dependence is primarily explicit amongst the past $M + 1$ samples of $y(n)$ (see Figure 2-4).

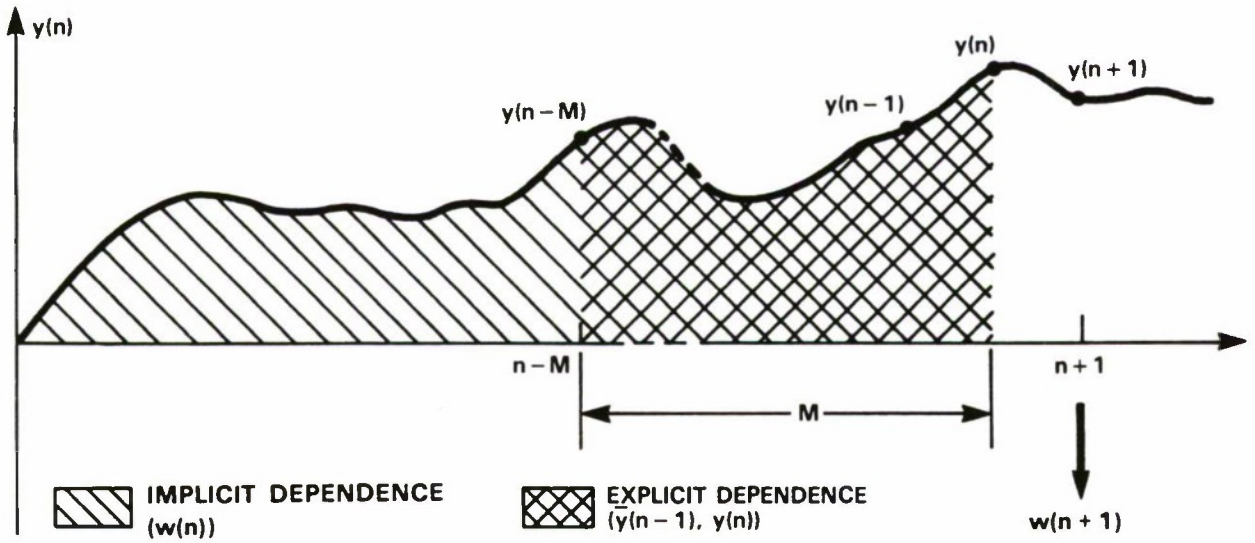


Figure 2-4. Dependence of the weight adaptions upon the input data.

Considering the LMS convergence properties, clearly by employing the instantaneous estimates, the weights are expected to fluctuate during the iterative process. However, under certain independence assumptions [7], or under specific statistical correlation properties of the signal [8], both the ensemble average of the LMS weight vector and the ensemble average of the mse converge to the optimal values, provided the rate of adaption is bounded by

$$0 < \eta < 1/\sum_{i=1}^{M} \lambda_i \tag{2.16}$$

where $\{\lambda_i\}$ are the eigenvalues of the correlation matrix. Moreover, this criterion [Equation (2.16)] is simplified by observing for a stationary $y(n)$ process

$$Mr(0) = tr[\bar{R}] = tr[\bar{Q}\bar{Q}^H \bar{R}] = tr[\bar{Q}^H \bar{R}\bar{Q}] = tr[\bar{\Lambda}] = \sum_{i=1}^{M} \lambda_i \tag{2.17}$$

where $\bar{Q}$ is the matrix of eigenvectors for $\bar{R}$ and $\bar{\Lambda}$ is the diagonal matrix containing the eigenvalues. Thus the convergence in mean and mean square of the AR coefficients, when $y(n)$ is a stationary process conforming to the various assumptions, is guaranteed provided

$$0 < \eta < \frac{1}{Mr(0)}. \tag{2.18}$$

2.5 SUMMARY OF ADAPTIVE AR MODELING ALGORITHM

The LMS algorithm suited for recursively determining the AR model coefficients describing a stochastic process $y(n)$ in the absence of complete a priori statistics is summarized below.

A Priori Parameters

$$M = \text{AR model order} \tag{2.19}$$

$$\eta = \text{adaption (learning) parameter } \left(0 < \eta < \frac{1}{M\, r(0)}\right) \tag{2.20}$$

Initial Conditions

$$\bar{w}(0) = \bar{0} \tag{2.21}$$

$$y(0) = 0 \tag{2.22}$$

Innovations

$$e(n) = y(n) - \bar{w}^H(n)\,\bar{y}(n-1) \tag{2.23}$$

$$= y(n) - \sum_{i=1}^{M} w_i^*(n)\, y(n-i)$$

AR Coefficients

$$\bar{w}(n+1) = \bar{w}(n) + 2\eta\, \bar{y}(n-1)e^*(n) \tag{2.24}$$

or $$w_i(n+1) = w_i(n) + 2\eta\, y(n-i)e^*(n). \tag{2.25}$$

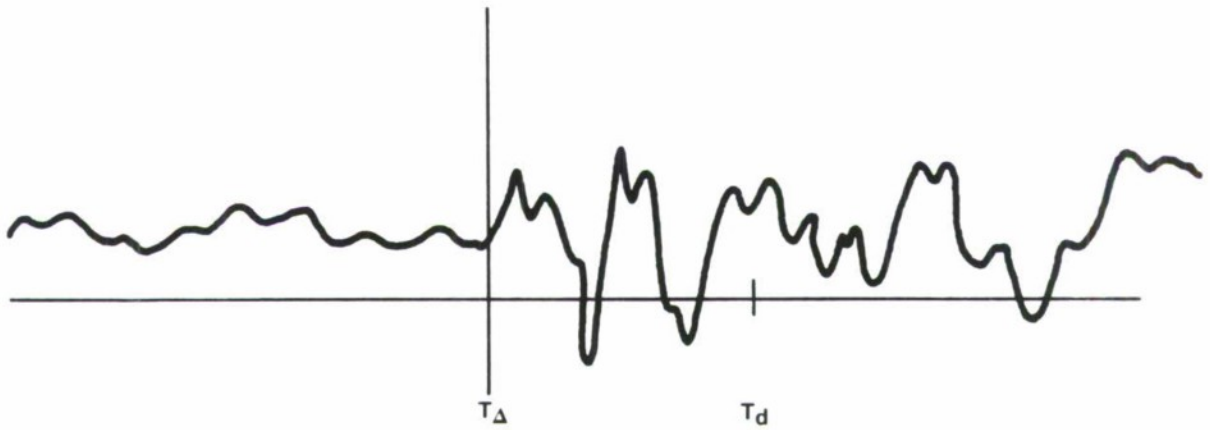
3. TRANSITION DETECTION

3.1 OBJECTIVE

Recall that the central objective of preprocessing a nonstationary signal is to produce feature vectors over near-stationary data segments. These feature vectors, subsequently used for signal classification, ideally provide data reduction while maintaining information content.

Consequently, identifying the near-stationary intervals comprising the nonstationary signal is of prime importance in feature vector extraction. And since the boundaries of the near-stationary segments coincide with a change in the spectral characteristics of the signal, a spectral transition detection algorithm is warranted.

In addition to detecting a change in spectral character, an algorithm is sought which also estimates the actual transition occurrence (T_Δ) with short delay $T_d - T_\Delta$ and few false alarms (see Figure 3-1).



109165-8

Figure 3-1. Piecewise stationary signal with spectral transition at T_Δ and subsequent detection at T_d.

Furthermore, in harmony with the previously mentioned AR modeling objectives, the adaptive algorithm is desired recursive and operational in environments where the signal statistics are unknown both before and after transition.

A candidate algorithm for transition detection is presented which couples nicely to the previously developed recursive AR modeling algorithm. The algorithm discussed is a dual window approach which counters many limitations of the classical single window technique that essentially tests how far from the white noise hypothesis are the innovations arising from the modeling.

3.2 DUAL WINDOW APPROACH

To counter the limitations of the classical single window approach (including large variances before transition, unpredictable behavior following transitions where a decrease in signal energy occurs, and requirement of a priori reference information), a technique utilizing both global and local windows is adopted. The approach introduced by Basseville and Benveniste [9] offers better behavior before transitions and yields larger drifts in the test statistic following detection, thereby improving detection capability.

The approach, like numerous other transition detection algorithms, employs a cumulative sum statistic of the form

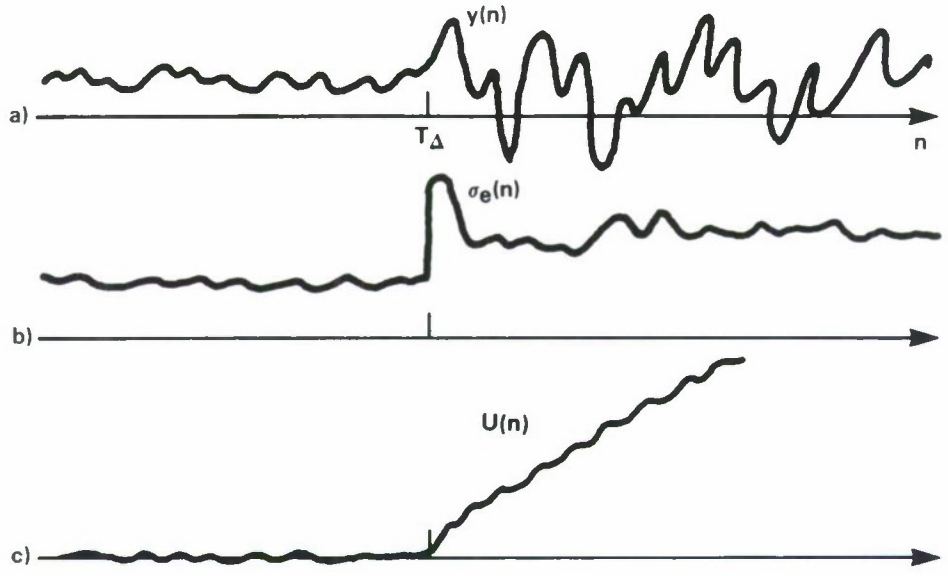
$$u(n) = \sum_{k=1}^{n} T(k) \tag{3.1}$$

[$T(k)$ to be determined] whose drift properties signal a change in spectral characteristics. The input driving the statistic is the modeling error or innovation process, as diagramed in Figure 3-2.

Figure 3-2. General transition detector based on cumulative sum statistic.

Basically, the integration effect provided by a cumulative sum statistic provides more reliable detection capability in noisy environments (i.e., where the true AR model coefficients are unknown). Thus, instead of detecting a change in absolute mean of the innovation process at a transition, rather a more sensitive change in drift of $u(n)$ is detected as shown in Figure 3-3.

The distinctive feature of this approach, however, is the utilization of two windows. The *global* reference window expands during the process allowing *all* information in the stationary segment under the distribution P_a to be included in the AR model building. Utilizing all information enables better modeling (estimation of the AR coefficients) of $y(n)$ before the transition. In contrast, the *fixed* local window uses only the most *recent* information. When utilizing a dual window transition detection approach in conjunction with recursive AR modeling, the *global and local windows can be effectively implemented by simply choosing appropriate learning parameters.* For example, when using two LMS filters (per Section 2.4) for recursively estimating the local and global AR coefficients simultaneously, the learning parameter for the global filter (η_a) is chosen smaller than the learning parameter for the local filter (η_b) within the constraint [Equation (2.18)]. Consequently, the global filter with long memory (small η_a) coincides with the expanding global window, while the local filter with short memory (large η_b) corresponds to the fixed local window.



109165-10

Figure 3-3. Properties of the (a) signal, (b) innovation, and (c) cumulative sum statistic at a spectral transition (true modeling coefficients are unknown).

Now as the region of transition is approached [see Figure 3-4 (b) and (c)], the information theoretic distance metric between the distribution laws p_a and p_b extracted from the global and local windows increases, since the global reference model with long memory is virtually unaffected by the most recent information. Once the distance measure exceeds a given threshold, the transition is detected and the reference window reinitialized [see Figure 3-4 (d)].

The particular distance measure used to gauge the discrepancy between the adaptive filters modeling the signal based upon complete (global window) and partial (local window) information is given by

$$T(k) = I_{p_a/p_b}(y(k)/\bar{y}(k-1)) - E\left\{I_{p_a/p_b}(y(k)/\bar{y}(k-1))\right\} \tag{3.2}$$

where

$$I_{p_a/p_b}(y(k)/\bar{y}(k-1)) = \log \frac{p_a(y(k)/\bar{y}(k-1))}{p_b(y(k)/\bar{y}(k-1))} \tag{3.3}$$

$$E_{p_a}\left\{I_{p_a/p_b}(y(k)/\bar{y}(k-1))\right\} = \int p_a(y/\bar{y}(k-1)) \log \frac{p_a(y/\bar{y}(k-1))}{p_b(y/\bar{y}(k-1))}\, dy \tag{3.4}$$

(conditional Kullback's information [10])

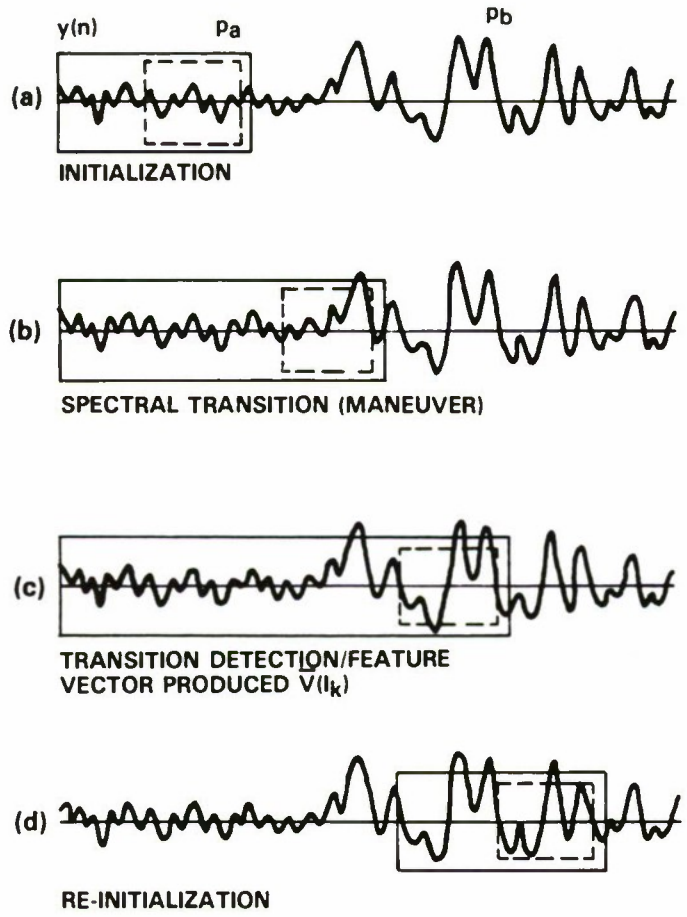


Figure 3-4. Dual window transition detector operation.

resulting in the cumulative sum statistic given by Equation (3.1). Notice from Equation (3.2) that the measure is seen to be a difference in the average and instantaneous discrepancies in the distributions representing the data from the global and local windows.

The desirable properties of this test statistic are revealed in part by examining the drifts in the conditional mean value before and after transition. Before transition

$$D_a = E_{p_a}\{u(n) - u(n-1)\} = E_{p_a}\{T(n)\} = 0, \tag{3.5}$$

while after the transition

$$D_b = E_{p_b}\{u(n) - u(n-1)\} \tag{3.6}$$

$$= \int \left[p_b(y / \bar{y}(n-1)) - p_d(y / \bar{y}(n-1)) \right] \log \frac{p_a(y / \bar{y}(n-1))}{p_b(y / \bar{y}(n-1))} \, dy$$

(conditional Kullback's divergence)

$$< 0 .$$

Thus, zero conditional drift in $u(n)$ occurs before the transition, while a negative drift occurs after the transition. Consequently, the detection of a statistical transition in the process $y(n)$ can be accomplished by detecting a nonzero drift in $u(n)$.

3.3 TRANSITION DETECTION ALGORITHM

Examining Equation (3.2), obtaining an explicit algorithm for the cumulative sum test statistic requires knowledge of the functional form of the distributions. Now if the input process is assumed jointly Gaussian, the resulting transition detection algorithm couples nicely with the previous recursive AR modeling algorithm.

In particular, once the recursive AR modeling coefficients have converged to the optimal values (equivalently, the adaptive filter's impulse response has converged to the optimal tap weights; $\bar{w} \rightarrow \bar{w}^{\circ}$, $\bar{R}\bar{w}^{\circ} = \bar{r}$ and $\sigma_e^2 = r(0) - \bar{w}^{\circ H} \bar{r}$), the cumulative sum test statistic for detecting statistical transitions is given by Equation (3.1) with (see Appendices B and C).

$$T(k) = \frac{1}{2} \left[1 - \frac{\sigma_{e_a}^2}{\sigma_{e_b}^2} + \frac{|e_b(k)|^2}{\sigma_{e_b}^2} - \frac{|e_a(k)|^2}{\sigma_{e_a}^2} - \frac{|e_b(k) - e_a(k)|^2}{\sigma_{e_b}^2} \right] \tag{3.7}$$

The associated schematic representing the algorithm Equation (3.7) is shown in Figure 3-5. Notice the statistic represents a normalized distance measure between the innovation processes derived from the local and global data windows. Specifically, the statistic is driven by both the squared difference and the difference in squares of the respective innovation processes.

3.4 TRANSITION TIME ESTIMATION ALGORITHM

Following detection of a spectral transition utilizing the cumulative sum statistic, the actual time of transition must be estimated. The basic estimation task is illustrated in Figure 3-6, where the objective is to provide an estimate of the transition time T_Δ with minimal delay $(T_d - T_\Delta)$ following detection.

An optimal approach developed by Hinkley [11] for minimizing the detection delay time $(T_d - T_\Delta)$, assuming a fixed average time between false alarms, involves assigning a drift bias δ to the test statistic. Thus, instead of estimating the point of departure from zero drift as depicted in Figure 3-6, the transition estimate $\hat{T}_\Delta$ is simply the time where the biased cumulative sum statistic attains minimal value in the neighborhood of the alarm per Figure 3-7.

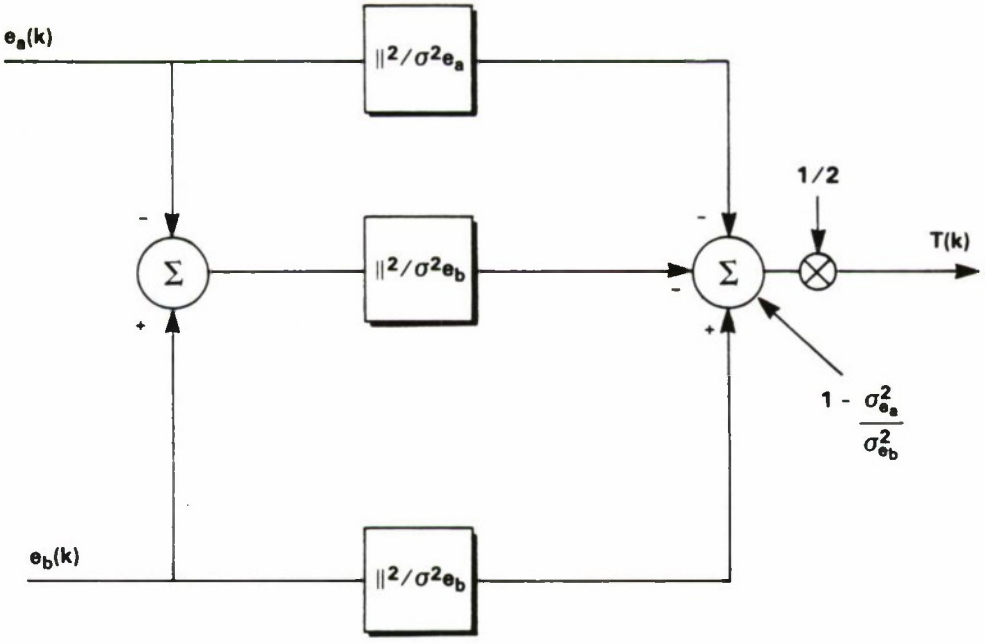


Figure 3-5. Transition detection algorithm architecture.

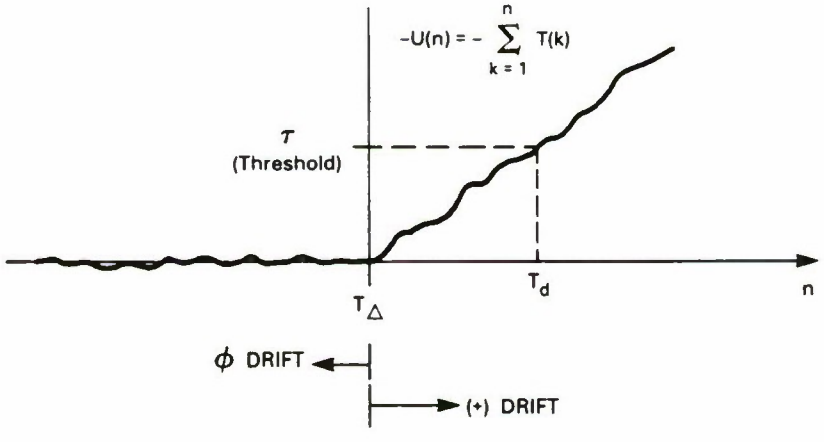


Figure 3-6. Cumulative sum statistic behavior at a transition.

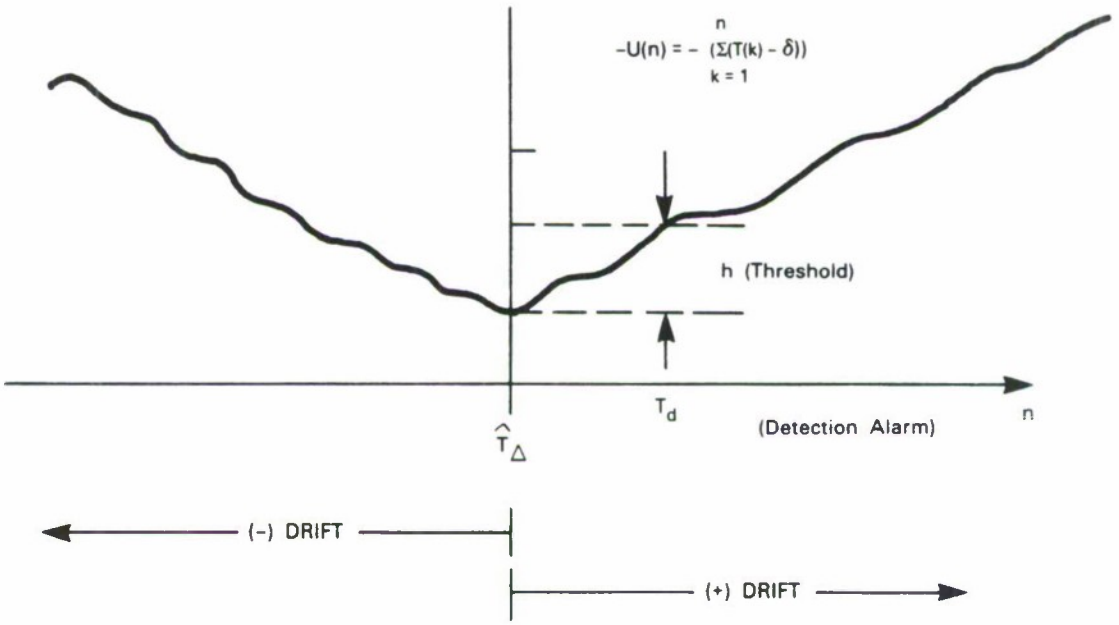


Figure 3-7. Biased cumulative sum statistic behavior at a transition.

Consequently, the complete transition detection and estimation algorithm is given by

Statistics

$$u(n) = \sum_{k=1}^{n} T(k) - \delta \tag{3.8}$$

$$T(k) = \frac{1}{2}\left[1 - \frac{\sigma_{e_a}^2}{\sigma_{e_b}^2} + \frac{|e_b(k)|^2}{\sigma_{e_b}^2} - \frac{|e_a(k)|^2}{\sigma_{e_a}^2} - \frac{|e_b(k) - e_a(k)|^2}{\sigma_{e_b}^2}\right] \tag{3.9}$$

Detection Rule

$$\begin{aligned} u(n) - m(n) \geq h \quad &\Rightarrow \text{detection} \\ < h \quad &\Rightarrow \text{no detection} \end{aligned} \tag{3.10}$$

where

$$m(n) = \min\ u(k) \qquad 0 \leq k \leq n \tag{3.11}$$

Estimate

$$\hat{T}_{\Delta} = n_o \ni u(n_o) = \min u(k) \qquad k \leq T_d \,. \tag{3.12}$$

4. SUMMARY OF PREPROCESSING ALGORITHMS

Combining the results from the adaptive recursive AR modeling and the dual window transition detection results in the following set of algorithms for preprocessing nonstationary data.

ADAPTIVE AR MODELING

Global Model (long memory)		**Local Model** (short memory)
	A Priori Parameters	
M_a	AR model order	M_b
η_a	adaption (learning) parameter $\left(0 < \eta_b < \frac{1}{M\, r(0)}\right)$	η_b
	$(0 < \eta_a < \eta_b)$	

Initial Conditions

$$\overline{w}_a(0) = \overline{0} \qquad \overline{w}_b(0) = \overline{0} \qquad (4.1a,b)$$

$$y(0) = 0 \qquad y(0) = 0 \qquad (4.2a,b)$$

Innovations

$$e_a(n) = y(n) - \sum_{i=1}^{M_a} w_{i_a}^*(n)\, y(n-i) \qquad e_b(n) = y(n) - \sum_{i=1}^{M_b} w_{i_b}^*(n)\, y(n-i) \qquad (4.3a,b)$$

AR Coefficients

$$w_{i_a}(n+1) = w_{i_a}(n) + 2\eta_a\, y(n-i_a)\, e_a^*(n) \qquad w_{i_b}(n+1) = w_{i_b}(n) + 2\eta_b\, y(n-i_b)\, e_b^*(n) \qquad (4.4a,b)$$

$$i_a = 1,2,\ldots M_a \qquad i_b = 1,2,\ldots M_b$$

TRANSITION DETECTION

A Priori Parameters

δ	drift bias	(empirically determined)
h	threshold	

Initial Conditions

$$u(0) = 0 \tag{4.5}$$

Cumulative Sum Statistic

$$T(n) = \frac{1}{2}\left[1 - \frac{\sigma_{e_a}^2}{\sigma_{e_b}^2} + \frac{|e_b(n)|^2}{\sigma_{e_b}^2} - \frac{|e_a(n)|^2}{\sigma_{e_a}^2} - \left|\frac{e_b(n) - e_a(n)}{\sigma_{e_b}^2}\right|^2\right] \tag{4.6}$$

$$u(n) = u(n-1) + T(n) - \delta \tag{4.7}$$

Detection Rule

$$u(n) - m(n) \geq h \quad \Rightarrow \text{ detect transition } (T_d) \tag{4.8}$$

$$< h \quad \Rightarrow \text{ no transition}$$

where

$$m(n) = \min\ u(k)$$

$$0 \leq k \leq n$$

Transition Time Estimate

$$\hat{T}_\Delta = n_o \ni u(n_o) = \min u(k) \qquad k \leq T_d. \tag{4.9}$$

In addition, if the variances of the respective innovation processes are unknown, the unbiased sample variance estimates below can be substituted in Equation (4.6).

$$\hat{\sigma}_e^2(n) = \frac{1}{n-1}\sum_{k=1}^{n}[e(k) - \bar{e}(n)]^2 \tag{4.10}$$

where

$$\bar{e}(n) = \frac{1}{n}\sum_{k=1}^{n} e(k) \tag{4.11}$$

or the recursive forms may be utilized

$$\hat{\sigma}_e^2(n) = \hat{\sigma}_e^2(n-1) + [\bar{e}(n) - \bar{e}(n-1)]^2 + \frac{1}{n-1}[e(n) - \bar{e}(n)]^2 \tag{4.12}$$

where

$$\bar{e}(n) = \frac{n-1}{n}\bar{e}(n-1) + \frac{1}{n}e(n). \tag{4.13}$$

The schematic representing the collective preprocessing algorithms is shown in Figure 4-1. As a final comment, the resulting feature vectors computed over the near-stationary intervals possess several properties instrumental in contributing to good classification performance. Regarding invariances, scale invariance is easily demonstrated by scaling both sides of Equation (2.1) by a constant, and observing the new process $z(n) = ay(n)$ yields the *same* AR coefficients. Also from the same equation, translation invariance is easily demonstrated by forming a new process $z(n) = y(n-k)$ and again observing that the coefficients are *identical*.

Furthermore, the characteristic equation $(1 - w_1^* z^{-1} - w_2^* z^{-2} - \ldots - w_M^* z^{-M} = 0)$ representing an AR asymptotic stationary (physical) process must have roots bounded in norm by unity. Hence, the AR coefficients themselves are bounded, although not necessarily by unity. Yet for many applications, including the present, the bound for the AR coefficients is empirically observed to be unity.

Consequently, in addition to the compact representation being a sufficient characterization of the stochastic process, the feature vectors inherently instill a degree of robustness due to the invariance properties.

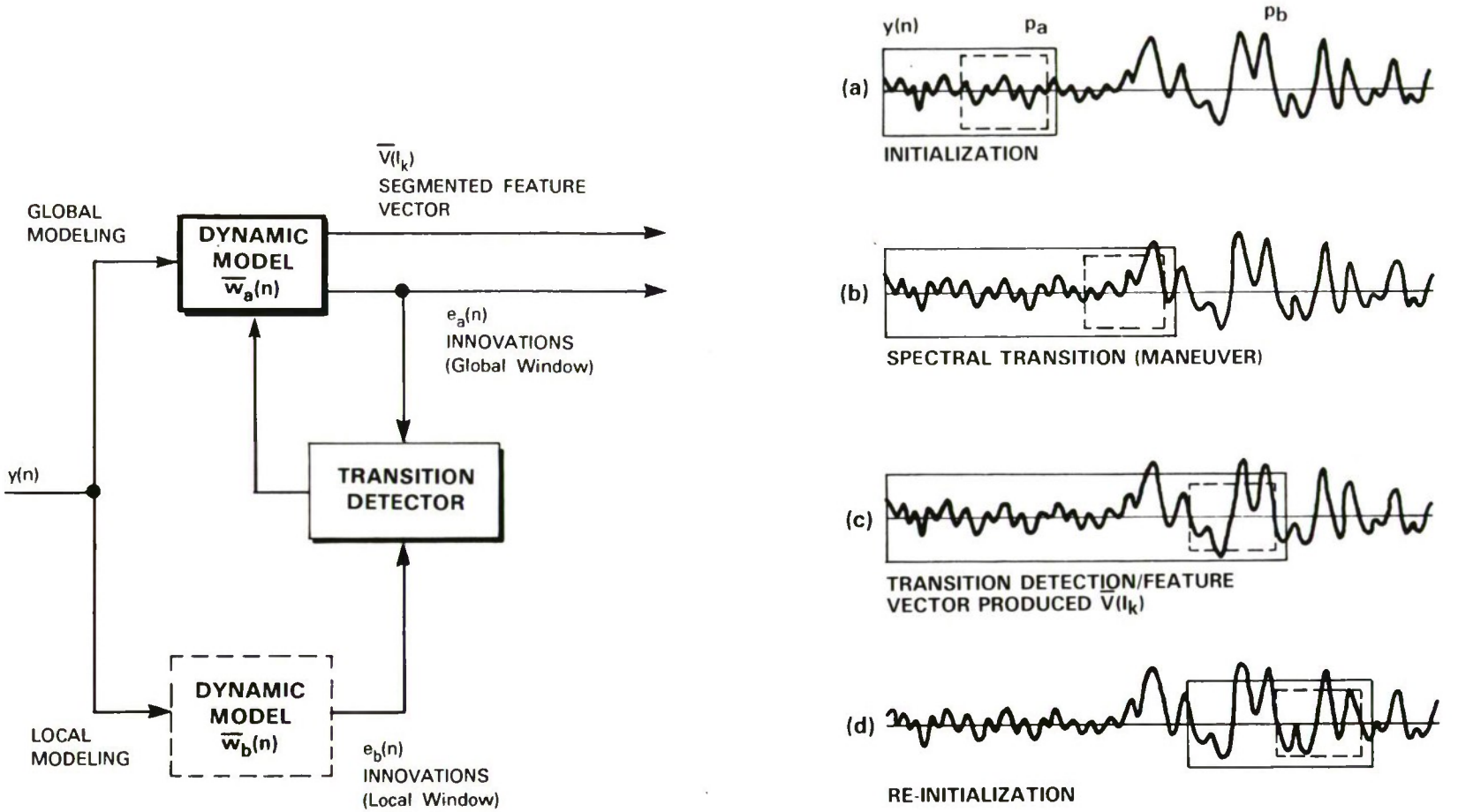


Figure 4-1. Schematic of adaptive recursive preprocessor for nonstationary signals.

118229-3

109165-2

5. RESULTS

Recall that the objective of preprocessing nonstationary signals is to produce a compact representation (feature vector) of the data which simultaneously reduces data and maintains information. More specifically, guided by the decomposition theorem the data are first detrended then segmented into near-stationary intervals. Next, feature vectors containing the AR modeling coefficients are adaptively computed over each interval.

An illustration of this adaptive preprocessing technique is shown in Figure 5-1. As the first transition in spectral characteristics is encountered (at time T_{Δ}^{1}), the cumulative sum statistic ($\mu(n)$)displays a nonzero drift. Upon reaching a threshold (τ) at time T_{d}^{1}, the transition time is estimated ($\hat{T}_{\Delta}^{1}$) and the AR modeling coefficients determined over the now-defined near-stationary segment I_1. The elements comprising the feature vector $\bar{V}(I_1)$ are computed by simply averaging the instantaneous AR coefficients obtained from the adaptive global AR modeling filter over the segment I_1. The process is then reinitialized at ($\hat{T}_{\Delta}^{1}$), and the subsequent feature vector $\bar{V}(I_2)$ over the next interval I_2 (defined by the transition detector) is computed similarly. The net result is a collection of feature vectors, each obtained over a near-stationary interval which can be utilized to characterize the nonstationary signal. Results from applying this general preprocessing technique to real nonstationary data follow. The nonstationary signal utilized is a radar cross section (RCS) versus time record obtained from a radar observing maneuvers from an object of interest. The spectral transitions observed in the data are physically produced by the object undergoing specific maneuvers. Hence, the applied objective of preprocessing is to automatically produce disparate feature vectors representing each of the maneuvers, which can then be used to drive an appropriate classification algorithm, thereby automatically detecting and classifying object maneuvers.

First, the detrending procedures necessary for the specific signals utilized are described. Next, the model order for the adaptive AR modeling filters is determined. Now with a fixed model order, transition detection and feature vector extraction results are presented. Finally, a summary of the preprocessing performance is presented.

5.1 DETRENDING

The RCS data employed are shown in Figure 5-2. An experienced radar analyst categorizes the data into the maneuvers *pitch*, *roll*, and *yaw* (each separated by a *stable* region). Such categorization serves as "truth" and is used later for evaluating the performance of the transition (maneuver) detector. Detrending begins by subtracting out the overall mean, physically corresponding to the mean RCS of the object which is typically known a priori. Next, the data are high-pass filtered to remove low-frequency trends which persist throughout the entire data record. The filtering is achieved by forming a new detrended sequence

$$y(n) \equiv y'(n) - y_{LP}(n) \tag{5.1}$$

where

$$y_{LP}(n) = \alpha y_{LP}(n-1) + (1-\alpha)\, y'(n) \tag{5.2}$$

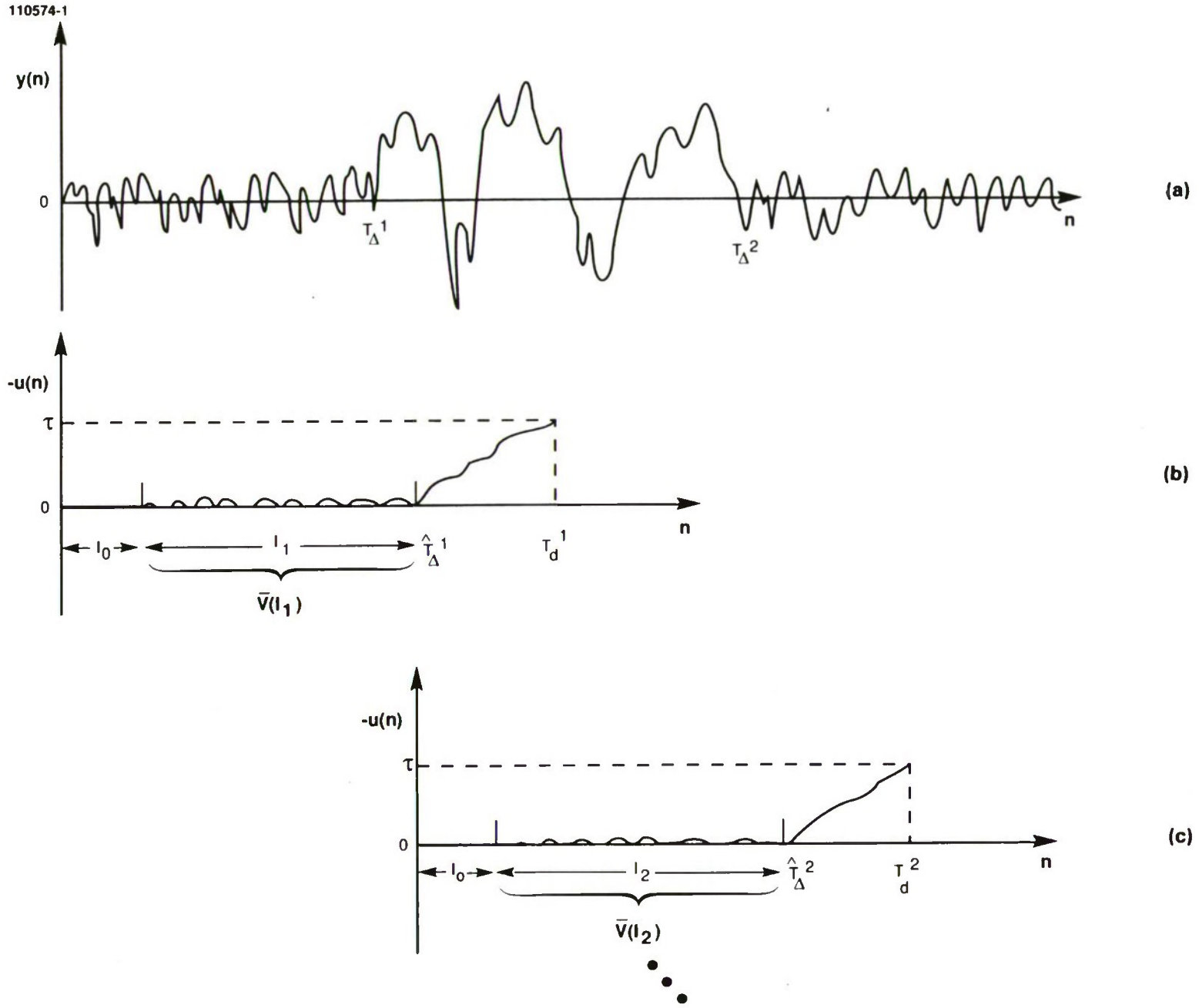


Figure 5-1. General adaptive preprocessing procedures (a) transition detected at time T_d; (b) transition time estimated ($\hat{T}_\Delta$); (c) AR coefficients computed $|V(l_k)|$; and reinitialization

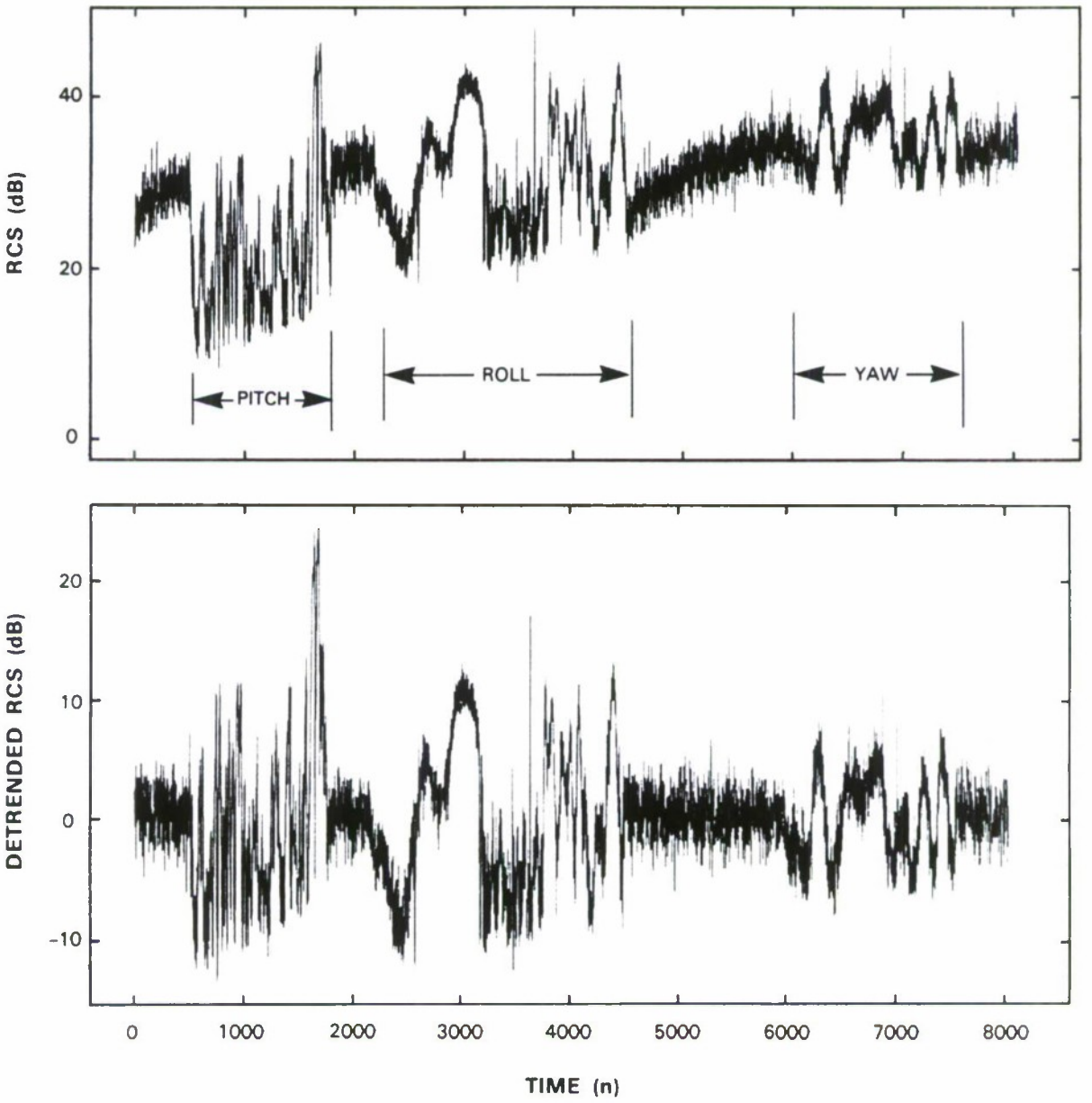


Figure 5-2. Original and detrended radar data.

is a unity DC-gain low-pass version of the zero mean RCS data $y'(n)$. Typically with such data, the low-frequency components common through the record can be removed by modest filtering ($\alpha = 0.9999$). Consequently, a zero mean, high-pass, detrended sequence results (as shown in Figure 5-2), and that drives the subsequent AR modeling filter and transition detector.

5.2 MODEL ORDER

Before the adaptive modeling filter can be driven by the detrended time series, the model order (equivalently, the number of tap weights for the adaptive transversal filter) must be determined. Ideally, the order is chosen to provide parsimonious modeling [12]. That is, a model order is sought which provides an acceptable compromise between model performance and complexity.

Two conventional criteria for model order selection were investigated. The AIC criterion, both theoretically intuitive and practically effective, minimizes the distance between the true and observed distributions. For the AR formulation, AIC is given by [13]

$$AIC\ (Q) = N\ \ln\ \sigma_e^2 + 2Q \tag{5.3}$$

where

$$Q = \text{AR model order} \tag{5.4}$$

$$N = \text{number of samples} \tag{5.5}$$

$$\sigma_e^2 = \text{modeling error variance.} \tag{5.6}$$

Notice the first term in Equation (5.3) serves to penalize poor modeling, while the second serves increased complexity (through the number of tap weights Q). The AIC optimal model order is simply the order that minimizes Equation (5.3). Although intuitive and often effective, the AIC optimal model order is not consistent, and hence does not necessarily converge to the true model order.

A consistent model order estimator investigated is the minimum description length (MDL) given by [14]

$$MDL(Q) = \frac{N}{2} \ln\ \sigma_e^2 + \frac{Q}{2} \ln\ N \tag{5.7}$$

This criterion minimizes the number of digits necessary to encode N observations and often results in a lower model order than AIC. Similarly, MDL optimal order is chosen to minimize Equation (5.7).

Both techniques for determining model order yielded identical optimal order. The similarity can be attributed to the relatively large number of data samples ($N \approx 8000$). Observing Equations (5.3) and (5.7), for large N the first terms dominate, causing AIC and MDL values to vary by only a constant factor. Consequently, minimal values denoting optimal order were achieved at identical orders. The results are shown in Figure 5-3 along with the AIC and MDL optimal model orders for each maneuver.

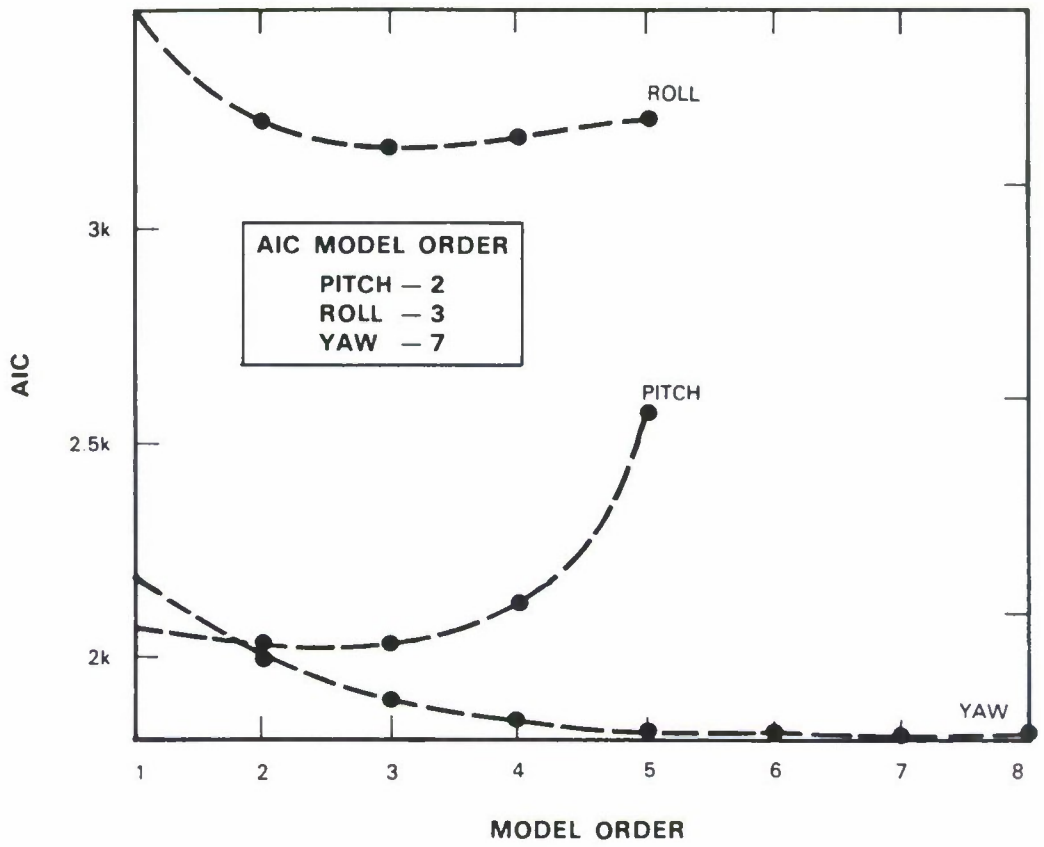


Figure 5-3. AIC model order results.

Unfortunately, the optimal model orders vary for each maneuver. Therefore, to achieve constant model order throughout all maneuvers (constant model order is desired for fast automatic preprocessor operation without a priori maneuver knowledge), an alternate technique for determining the "best" model order is required. Specifically, the best model order chosen represents the lowest order model which provides good discrimination amongst the feature vectors (AR coefficients) for the various maneuvers. From Table 5-1, the best model order for the RCS data is seen to be 4, since little disparity amongst the feature vectors exists for lower orders. Note that M = 4 is also the average model order amongst the optimal orders selected by AIC and MDL for the maneuvers.

TABLE 5-1

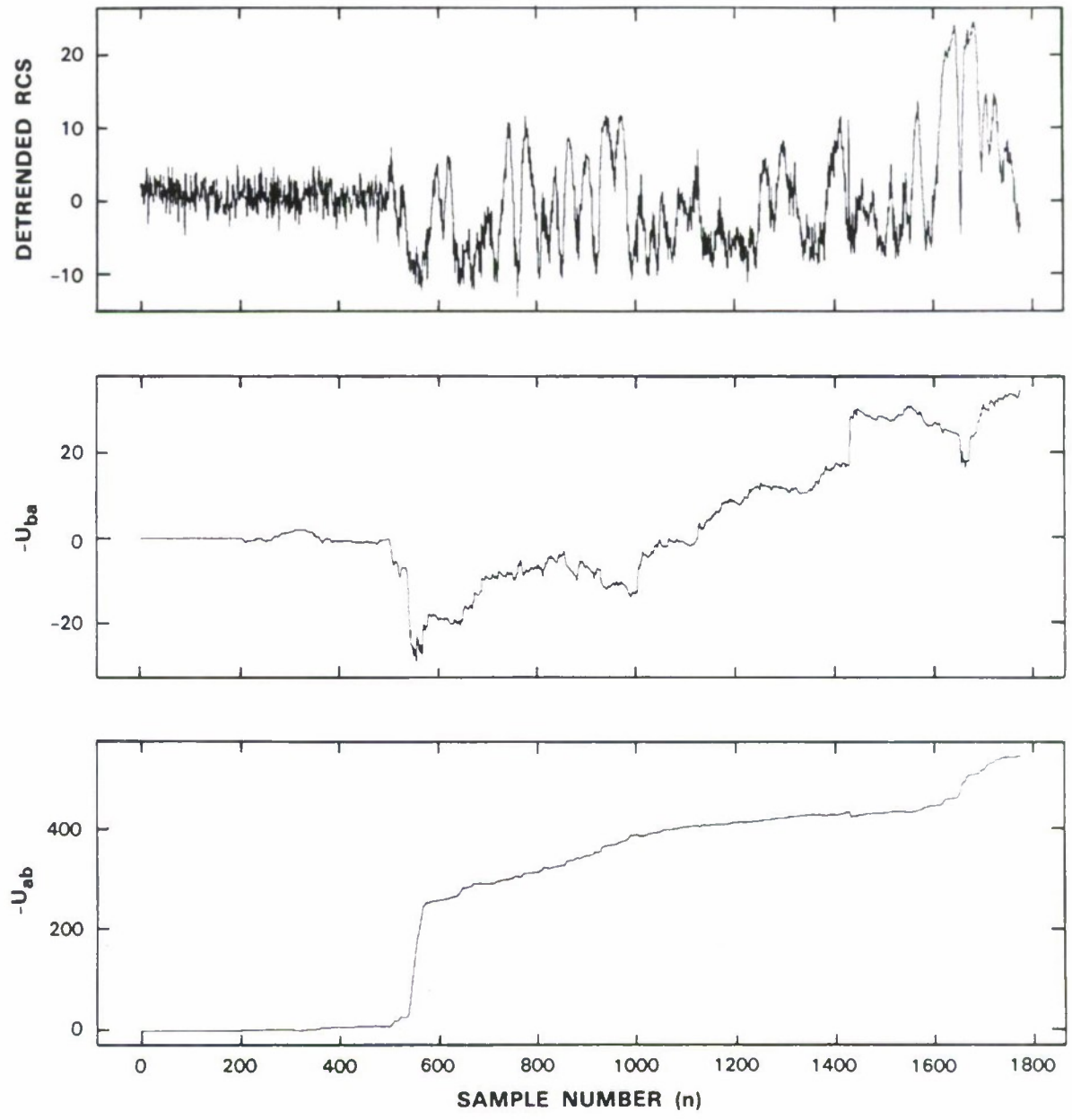
AR Cofficients for Varying Model Order

Model Order	STABLE	PITCH	ROLL	YAW
1	0	0.93	0.91	0.79
2	0	0.61	0.57	0.47
	0	0.29	0.39	0.37
3	0	0.61	0.41	0.37
	0	0.30	0.25	0.24
	0	0.04	0.22	0.25
4	0	0.58	0.39	0.32
	0	0.32	0.23	0.19
	0	0.04	0.14	0.17
	0	-0.06	0.06	0.21

5.3 PREPROCESSOR RESULTS

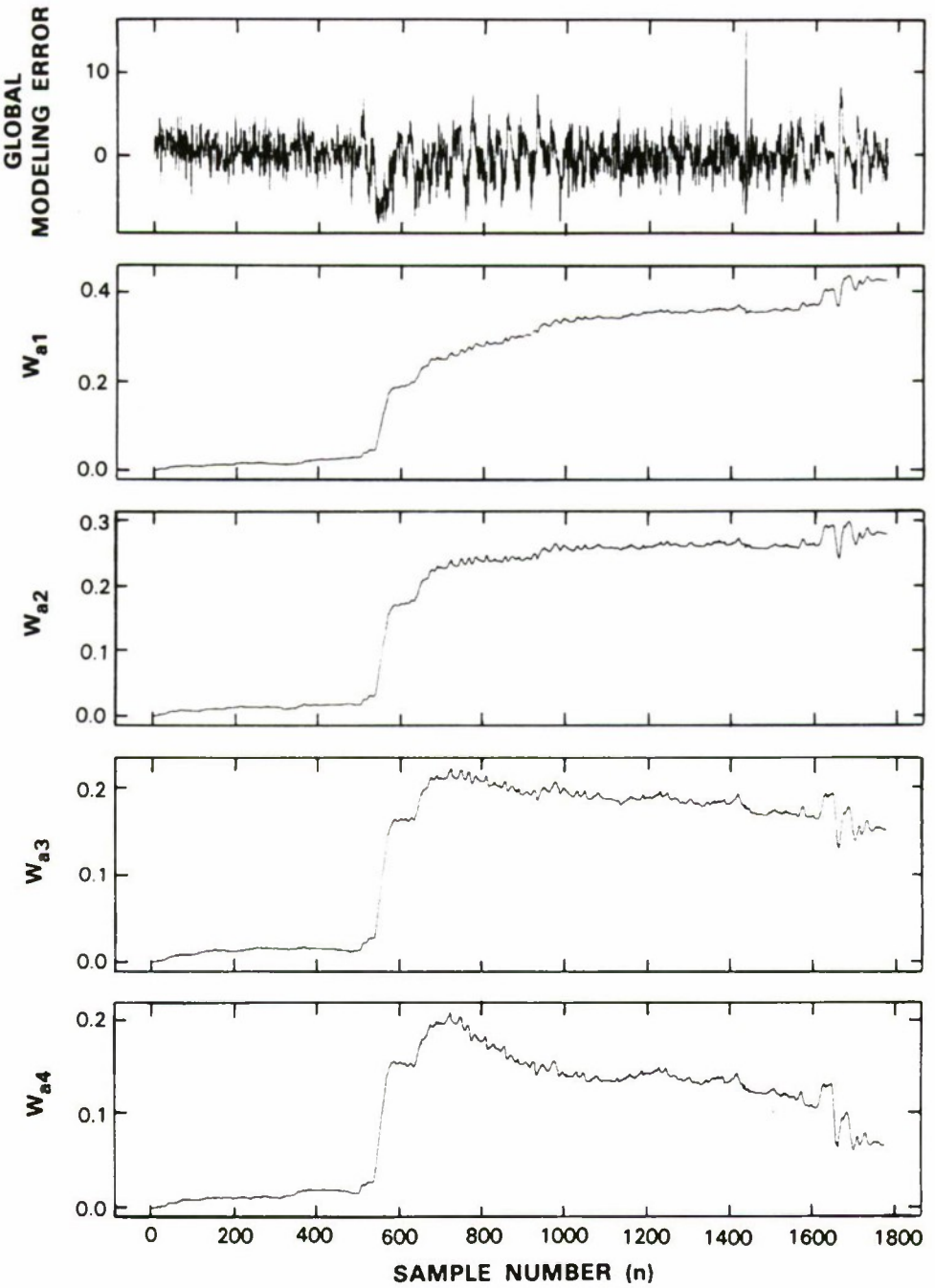
With the RCS data detrended and model order now selected, transition detection and feature vector extraction can be performed. Expanded results are presented in the region of each transition, thereby enabling both evaluation of the transition detector and the production of the feature vectors. For each transition, a portion of the detrended signal is shown with two cumulative sum statistics. The first, u_{ab}, follows from Equation (3.8) and exhibits the desired drift behavior for transitions from low to high variance. The second statistic u_{ba} is computed in parallel by simply interchanging the roles of η_a and η_b in Equation (4.4), offering better drift behavior for high- to low-variance transitions. A threshold of $\tau = 100$ was used throughout for both statistics. On subsequent graphs for each transition, the instantaneous AR coefficients (tap weights of the adaptive transversal filters) for both the global and local modeling filters with accompanying model error (innovations) are shown. Averaging the AR coefficients over the detected intervals constitutes the corresponding feature vector.

Stable-Pitch Transition: Results are shown in Figures 5-4 to 5-6. The transition occurs at n = 500 and is detected ($-u_{ab}$ exceeds threshold) near 600. Subsequent estimation of the transition location $\hat{f}_{\Delta}^{1}$ [initial point of positive drift in $-u_{ab}$, or equivalently, the local minimum of Equation (3.8)] is within samples of the actual location. The feature vector representing the *stable* region is computed by averaging the instantaneous AR global model coefficients over the first interval detected $(0 - \hat{f}_{\Delta}^{1})$, yielding approximately $\bar{V}^{T}(I_1) = (0, 0, 0, 0)$. Hence the stable region is modeled as white noise.



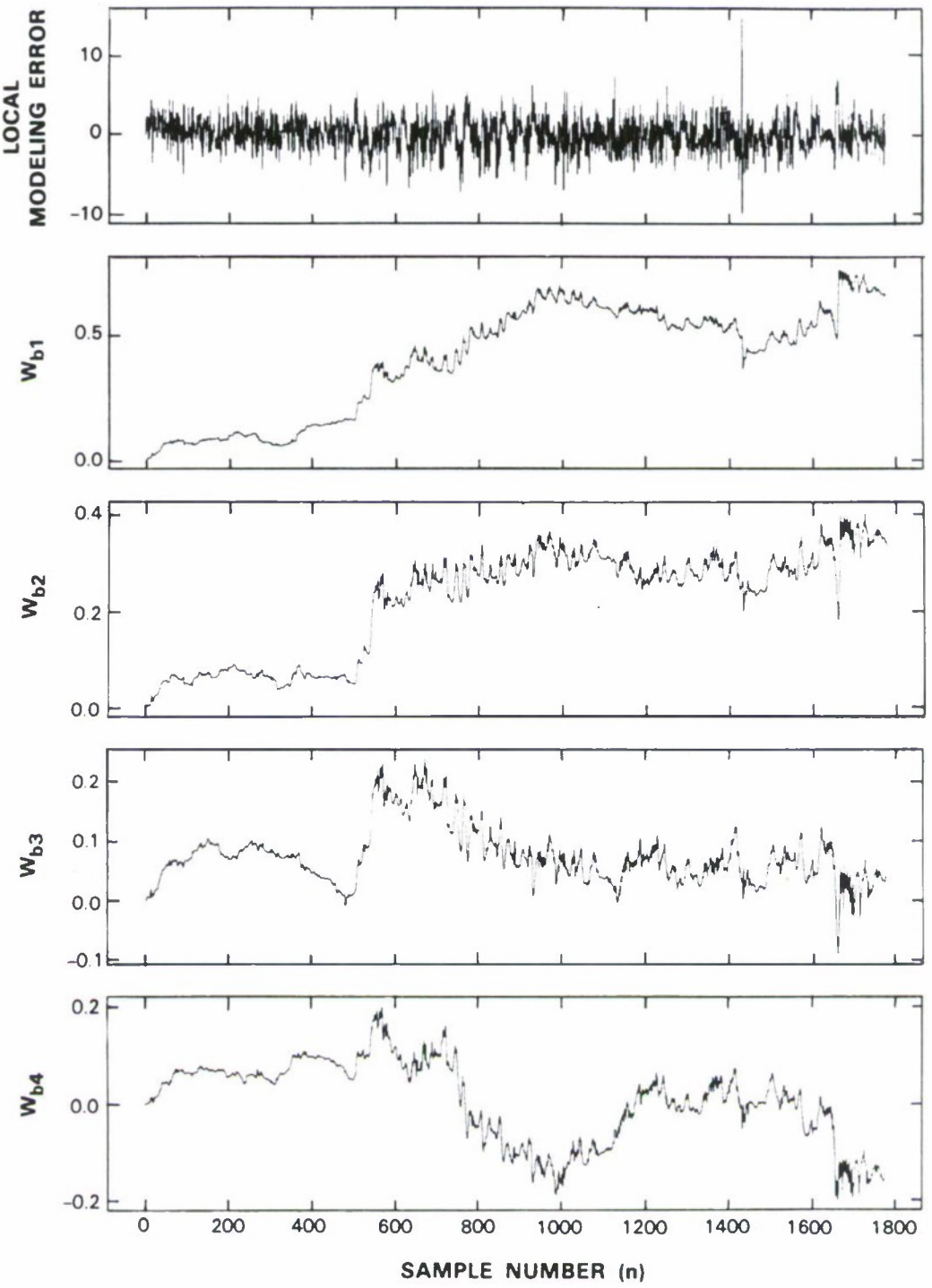
116884-21

Figure 5-4. Stable-pitch transition detection results.



116884-20

Figure 5-5. Stable-pitch global modeling results.



116884-22

Figure 5-6. Stable-pitch local modeling results.

Pitch-Stable Transition: Results are shown in Figures 5-7 to 5-12. The transition occurs at n = 1276, and the appropriate test statistic u_{ba} (for high- to low-variance transitions) fails to surpass threshold, thereby being incapable of detecting the transition within acceptable delay time. However, by examining the behavior of u_{ba} , modifications can be employed to ensure detection.

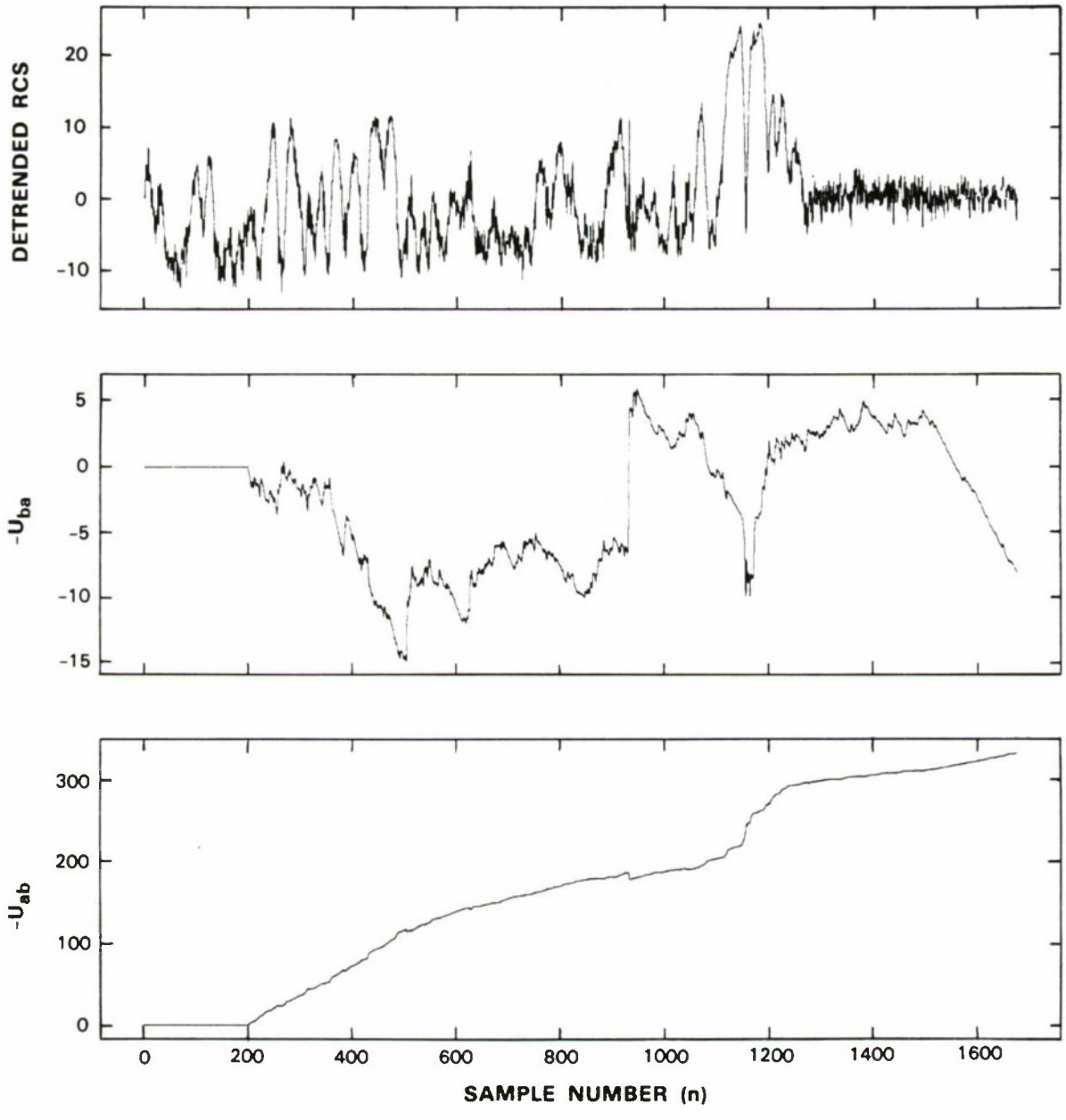
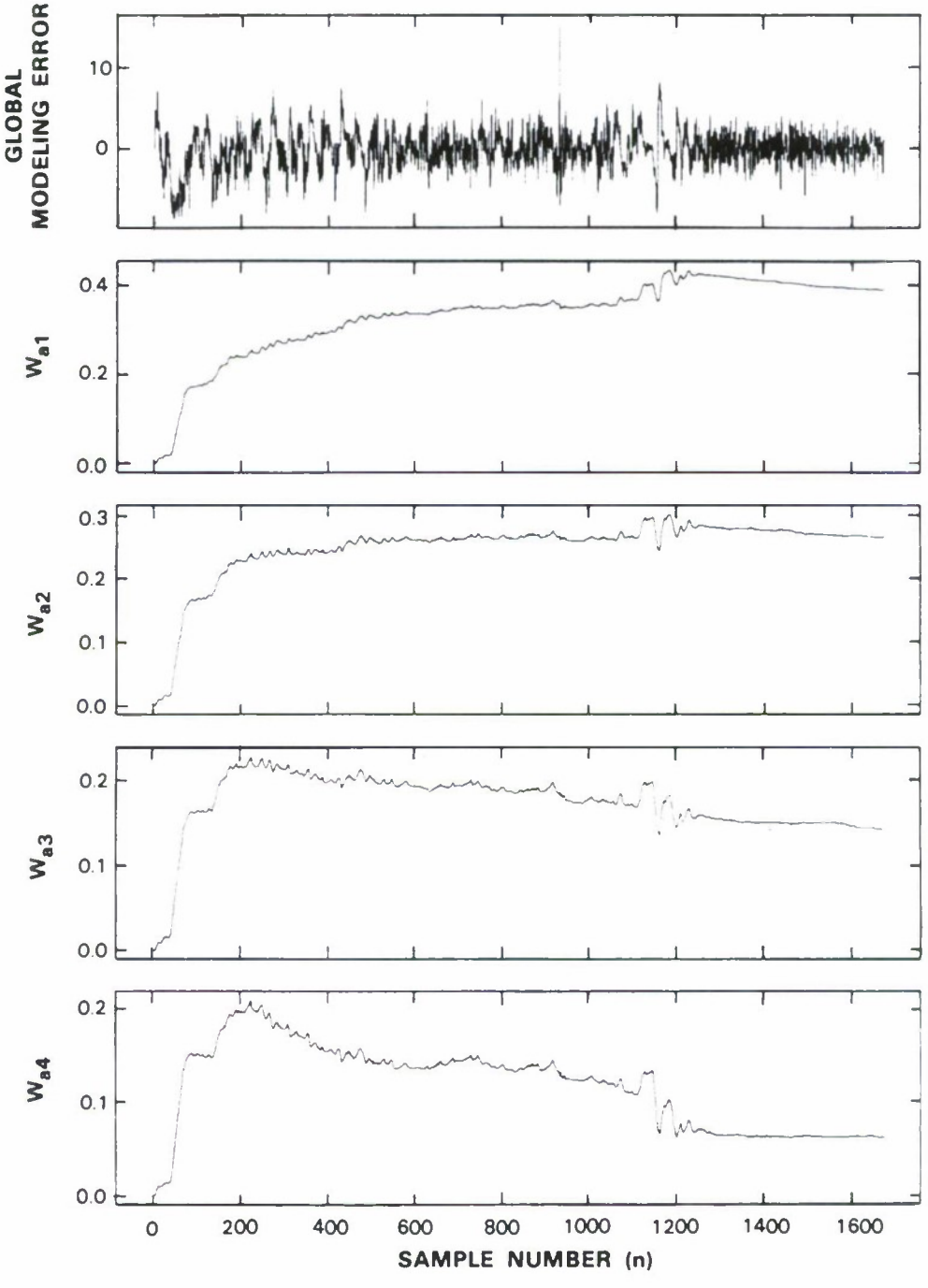
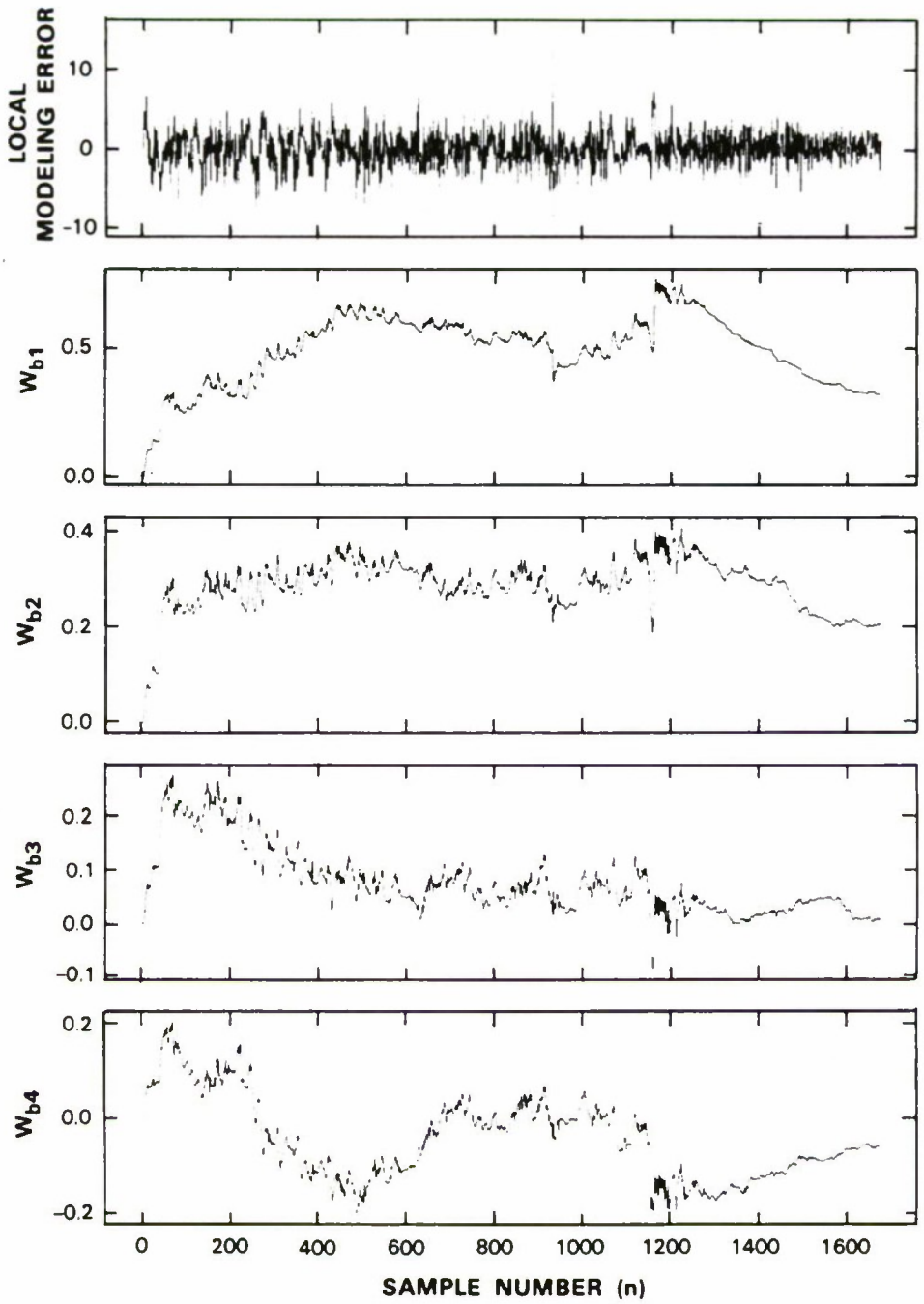


Figure 5-7. Pitch-stable transition detection results.



116884-12

Figure 5-8. Pitch-stable global modeling results.



116884-13

Figure 5-9. Pitch-stable local modeling results.

The behavior observed is due to the test statistic $u\,(n\,)$ having an asymmetrical unconditional drift. That is, the drift in $u\,(n\,)$ experienced by a signal transition from a distribution p_a to p_b is not equivalent in magnitude to the drift incurred from the transition p_b to p_a. Specifically, for the signal analyzed (see Figure 5-7), a transition from high to low variance (signal power) yields slow linear behavior in u_{ba} following the transition.

The linear property follows from the post-transition signal (having low variance) being relatively small in magnitude. From Equation (4.3), small $y\,(n\,)$ yields small innovations $e\,(n\,)$, since typically $|w_i| \leq 1$. Furthermore, by examining Equation (4.4), small innovations $e\,(n\,)$ coupled with small signals $\bar{y}\,(n-1)$, cause very little adaption in filter tap weights $(\bar{w}\,)$; (as verified by Figures 5-8 and 5-9). Now with the filters practically fixed $[e_a(k\,) \approx e_b(k\,)]$ and the innovations $e_a(k\,)$ and $e_b(k\,)$ small, the statistic $T\,(k\,)$ is practically reduced to [see Equation (4.6)]

$$T\,(k\,) \approx \frac{1}{2}\left[1 - \sigma_{e_a}^2/\sigma_{e_b}^2\right], \tag{5.8}$$

and consequently, the cumulative sum statistic reduces to

$$u_{ba}(n) \approx \frac{1}{2}\sum_{k=1}^{n}\left(1 - \sigma_{e_b}^2/\sigma_{e_a}^2\right) \tag{5.9}$$

resulting in the linear behavior for $\sigma_{e_a}^2 > \sigma_{e_b}^2$.

One approach to circumventing such behavior is to normalize the signal. Now the post-transition signal is sufficiently large to drive the filter adaption through the innovation process. Results from the normalized *pitch-stable* transition are shown in Figures 5-10 to 5-12. Notice that the transition is detected near 1330, and the transition estimation $\hat{T}_\Delta^2$ is within samples of the actual location.

The feature vector representing the pitch region is computed by averaging the instantaneous AR coefficients over the second interval $(\hat{T}_\Delta^1 - \hat{T}_\Delta^2)$ yielding $\bar{V}^T\,(I_2) = (0.58, 0.32, 0.04, -0.06)$. Having only two significantly nonzero components tends to substantiate the AIC and MDL optimal second order.

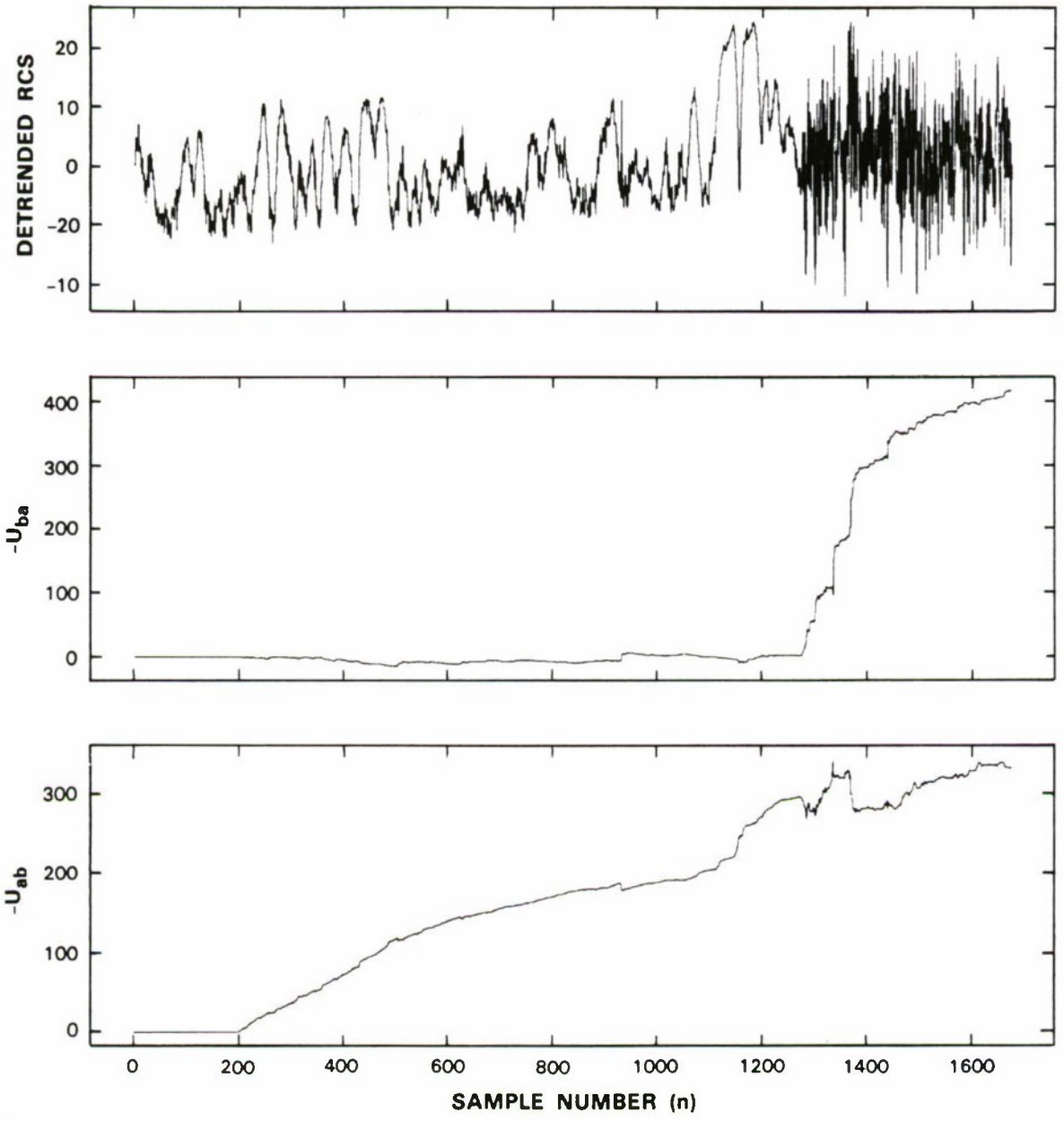
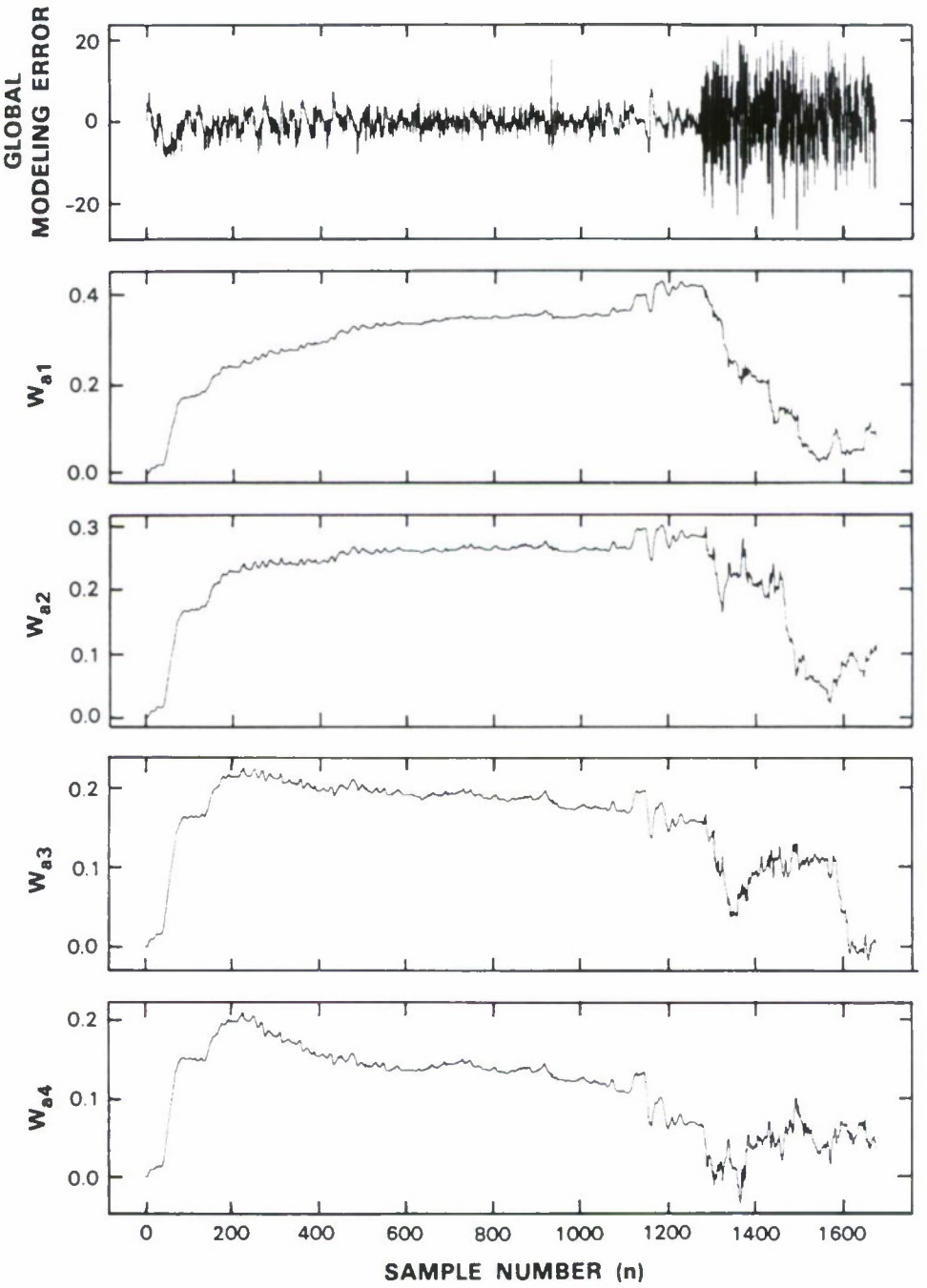
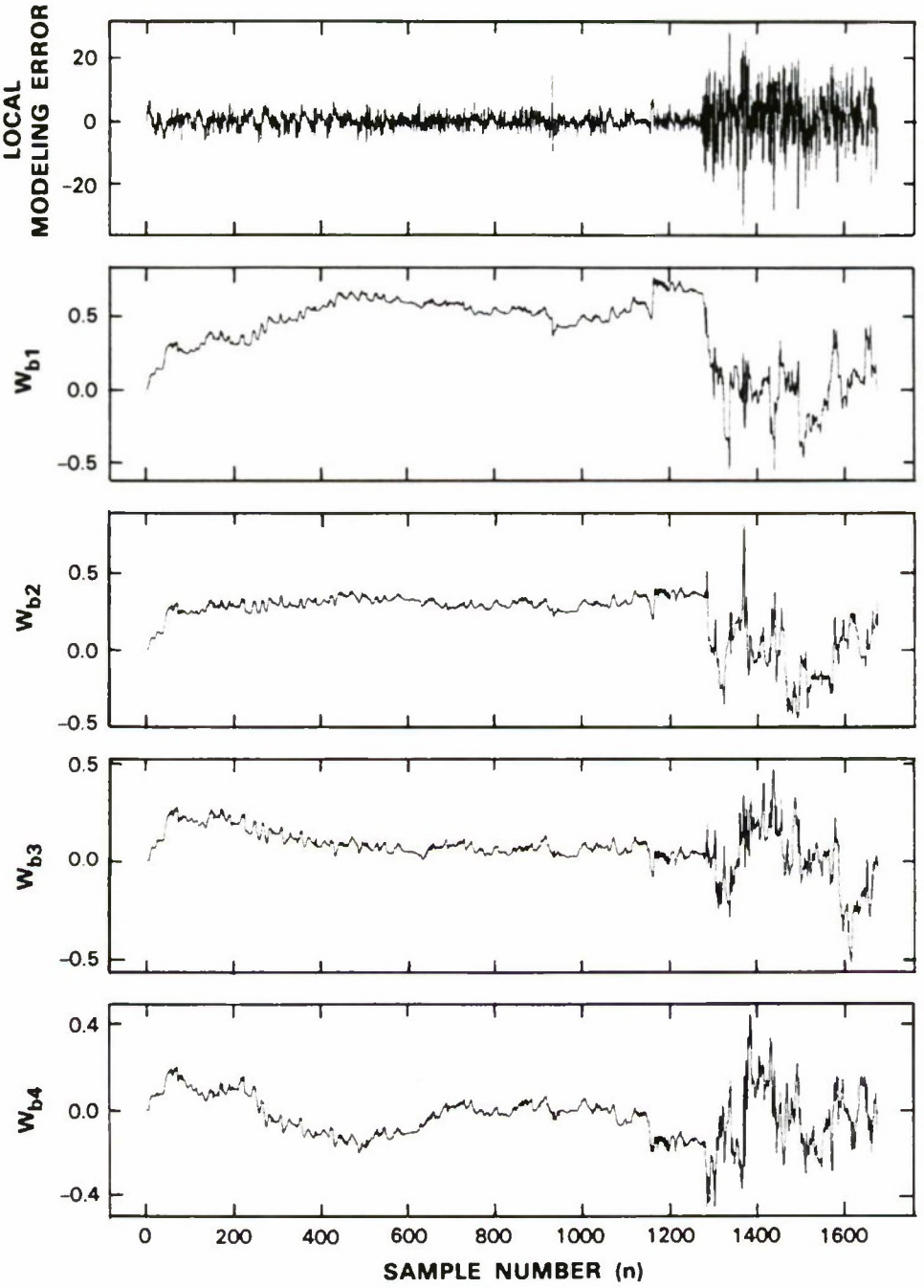


Figure 5-10. Normalized pitch-stable transition detection results.



116884-18

Figure 5-11. Normalized pitch-stable global modeling results.



116884-19

Figure 5-12. Normalized pitch-stable local modeling results.

Stable-Roll Transition: Results are shown in Figures 5-13 to 5-15. The transition occurs at n = 400, detected near 500, and location estimated $\hat{T}_{\Delta}^{3}$ within samples of the actual occurrence. Consequently, another feature vector representing the stable region is computed over the third interval $(\hat{T}_{\Delta}^{2} - \hat{T}_{\Delta}^{3})$ similarly yielding approximately $\bar{V}^{T}(I_3) = \bar{0}$.

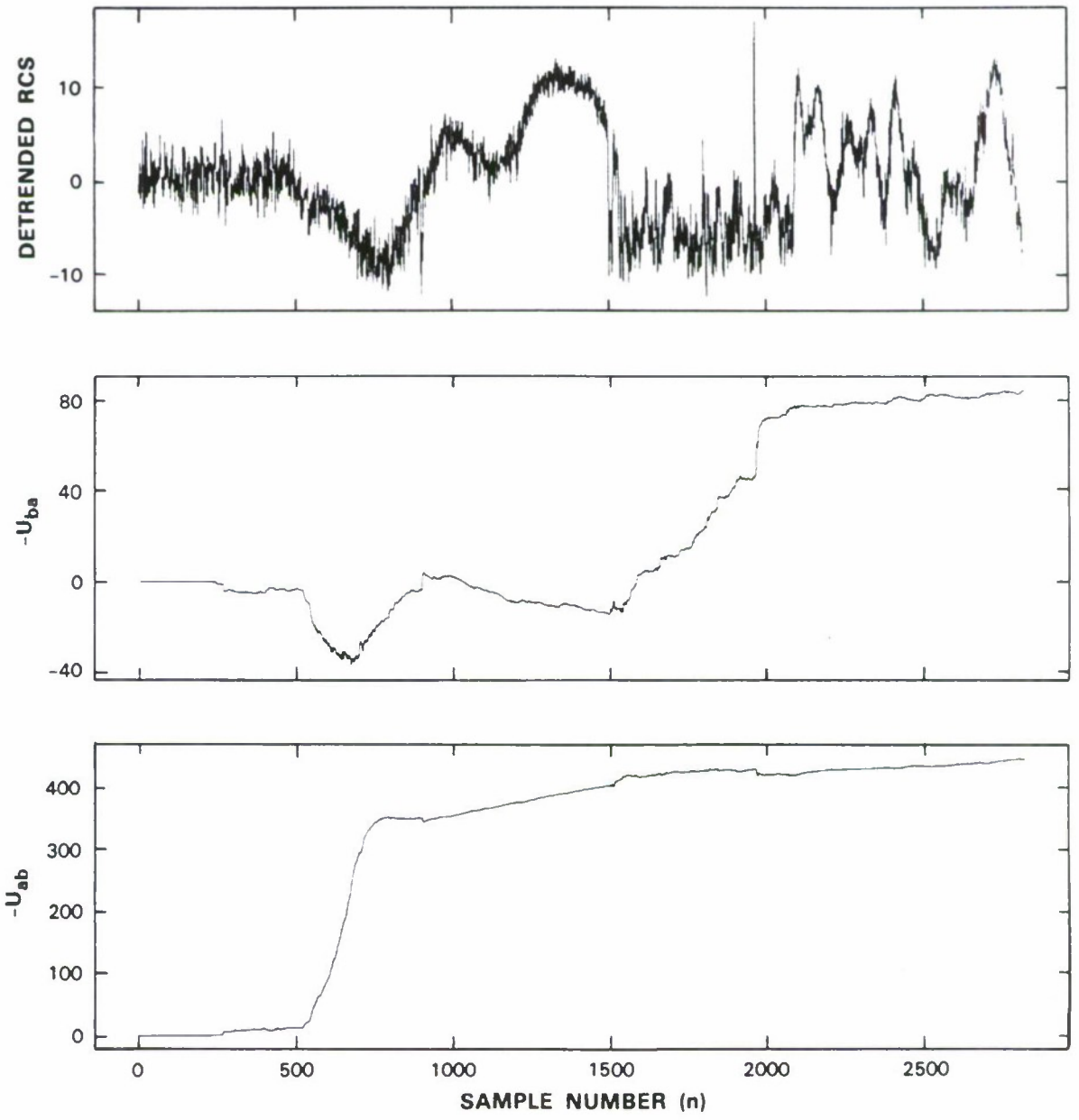
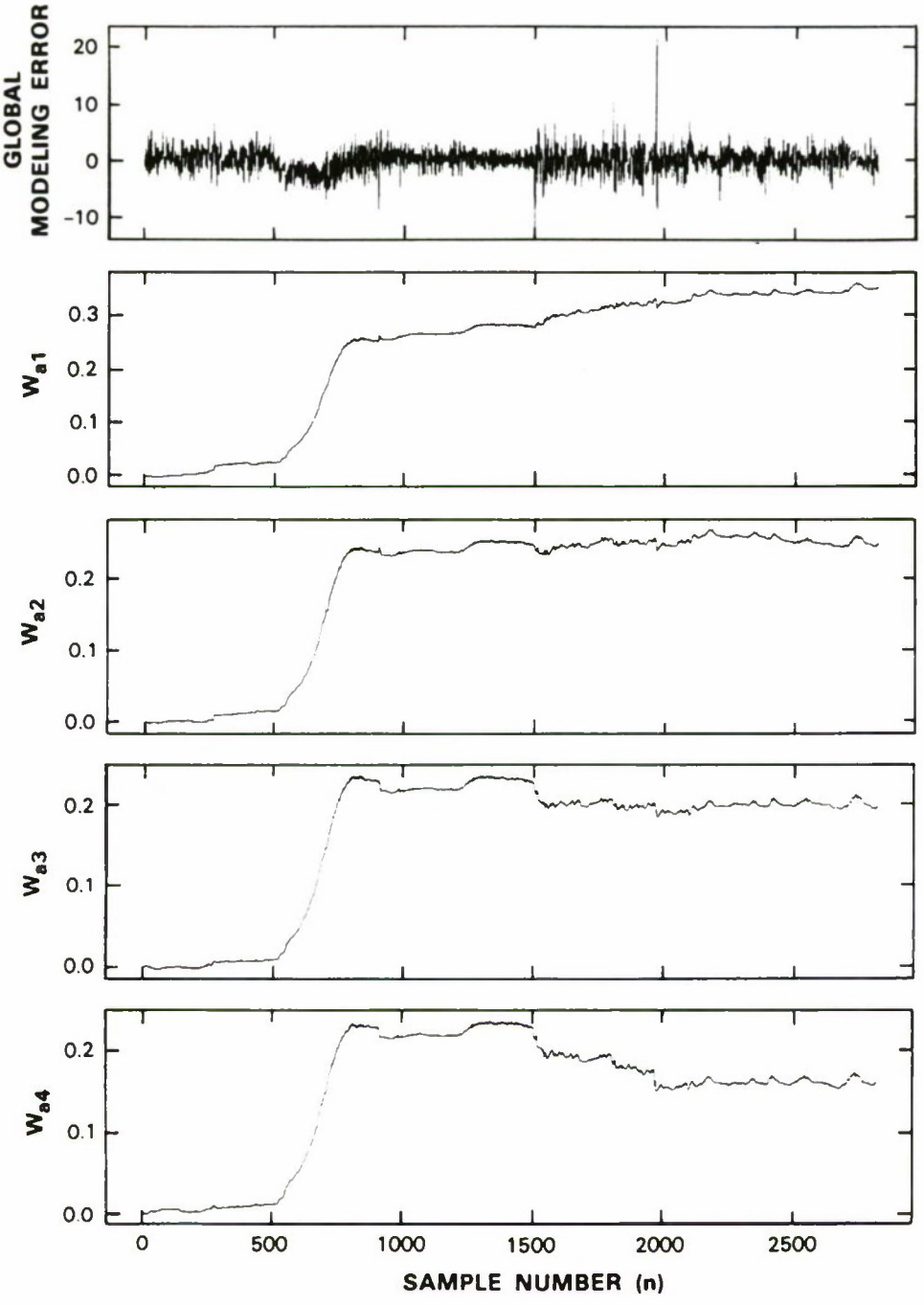
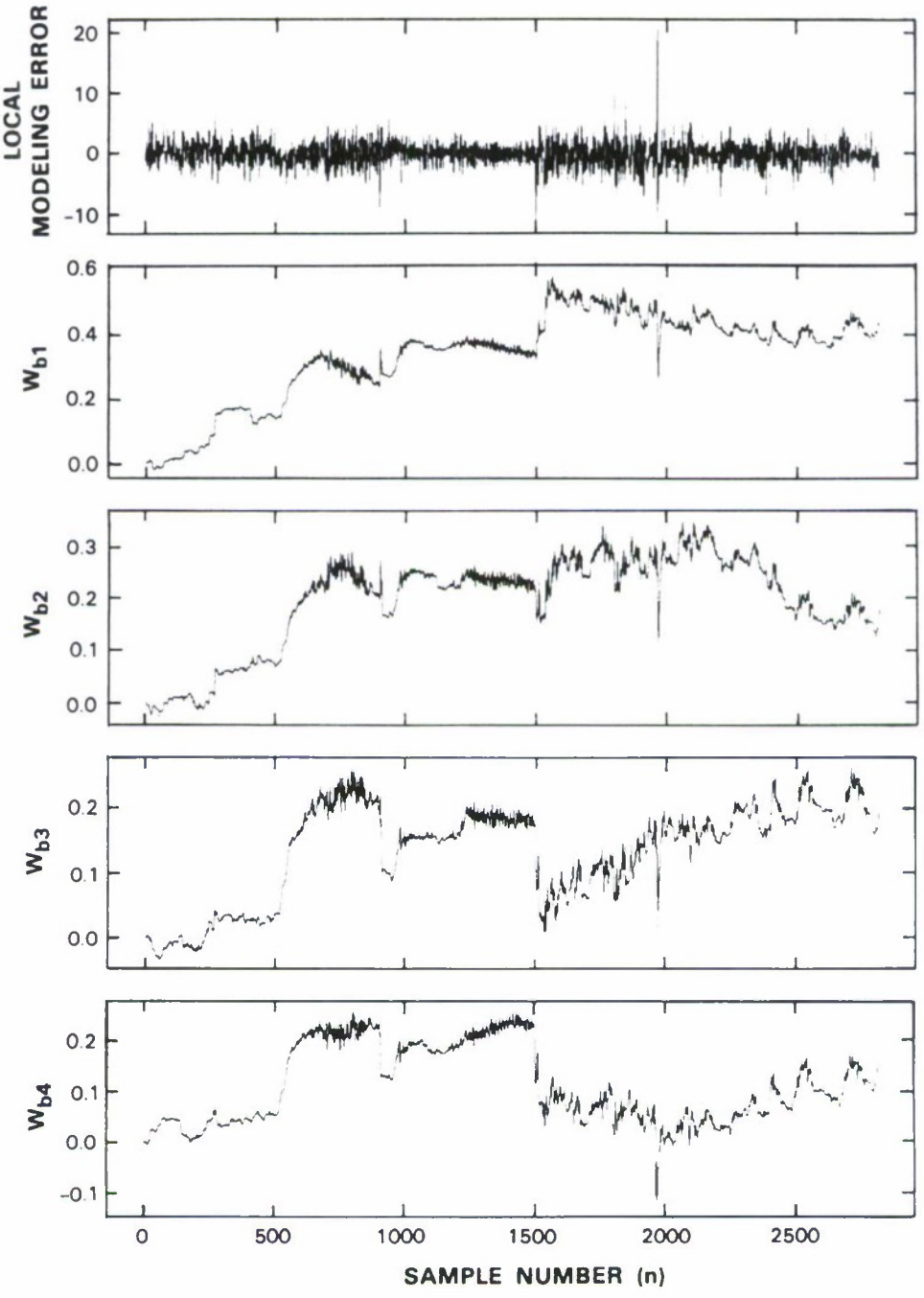


Figure 5-13. Stable-roll transition detection results.



116884-15

Figure 5-14. Stable-roll global modeling results.



116884-16

Figure 5-15. Stable-roll local modeling results.

Roll-Stable Transition: Results are shown in Figures 5-16 to 5-19. The transition occurs at n = 2319, and the appropriate statistic u_{ba} (for high- to low-variance transition) fails to surpass threshold. However, for the same reasons described in the pitch-stable transition section, the performance can be improved by normalizing the signal.

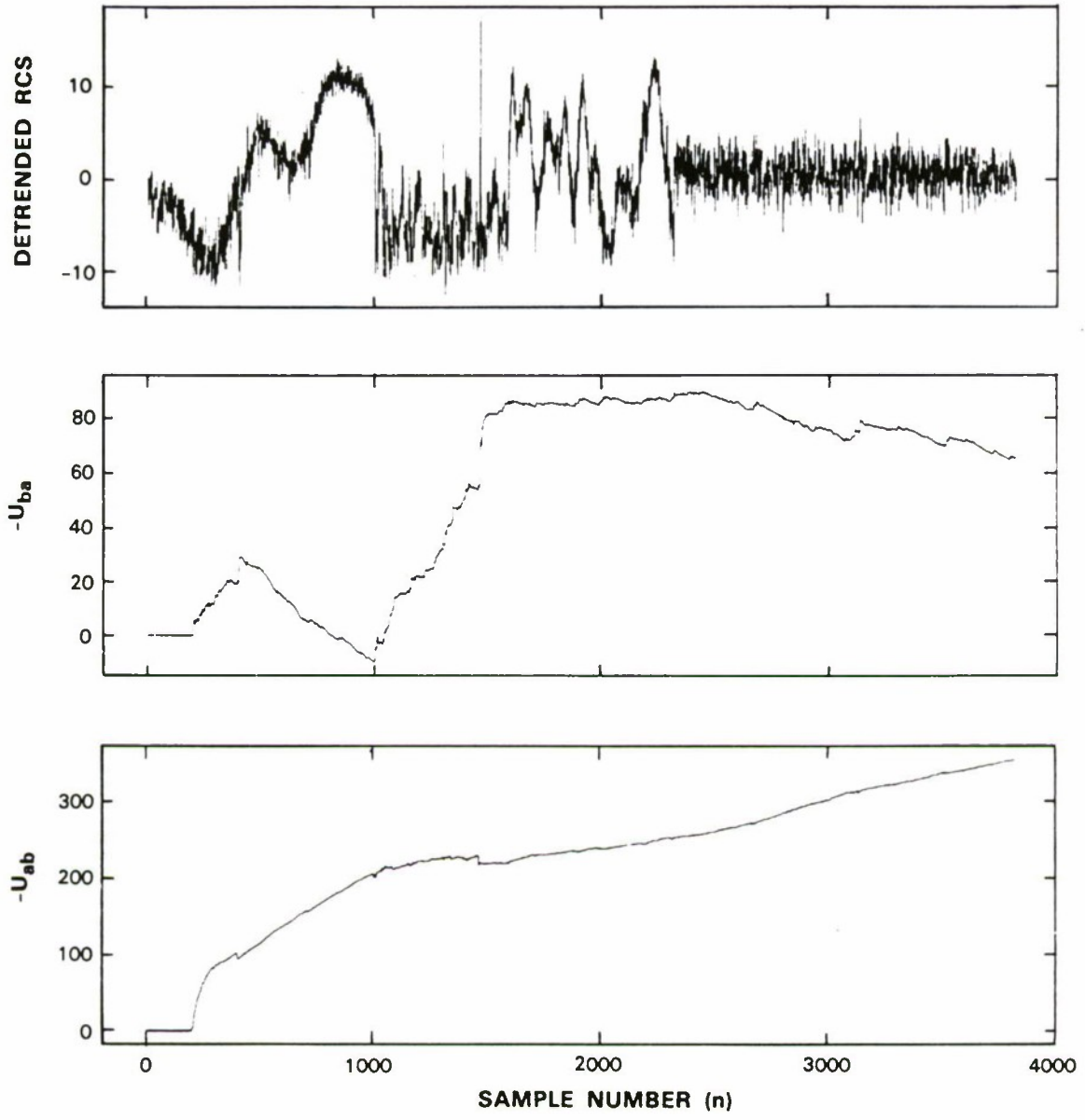
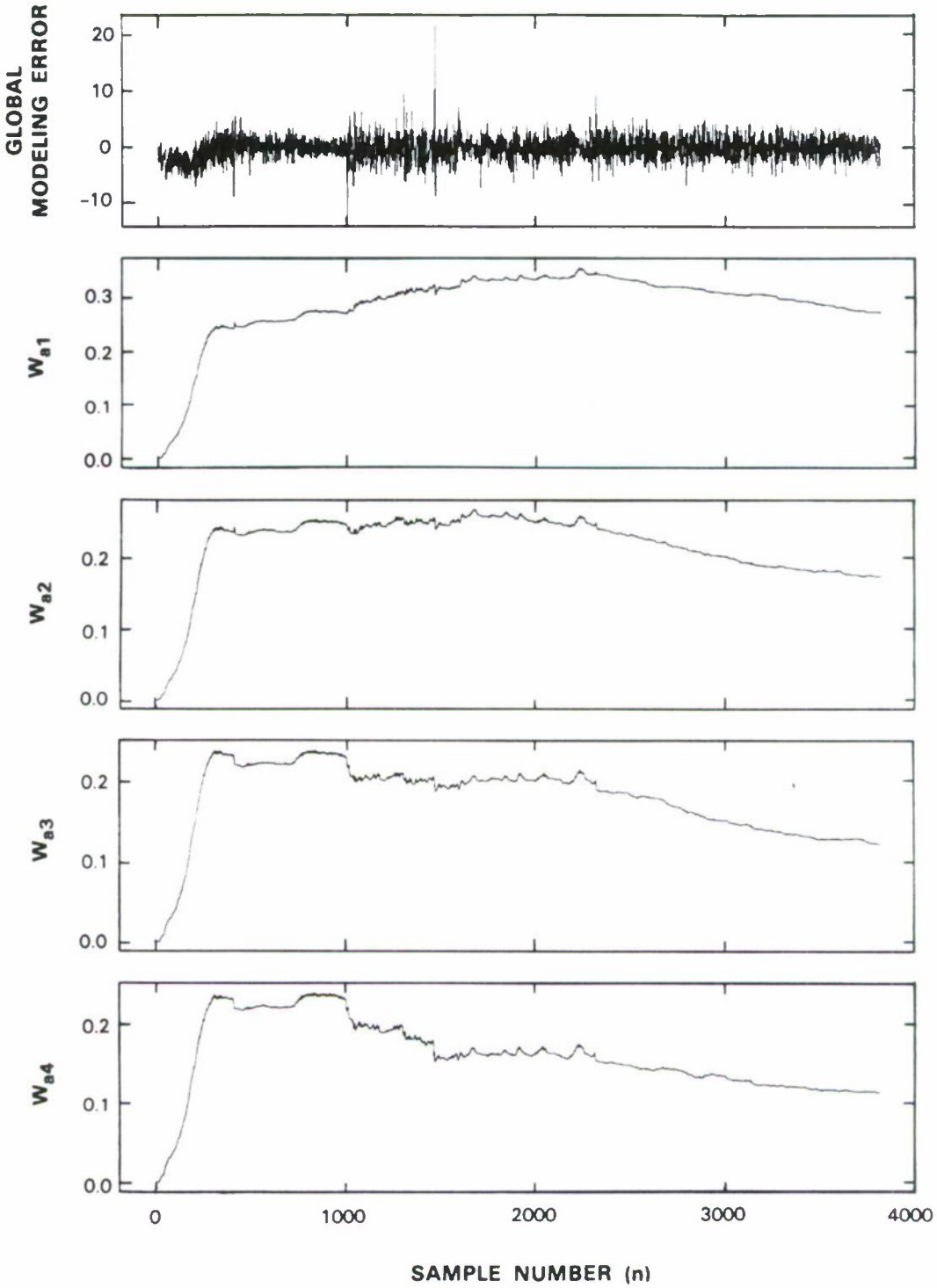


Figure 5-16. Roll-stable transition detection results.



116884-9

Figure 5-17. Roll-stable global modeling results.

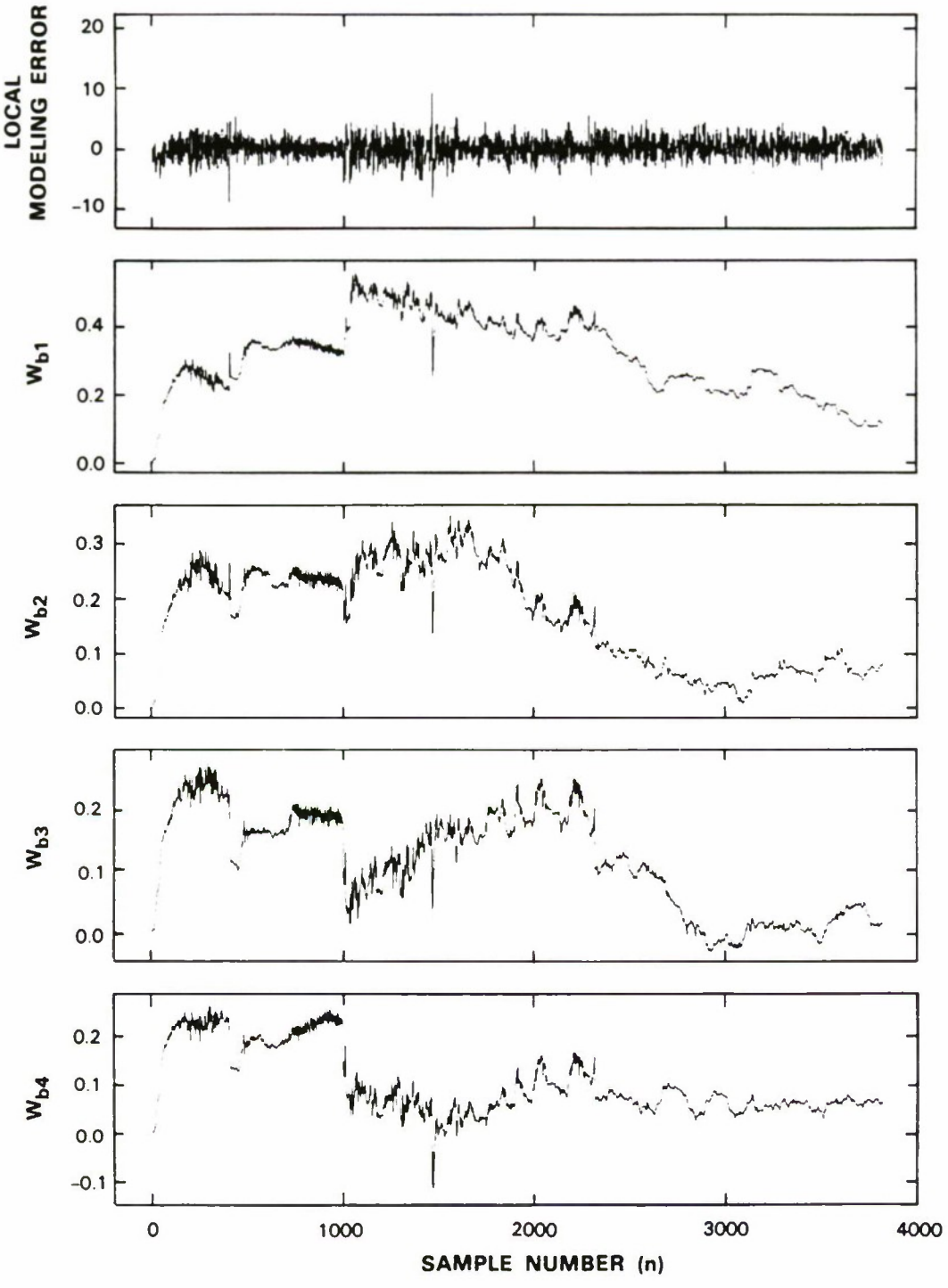


Figure 5-18. Roll-stable local modeling results.

Normalized roll-stable transition results are shown in Figures 5-19 to 5-21. Notice that two significant regions of positive drift are exhibited by $-u_{ba}$, although only the last drift exceeds threshold. The latter coincides with the actual transition, yielding an estimated $\hat{T}_{\Delta}^{4}$ near the truth. The former initiated at n = 1000 appears to indicate a false transition. However, conversations with an experienced analyst [15] revealed that typical *roll* maneuvers comprise three submaneuvers, representing acceleration, constant velocity, and deceleration. Consequently, the transition detector is attempting to detect submaneuver transitions as well.

The feature vector representing the roll maneuver over the interval $(\hat{T}_{\Delta}^{3} - \hat{T}_{\Delta}^{4})$ is $\bar{V}^{T}(I_3)$ $= (0.39, 0.23, 0.14, 0.06)$. Here again, the relative magnitudes of the feature vector components substantiate the optimal AIC and MDL optimal model order.

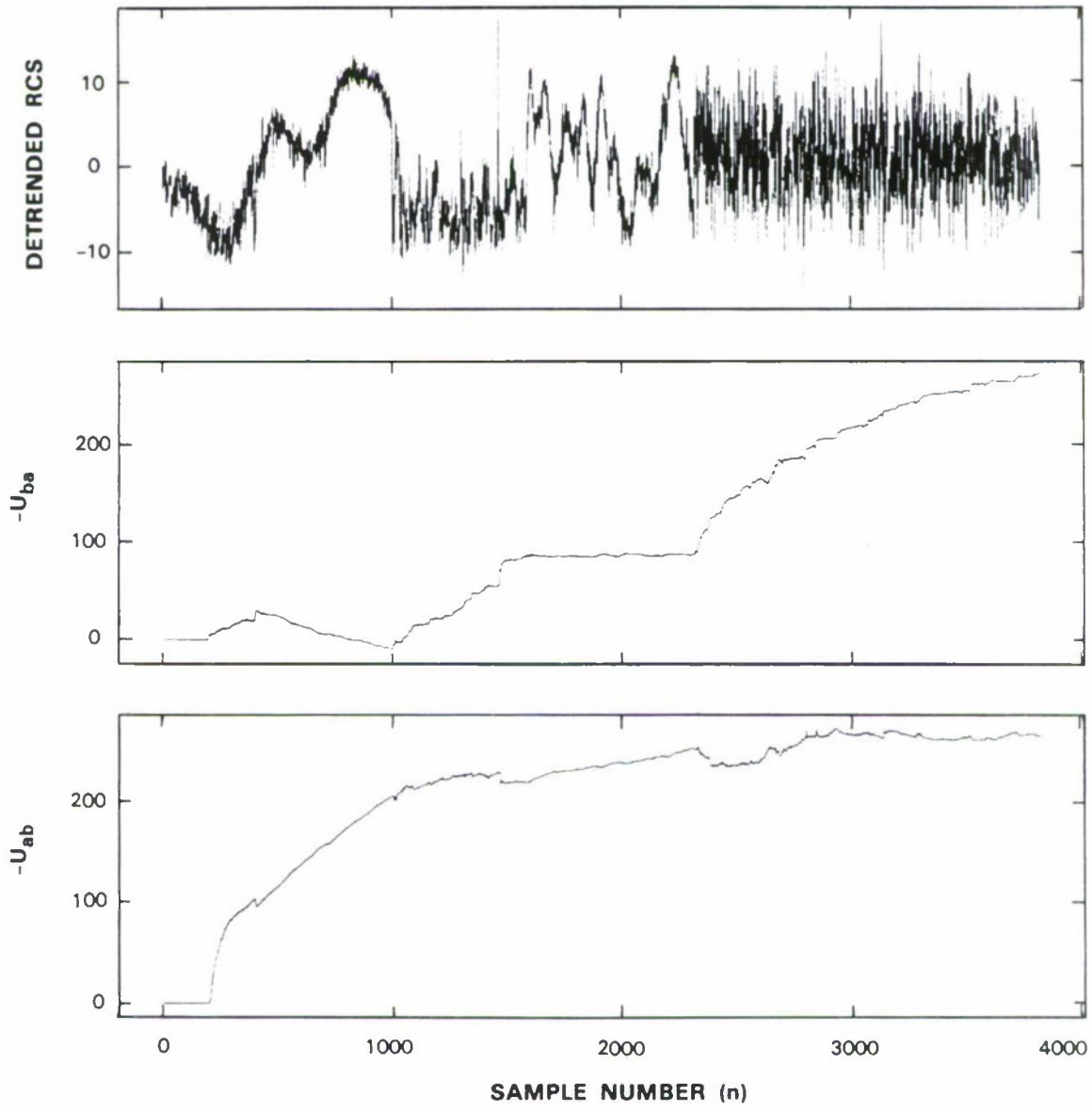


Figure 5-19. Normalized roll-stable transition detection results.

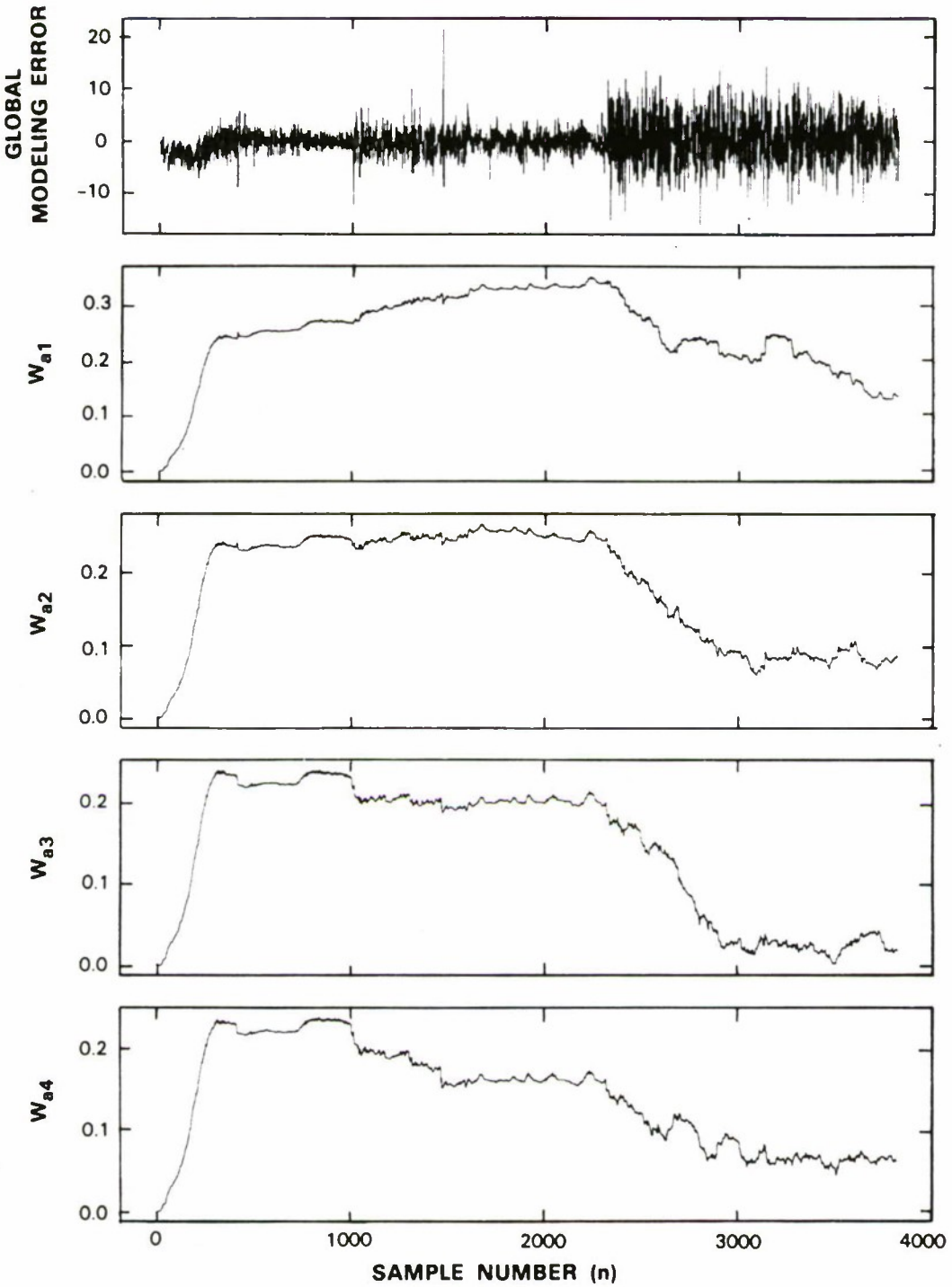
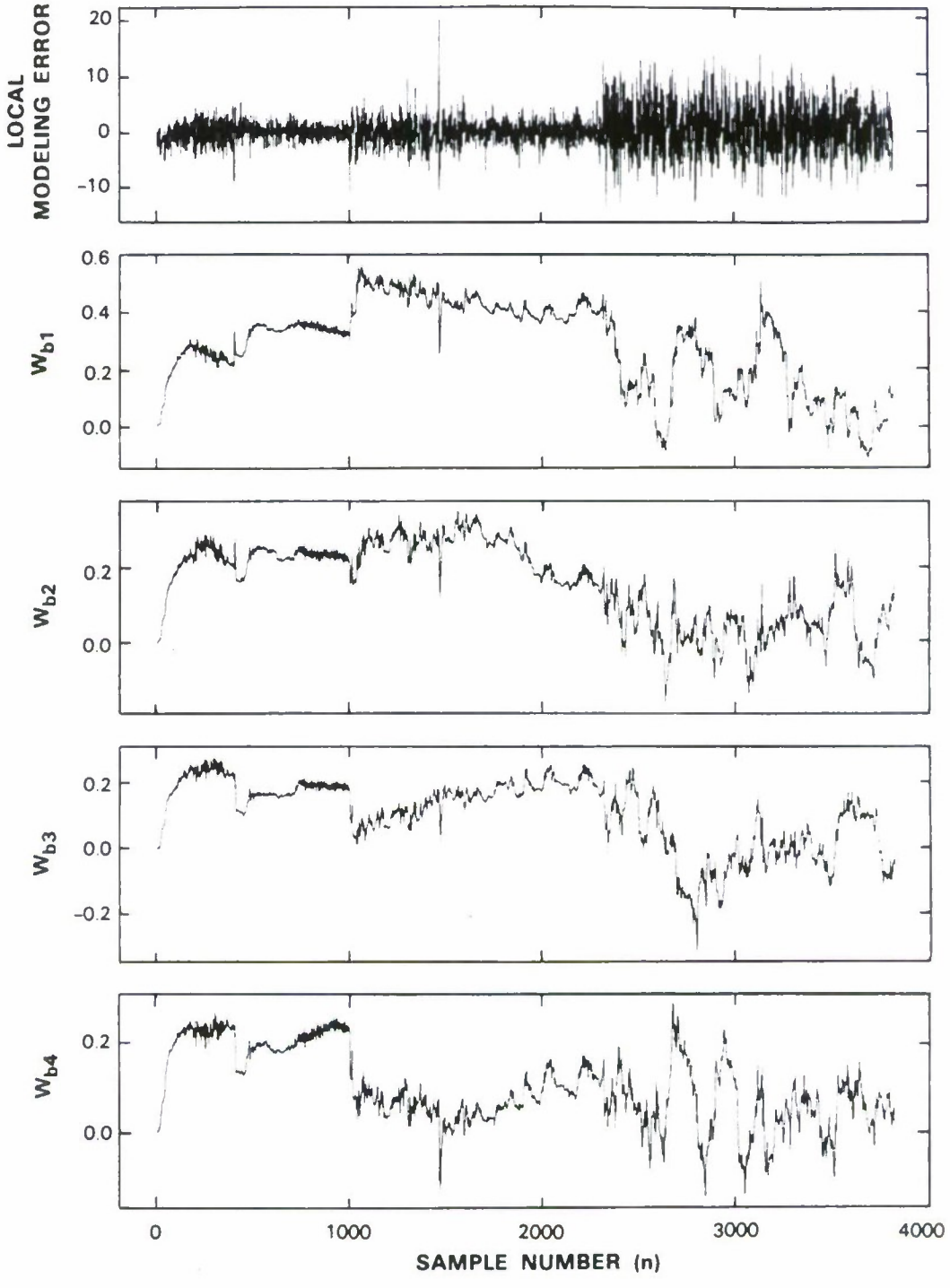


Figure 5-20. Normalized roll-stable global modeling results.



116884-7

Figure 5-21. Normalized roll-stable local modeling results.

Stable-Yaw: Finally, these results are shown in Figures 5-22 to 5-24. The transition occurs at n = 1502, detected near 1600, and again the location $\hat{T}_{\Delta}^{4}$ estimated within samples of the true transition.

Feature vectors representing a *stable* region and final *yaw* maneuver are calculated over the intervals $(\hat{T}_{\Delta}^{3} - \hat{T}_{\Delta}^{4})$ and $(\hat{T}_{\Delta}^{4} - N)$, respectively. Results are $\bar{V}^{T}(I_4) = \bar{0}$ and $\bar{V}^{T}(T_5) = (0.32, 0.19, 0.17, 0.21)$.

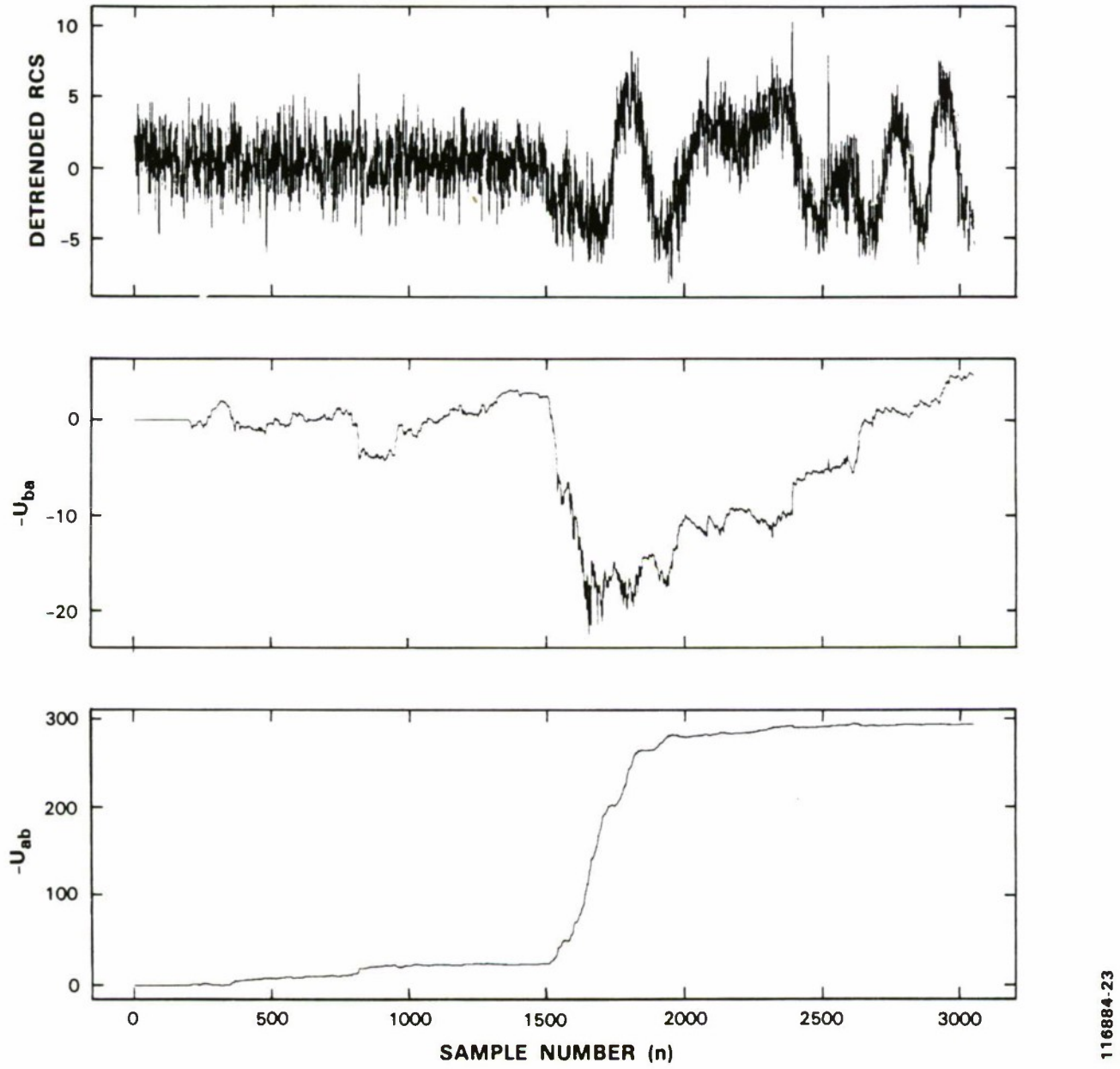
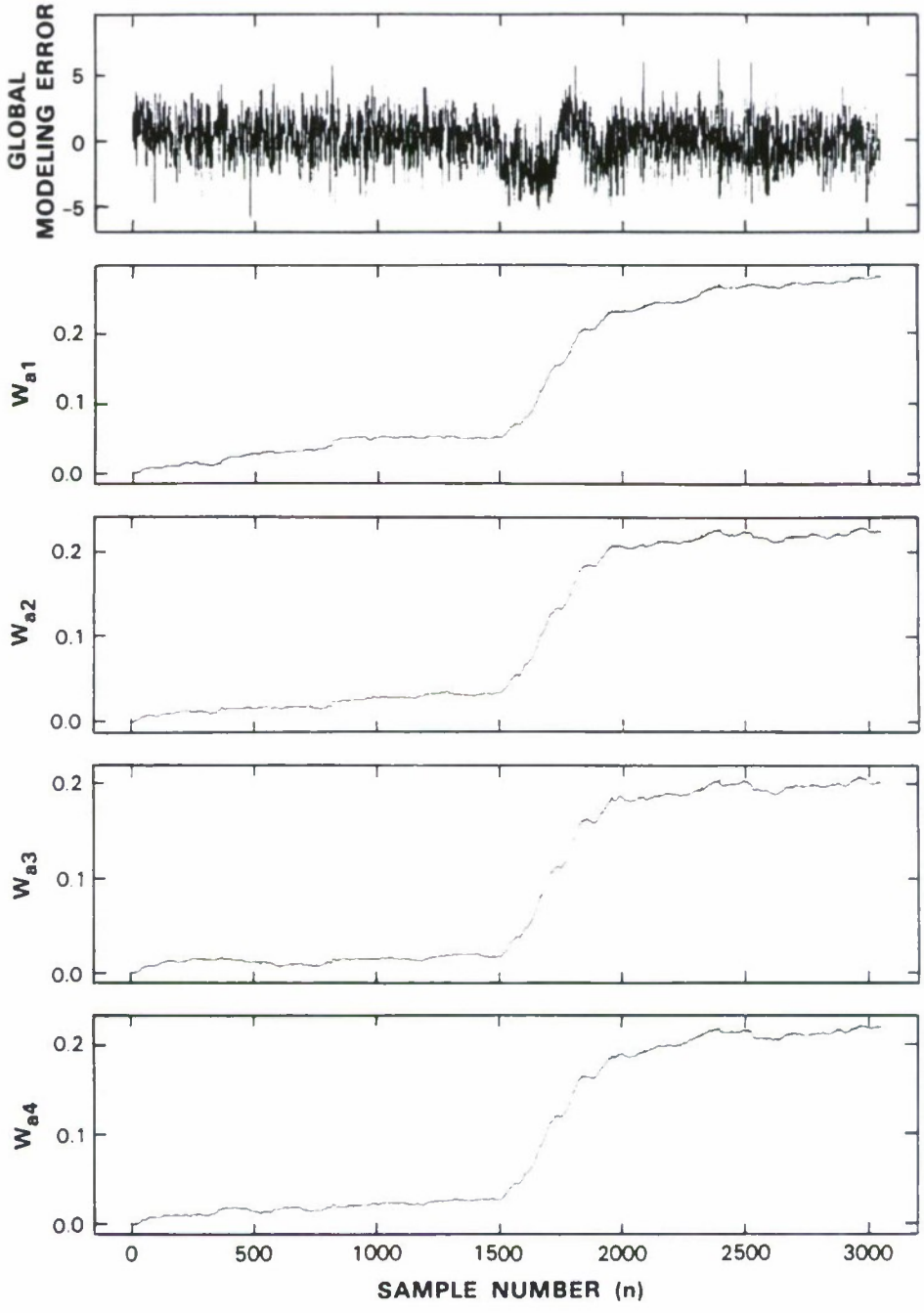
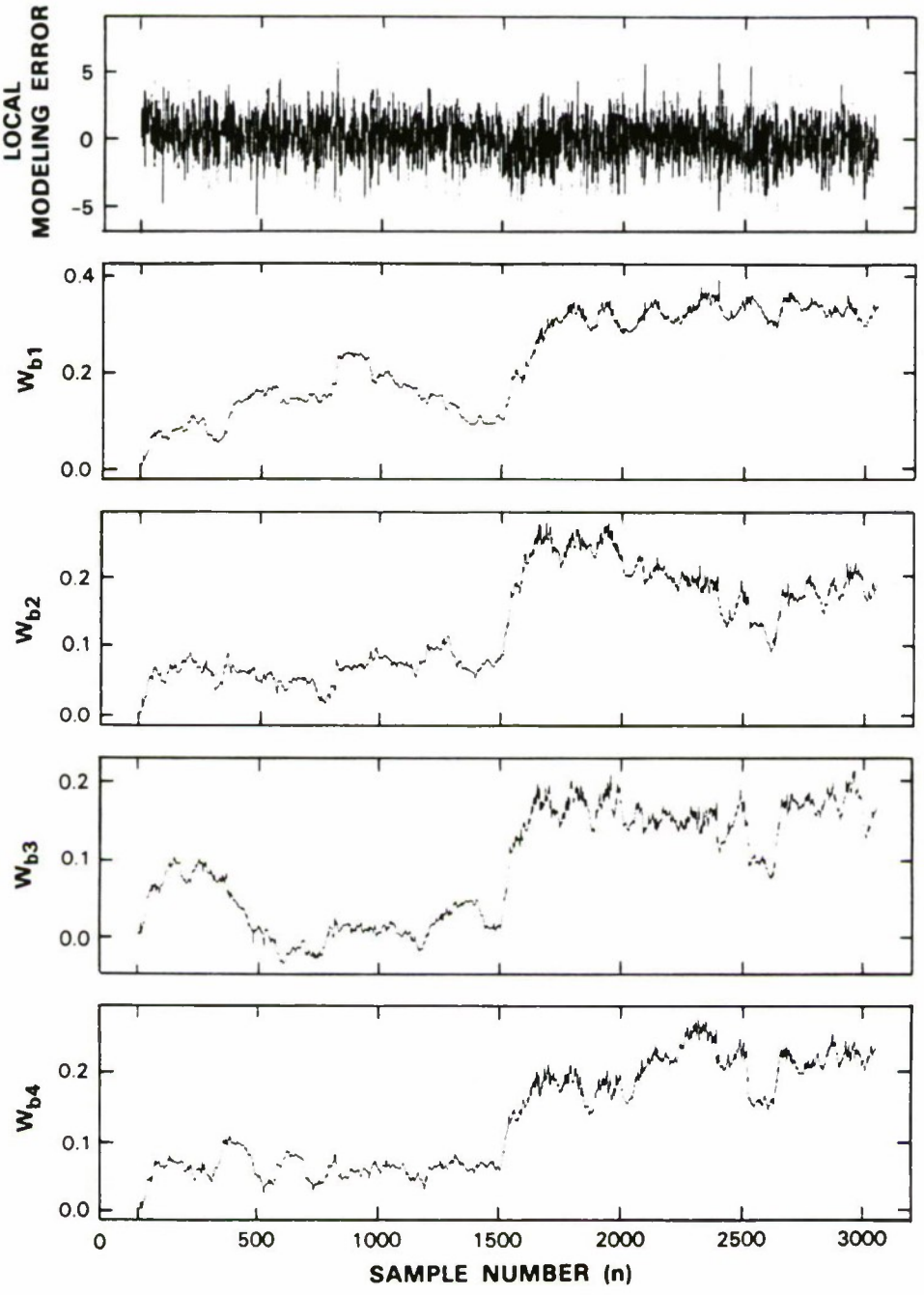


Figure 5-22. Stable-yaw transition detection results.



116884-24

Figure 5-23. Stable-yaw global modeling results.



116884-25

Figure 5-24. Stable-yaw local modeling results.

5.4 SUMMARY

Application of the preprocessing algorithms (Section 4), incorporating spectral transition detection and adaptive AR filtering (feature vector extraction), to nonstationary RCS data was reported. All filtering and detection parameters [including learning rates (η_a, η_b), filter orders (M_a, M_b) bias (δ), and a threshold (h)] were held constant throughout signal preprocessing. Therefore, the performance described is indicative of what can be achieved with a fixed set of parameters.

The data were detrended by removing the mean value and then selectively high-pass filtered. Results from detecting spectral transitions (physically representing object maneuvers) embedded in the detrended signal are summarized in Figure 5-25. All transitions were detected within samples of the actual locations. Detection performance was improved for signal transitions from high to low variance (*pitch-stable, roll-stable*) by normalizing the signal.

The overall performance of the adaptive preprocessor on real nonstationary data can be illustrated by Figure 5-26. The 8000-sample data were reduced to 7 disparate feature vectors, each of dimension 4. Moreover, the amount of information maintained in the feature vector representation can be gauged by the innovations (modeling error) sequence shown in Figure 5-26. With the exception of the *roll* maneuver, where the transition detector attempted to detect the *roll* submaneuvers, the sequence is near white, designating near-complete statistical information extraction.

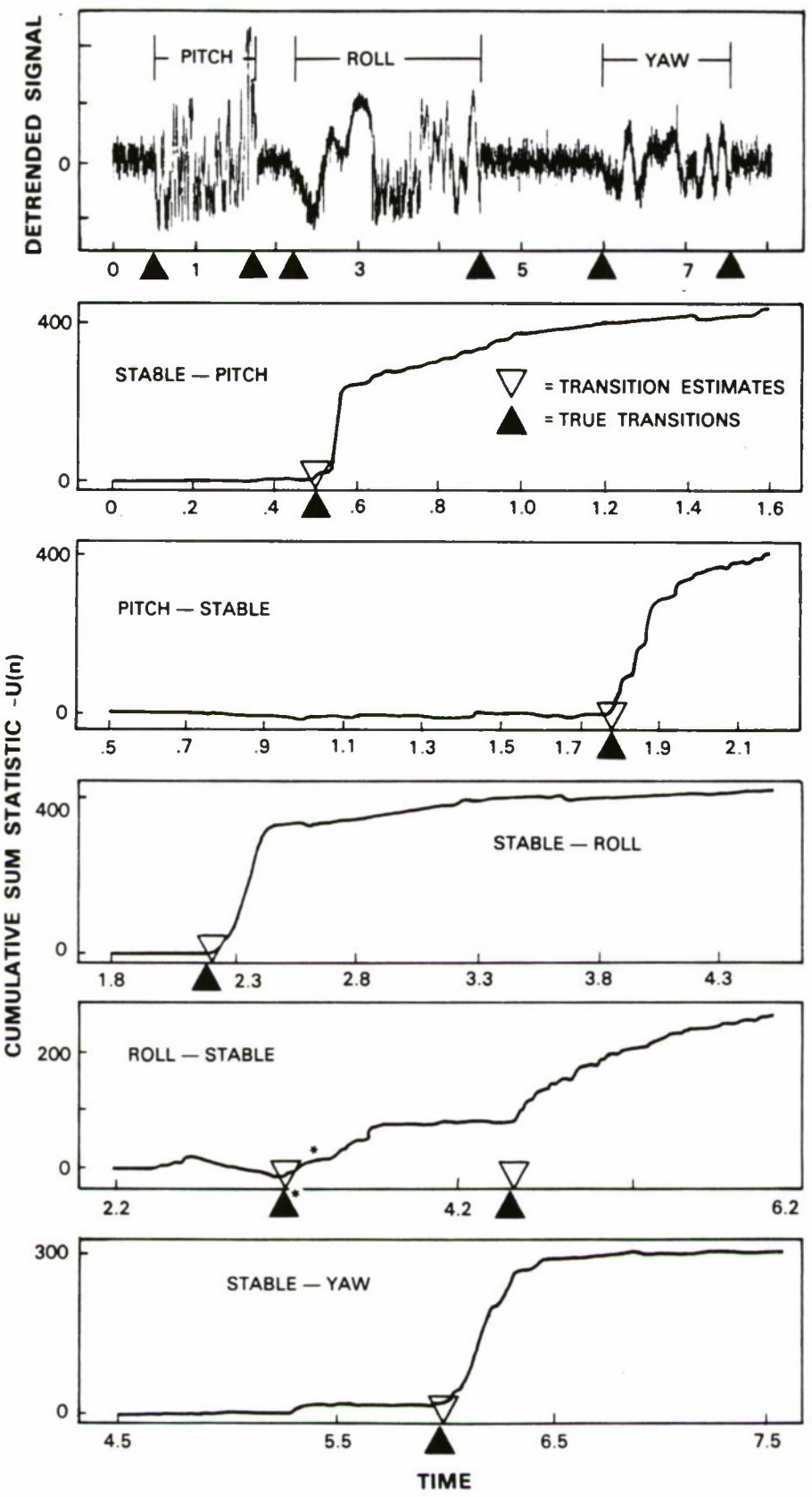


Figure 5-25. Summary of transition detection results.

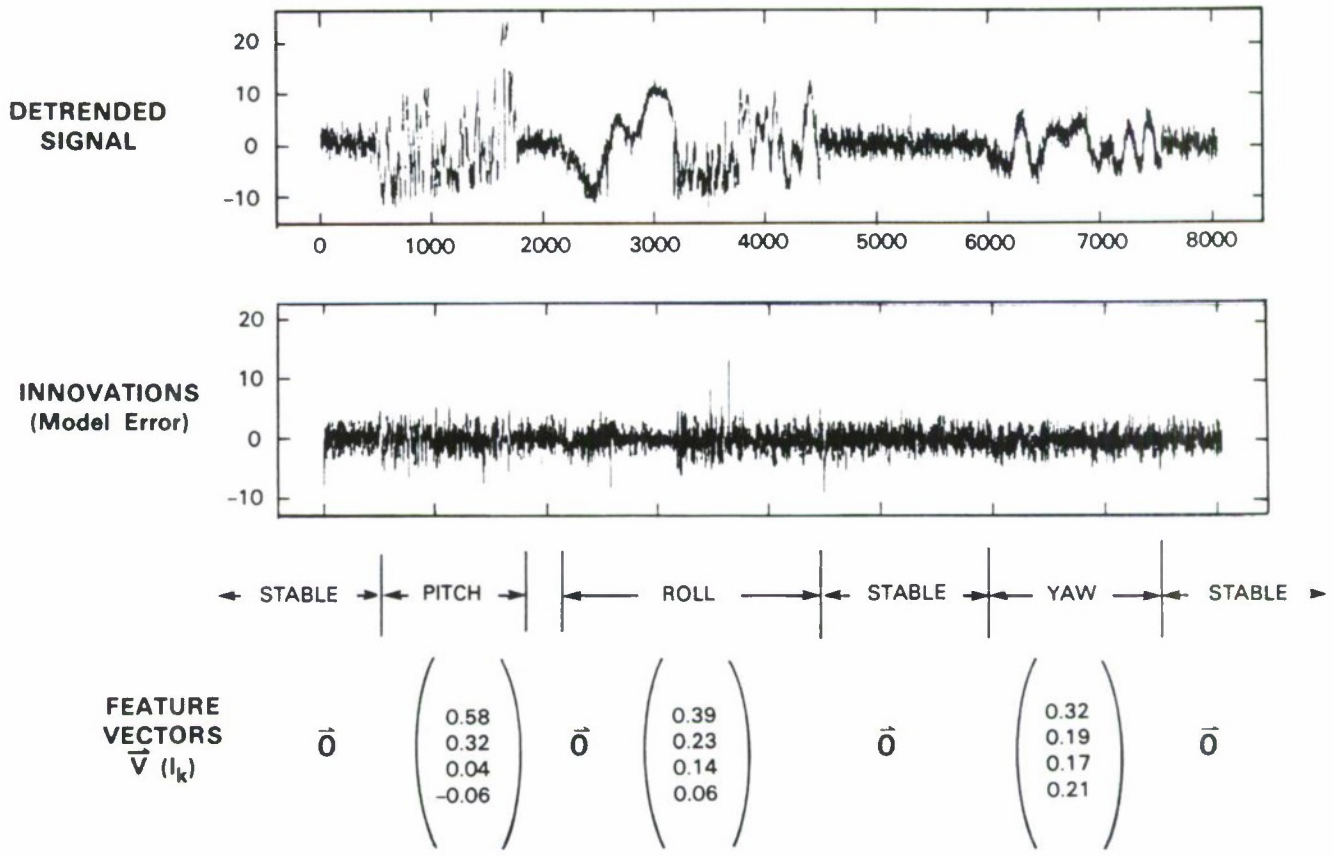


Figure 5-26. Adaptive preprocessor performance.

6. CONCLUSION

An adaptive preprocessor was introduced to provide a compact representation of nonstationary data. The representation, consisting of AR coefficient vectors computed over near-stationary segments, is scale and translation invariant, normalized, and statistically sufficient.

Guided by a decomposition theorem, the preprocessor was constructed using adaptive AR modeling filters and a transition detector. The information-theoretic transition detector driven by the parallel adaptive AR modeling filters successfully detected all transitions within the radar signature analyzed, thereby yielding segments of near-stationary data. Also, the feature vectors produced by the modeling filters over the detected near-stationary segments were sufficiently distinct, thus supporting automatic classification. Practically all the sensor statistical data information (throughout 8000 samples) was retained in the 7 compact feature vectors (each of dimension 4) as substantiated by the resulting near-white innovation process.

Regarding future research suggestions, both components of the preprocessor could be enhanced. For example, alternative recursive least-squares algorithms with faster convergence times might be employed for the adaptive AR modeling filter, currently utilizing the LMS adaption algorithm. Also, new transition detector statistics might be explored which yield symmetrical behavior, thereby avoiding the normalization requirement when encountering high- to low-variance transitions. Or alternatively, automatic gain control circuitry could possibly be employed to automatically normalize the data.

REFERENCES

1. H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, Sweden: Almquist and Wieksells (1938).
2. S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall (1986).
3. G. Yule, "On a method of investigating periodicities in distributed series, with special reference to Wolfer's sunspot numbers, "*Philos. Trans. R. Soc. London*, **A226**, 267-298 (1927).
4. G. Walker, "On periodicity in series of related terms," *Proc. R. Soc. London, Ser.* **A131**, 518-532 (1931).
5. N. Wiener, *Exploration, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, Cambridge, Mass: MIT Press (1942).
6. B. Widrow and M. Hoff, Jr., "Adaptive switching circuits," *IRE WESCON Conv. Rec.*, Part 4, 96-104 (1960).
7. J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, **58**, 963-993 (1979).
8. S. K. Jones, R. K. Cavin, III, and W. M. Reed, "Analysis of error-gradient adaptive linear equalizers for a class of stationary-dependent processes," *IEEE Trans. Inf. Theory*, **IT-28**, 318-329 (1982).
9. M. Basseville and A. Benveniste, "Sequential detection of abrupt changes in spectral characteristics of noise," *IEEE Trans. Inf. Theory*, **IT-29**, 709-724 (1983).
10. S. Kullback, *Information Theory and Statistics*, New York: Wiley (1959).
11. D. Hinkley, "Inference about the change-point from cumulative sum-tests, *Biometrika*, **58**, 509-523 (1971).
12. G. Box and G. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day: San Francisco, Calif. (1976).
13. H. Akaike, "A new look at statistical model identification," *IEEE Trans. Autom. Control*, **AC-19**, 716-722 (1974).
14. J. Rissanen, "Modeling by shortest data description," *Automatica*, **Vol. 14**, 465-471 (1978).
15. H. Inada, personal communication (1988).

APPENDIX A
AR AUTOCORRELATION

Assuming the process $y(n)$ is stationary, the autocorrelation function is

$$r(m) \equiv E\{y(n+m)\,y^*(n)\}, \tag{A.1}$$

recall

$$y(n) = \sum_{i=1}^{M} w_i^* y(n-i) + e(n) \tag{A.2}$$

so that

$$r(m) = E\left\{\left(\sum_{i=1}^{M} w_i^* y(n+m-i) + e(n+m)\right) y^*(n)\right\} \tag{A.3}$$

$$= \sum_{i=1}^{M} w_i^* r(m-i) + E\{e(n+m)\,y^*(n)\}$$

To evaluate the last term, consider the linear filter interpretation [i.e., $y(n)$ is the output of infinite impulse response (IIR) filter]

$$y(n) = h(n) * e(n) \tag{A.4}$$

$$= \sum_{k=0}^{\infty} h^*(k)\,e(n-k).$$

Hence,

$$r(m) = \sum_{i=1}^{M} w_i^* r(m-i) + E\left\{e(n+m)\sum_{k=0}^{\infty} h(k)\,e^*(n-k)\right\} \tag{A.5}$$

$$= \sum_{i=1}^{M} w_i^* r(m-i) + \sum_{k=0}^{\infty} h(k) E\{e(n+m)e^*(n-k)\},$$

and under the white noise residual assumption

$$E\{e(n+m)e^*(n-k)\} = \sigma_e^2 \delta_{m+k} = \begin{cases} \sigma_e^2 & k = -m \\ 0 & k \neq -m \end{cases}; \tag{A.6}$$

consequently,

$$r(m) = \sum_{i=1}^{M} w_i^* r(m-i) + \sum_{k=0}^{\infty} h(k)\sigma_e^2 \delta_{m+k}, \quad m = 0, 1, 2, \ldots \tag{A.7}$$

$$\therefore \quad r(m) = \begin{cases} \sum_{i=1}^{M} w_i^* r(-i) + h(0)\sigma_e^2 & m = 0 \\ \sum_{i=1}^{M} w_i^* r(m-i) & m > 0 \\ r^*(-m) & m < 0 \end{cases}.$$

Finally, substituting the lags $m = 1, 2, \ldots M$ into the autocorrelation expression yields the linear matrix equation for the AR coefficients

$$\begin{aligned} r(1) &= w_1^* r(0) + w_2^* r(-1) + \ldots + w_M^* r(1-M) \\ r(2) &= w_1^* r(1) + w_2^* r(0) + \ldots + w_M^* r(2-M) \\ &\vdots \\ r(M) &= w_1^* r(M-1) + w_2^* r(M-2) + \ldots + w_M^* r(0) \end{aligned}$$

$$\therefore \begin{pmatrix} r(0) & r(-1) & \ldots & r(1-M) \\ r(1) & r(0) & \ldots & r(2-M) \\ \vdots & & & \\ r(M-1) & r(M-2) & \ldots & r(0) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{pmatrix}^* = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(M) \end{pmatrix}. \tag{A.8}$$

Upon defining the correlation matrix

$$R \equiv E\left\{ \bar{y}(n-1)\, \bar{y}^{H}(n-1) \right\}$$

$$= E\left\{ \begin{pmatrix} y(n-1) \\ y(n-2) \\ \vdots \\ y(n-M) \end{pmatrix} \left(y(n-1)\ y(n-2) \ldots y(n-M) \right)^{*} \right\}$$

$$= \begin{pmatrix} r(0) & r(1) & r(M-1) \\ r(-1) & r(0) & r(M-2) \\ \vdots & & \\ r(-M+1) & r(-M+2) \cdots & r(0) \end{pmatrix} \tag{A.9}$$

and

$$\bar{r} \equiv \begin{pmatrix} r(-1) \\ r(-2) \\ \vdots \\ r(-M) \end{pmatrix} \tag{A.10}$$

the AR coefficients are seen to satisfy

$$\bar{R}^{T} \bar{w}^{*} = \bar{r}^{*} \quad or \quad \bar{R}^{H} \bar{w} = \bar{r} \tag{A.11}$$

$$\therefore \quad \bar{R}\bar{w} = \bar{r}$$

since the correlation matrix $\bar{R}$ is Hermitian.

APPENDIX B
CONDITIONAL DISTRIBUTION OF PARTITIONED GAUSSIAN RANDOM VECTOR

Let a Gaussian random vector be partitioned as

$$\bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \tag{B.1}$$

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_1} \end{pmatrix} \tag{B.2}$$

$$\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_1} \end{pmatrix} \tag{B.3}$$

having a joint distribution

$$p(\bar{z}) = \frac{1}{(2\pi)^{n/2} |\bar{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{z}-\bar{z}')^H \bar{\Sigma}^{-1} (\bar{z}-\bar{z}')} \tag{B.4}$$

and a partitioned covariance matrix given by

$$\bar{\Sigma} \equiv E\left\{(\bar{z} - \bar{z}')(\bar{z} - \bar{z}')^H\right\} = \begin{pmatrix} \bar{\Sigma}xx & \bar{\Sigma}xy \\ \bar{\Sigma}yx & \bar{\Sigma}yy \end{pmatrix} \tag{B.5}$$

where

$$\bar{z}' = E\{\bar{z}\}. \tag{B.6}$$

The aim is to express the conditional distribution $p(\bar{x}/\bar{y})$ in terms of the partitioned covariances $\bar{\Sigma}xx$, $\bar{\Sigma}xy$, $\bar{\Sigma}yx$, and $\bar{\Sigma}yy$. Such a task is easily accomplished if $\bar{\Sigma}$ can be massaged into matrix diagonal

form. That is, a transformation on $\bar{z}$ is sought to simplify the exponential argument in Equation (B.4) yielding

$$(\bar{z} - \bar{z}')^H \bar{\Sigma}^{-1}(\bar{z} - \bar{z}') = (\bar{s} - \bar{s}')^H \begin{pmatrix} \bar{C}_{11} & \bar{0} \\ \bar{0} & \bar{C}_{22} \end{pmatrix}^{-1} (\bar{s} - \bar{s}')$$

$$= (\bar{s} - \bar{s}')^H \begin{pmatrix} \bar{C}_{11}^{-1} & \bar{0} \\ \bar{0} & \bar{C}_{22}^{-1} \end{pmatrix} (\bar{s} - \bar{s}')$$

$$= (\bar{s}_1 - \bar{s}_1')^H \bar{C}_{11}^{-1}(s_1 - \bar{s}_1') + (\bar{s}_2 - \bar{s}_2')\bar{C}_{22}^{-1}(\bar{s}_2 - \bar{s}_2') \tag{B.7}$$

where $\bar{s}$ is the appropriately partitioned vector

$$\bar{s} = \begin{pmatrix} \bar{s}_1 \\ \bar{s}_2 \end{pmatrix}. \tag{B.8}$$

The following identity is used to provide the partitioned diagonalization. For $\bar{\Sigma}$ positive definite and $\bar{\Sigma}_{xx}$ square, let

$$\bar{B} = \begin{pmatrix} \bar{I} & -\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1} \\ \bar{0} & \bar{I} \end{pmatrix} \tag{B.9}$$

then

$$\bar{C} = \bar{B}\bar{\Sigma}\bar{B}^H = \begin{pmatrix} \bar{\Sigma}_{xx} - \bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\bar{\Sigma}_{yx} & \bar{0} \\ \bar{0} & \bar{\Sigma}_{yy} \end{pmatrix} = \begin{pmatrix} \bar{C}_{11} & \bar{0} \\ \bar{0} & \bar{C}_{22} \end{pmatrix} \tag{B.10}$$

and is easily verified upon substitution.

Now, $\bar{\Sigma}^{-1}$ can be computed in partitioned form

$$\bar{\Sigma}^{-1} = \left(\bar{B}^H \bar{C} \bar{B}\right)^{-1} = \bar{B}^{-1} \bar{C}^{-1} \left(\bar{B}^H\right)^{-1} = \bar{B}^H \bar{C}^{-1} \bar{B} = \bar{B}^H \begin{pmatrix} \bar{C}_{11}^{-1} & \bar{0} \\ \bar{0} & \bar{C}_{22}^{-1} \end{pmatrix} \bar{B}. \tag{B.11}$$

Also, the determinant can be computed by

$$|\bar{\Sigma}| = |\bar{B}^H \bar{C} \bar{B}| = |\bar{B}^H| |\bar{C}| |\bar{B}| = |\bar{C}_{11}| |\bar{C}_{22}|. \tag{B.12}$$

Consequently, using Equations (B.11) and (B.12), the joint distribution [Equation (B.4)] becomes

$$p(\bar{z}) = \frac{1}{(2\pi)^{n/2} |\bar{C}_{11}|^{\frac{1}{2}} |\bar{C}_{22}|^{\frac{1}{2}}} e^{-\frac{1}{2}(z - \bar{z}')\bar{B}^H \bar{C}^{-1} \bar{B} (z - \bar{z}')} \tag{B.13}$$

and is written compactly utilizing the coordinate transformation

$$\bar{s} = \bar{B}\bar{z} = \begin{pmatrix} \bar{x} - \bar{\Sigma}_{xy} \bar{\Sigma}_{yy} \bar{y} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \bar{s}_1 \\ \bar{s}_2 \end{pmatrix} \tag{B.14}$$

so that

$$p(\bar{s}) = \frac{1}{(2\pi)^{n/2} |\bar{C}_{11}|^{\frac{1}{2}} |\bar{C}_{22}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{s} - \bar{s}')^H \bar{C}^{-1} (\bar{s} - \bar{s}')} \tag{B.15}$$

$$= \frac{1}{(2\pi)^{n/2} |\bar{C}_{11}|^{\frac{1}{2}} |\bar{C}_{22}|^{\frac{1}{2}}} e^{-\frac{1}{2}\left[(\bar{s}_1 - \bar{s}'_1)^H \bar{C}_{11}^{-1} (\bar{s}_1 - \bar{s}'_1) + (\bar{s}_2 - \bar{s}'_2)^H \bar{C}_{22}^{-1} (\bar{s}_2 - \bar{s}'_2)\right]}$$

Finally, upon substituting Equation (B.10) into (B.16), the joint distribution in partitioned form becomes

$$p(\bar{x},\bar{y}) = \frac{1}{(2\pi)^{n/2}\left|\bar{\Sigma}_{xx} - \bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\bar{\Sigma}_{yx}\right|^{\frac{1}{2}}\left|\bar{\Sigma}_{yy}\right|^{\frac{1}{2}}} \tag{B.16}$$

$$\bullet\ e^{-\frac{1}{2}\left[(\bar{x}-\bar{x}')-\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}(\bar{y}-\bar{y}')\right]^{H}\left(\Sigma_{xx}-\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\bar{\Sigma}_{yx}\right)^{-1}\left((\bar{x}-\bar{x}')-\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}(\bar{y}-\bar{y}')\right)}$$

$$\bullet\ e^{-\frac{1}{2}\left[(\bar{y}-\bar{y}')^{H}\bar{\Sigma}_{yy}^{-1}(\bar{y}-\bar{y}')\right]}$$

while the conditional distribution in partitioned form becomes

$$p(\bar{x}/\bar{y}) = \frac{p(\bar{x},\bar{y})}{p(\bar{y})} = \frac{1}{(2\pi)^{n_x/2}\left|\bar{\Sigma}_{xx} - \bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\bar{\Sigma}_{yx}\right|^{\frac{1}{2}}} \tag{B.17}$$

$$\bullet\ e^{-\frac{1}{2}\left[\bar{x}-\left(\bar{x}'+\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}(\bar{y}-\bar{y}')\right)\right]^{H}\left(\Sigma_{xx}-\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\Sigma_{yx}\right)^{-1}\left[\bar{x}-\left(\bar{x}'+\bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}(\bar{y}-\bar{y}')\right)\right]}$$

APPENDIX C
DERIVATION OF TEST STATISTIC FOR AR GAUSSIAN PROCESS

The test statistic given by

$$u(n) = \sum_{k=1}^{n} T(k) \tag{C.1}$$

where

$$T(k) = \log \frac{p_a(y(k)/\bar{y}(k-1))}{p_b(y(k)/\bar{y}(k-1))} - \int p_a(y(k)/\bar{y}(k-1)) \log \frac{p_a(y(k)/\bar{y}(k-1))dy}{p_b(y(k)/\bar{y}(k-1))} \tag{C.2}$$

is derived for the zero mean AR Gaussian process $y(n)$

$$y(n) = \bar{w}^H \bar{y}(n-1) + e(n) \tag{C.3}$$

where $e(n)$ is a white Gaussian noise process with variance σ_e^2. First, the conditional distribution needed for deriving the test statistic Equation (C.2) is obtained by defining the $(M+1) \times 1$ partitioned random vector

$$\bar{z} \equiv \begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} = \begin{pmatrix} y(k) \\ \bar{y}(k-1) \end{pmatrix} \tag{C.4}$$

with appropriate partitioned covariance matrices

$$\bar{\Sigma}_{uu} \equiv E\left\{ y(k)\, y(k)^H \right\} = r(0) \tag{C.5}$$

$$\bar{\Sigma}_{vu} \equiv E\left\{ \bar{y}(k-1)\, y(k)^H \right\} = \bar{r} \tag{C.6}$$

$$\bar{\Sigma}_{uv} = \bar{\Sigma}_{vu}^H = \bar{r}^H \tag{C.7}$$

$$\bar{\Sigma}_{vv} \equiv E\left\{ \bar{y}(k-1)\, \bar{y}(k-1)^H \right\} = \bar{R}. \tag{C.8}$$

Next, the results of Appendix B are applied (substituting $\bar{u} = \bar{x}$, $\bar{v} = \bar{y}$) yielding

$$p(y(k)/\bar{y}(k-1)) = K' e^{-1/2\left[y(k)-\bar{r}^H \bar{R}^{-1} \bar{y}(k-1)\right]^H \left(r(0)-\bar{r}^H \bar{R}^{-1} \bar{r}\right)^{-1} \left[y(k)-\bar{r}^H \bar{R}^{-1} \bar{y}(k-1)\right]} \tag{C.9}$$

where

$$K' = \left(2\pi \left| r(0) - \bar{r}^H \bar{R}^{-1} \bar{r} \right|\right)^{-1/2}. \tag{C.10}$$

Employing the AR modeling assumption, (i.e., $\bar{R}\bar{w} = \bar{r}$ and $\sigma_e^2 = r(0) - \bar{w}^H \bar{r}$), the distribution reduces to

$$p(y(k)/\bar{y}(k-1)) = Ke^{-\frac{|e(k)|^2}{2\sigma_e^2}} \tag{C.11}$$

where

$$K = \left(2\pi\sigma_e^2\right)^{-\frac{1}{2}} \tag{C.12}$$

and

$$e(k) = y(k) - \bar{w}^H \bar{y}(k-1) \tag{C.13}$$

represents the AR modeling error (innovation) process [Equation (2.9)]. Therefore, the steady state test statistic [Equation (C.2)] for the AR Gaussian process [Equation (C.3)] is

$$T(k) = \log \frac{K_a}{K_b} - \frac{|e_a(k)|^2}{2\sigma_{e_a}^2} + \frac{|e_b(k)|^2}{2\sigma_{e_b}^2} \tag{C.14}$$

$$-\left(\log\left(\frac{K_a}{K_b}\right)-\frac{1}{2\sigma_{e_a}^2}\int_{-\infty}^{\infty}|e_a(y)|^2 p(e_a(y))dy+\frac{1}{2\sigma_{e_b}^2}\int_{-\infty}^{\infty}|e_b(y)|^2 p(e_a(y))dy\right)$$

where

$$e_a(y) = y - \bar{w}_a^H \bar{y}(k-1) \tag{C.15}$$

$$e_b(y) = y - \bar{w}_b^H \bar{y}(k-1) \tag{C.16}$$

$$p(e_a(y)) = N\left(0, \sigma_{e_a}^2\right). \tag{C.17}$$

Evaluating the first integral,

$$I_1 = \frac{1}{2\sigma_{e_a}^2}\int_{-\infty}^{\infty}|e_a(y)|^2 p(e_a(y))dy \; ; \tag{C.18}$$

with a change of variables

$$I_1 = \frac{1}{2\sigma_{e_a}^2}E\left\{|e_a|^2\right\}$$

$$= \frac{1}{2\sigma_{e_a}^2}\left(\sigma_{e_a}^2\right) = \frac{1}{2}. \tag{C.19}$$

Evaluating the second integral,

$$I_2 = \frac{1}{2\sigma_{e_b}^2}\int|e_b(y)|^2 p(e_a(y))dy \tag{C.20}$$

$$= \frac{1}{2\sigma_{e_b}^2}\int\Big(|e_b(y) - e_a(y)|^2 + e_a(y)[e_b(y) - e_a(y)]^*$$

$$+ \left[e_b(y) - e_d(y) \right] e_a^*(y) + \left| e_a(y) \right|^2 \Big) p(e_d(y)) dy$$

$$= \frac{1}{2\sigma_{e_b}^2} \Big\{ \left| e_b(k) - e_a(k) \right|^2 + \left(e_b(k) - e_a(k) \right)^* E\{e_a\}$$

$$+ \left(e_b(k) - e_a(k) \right) E^*\{e_a\} + \sigma_{e_a}^2 \Big\}$$

$$= \frac{1}{2\sigma_{e_b}^2} \left\{ \left| e_b(k) - e_a(k) \right|^2 + \sigma_{e_a}^2 \right\}$$

Combining the results, Equation (C.14) becomes

$$T(k) = \frac{1}{2} \left[1 - \frac{\sigma_{e_a}^2}{\sigma_{e_b}^2} + \frac{\left| e_b(k) \right|^2}{\sigma_{e_b}^2} - \frac{\left| e_a(k) \right|^2}{\sigma_{e_a}^2} - \frac{\left| e_b(k) - e_a(k) \right|^2}{\sigma_{e_b}^2} \right]. \qquad \text{(C.21)}$$

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report 849	5. MONITORING ORGANIZATION REPORT NUMBER(S) ESD-TR-88-329
6a. NAME OF PERFORMING ORGANIZATION Lincoln Laboratory, MIT 6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION Electronic Systems Division
6c. ADDRESS *(City, State, and Zip Code)* P.O. Box 73 Lexington, MA 02173-0073	7b. ADDRESS *(City, State, and Zip Code)* Hanscom AFB, MA 01731
8a. NAME OF FUNDING/SPONSORING ORGANIZATION HQ AF Systems Command 8b. OFFICE SYMBOL *(If applicable)* FTD/SQDRA	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-85-C-0002
8c. ADDRESS *(City, State, and Zip Code)* Andrews AFB Washington, DC 20334-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 12424F, 31310F, 63223C PROJECT NO.: 80 TASK NO.: WORK UNIT ACCESSION NO.:
11. TITLE *(Include Security Classification)* Adaptive Preprocessing of Nonstationary Signals	
12. PERSONAL AUTHOR(S) Mitchell D. Eggers and Timothy S. Khoun	
13a. TYPE OF REPORT Technical Report 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT *(Year, Month, Day)* 1989, May 9 15. PAGE COUNT 86
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

adaptive filters; autoregressive modeling; time series analysis; transition detection

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

As the number and bandwidth of sensors increases, an acute demand for preprocessing sensor data obtained for machine-based decision making arises. Especially in a data fusion context, the data from numerous sensors must first be preprocessed to prevent saturation of the decision making mechanism — albeit man or machine.

Presented is a general preprocessing approach which provides a compact representation (feature vector) of sensor data. The approach, supported by a signal decomposition theorem, adaptively models in recursive fashion, the detrended sensor data as an autoregressive (AR) process of sufficiently high order. Provisions are included to accommodate nonstationary data hy incorporating an information-theoretic transition detector to identify the segments of near-stationary data. Together, feature vectors (AR coefficients) are produced over near-stationary data segments which are scale invariant, translation invariant, normalized, and represent sufficient statistics. Furthermore, the merit of the preprocessor is quantitatively determined in a continuous manner from the resulting innovations (modeling error process).

Specific application results utilizing nonstationary radar data demonstrate the ability to simultaneously reduce data and maintain information content, without requiring a priori statistics and/or expert rules.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Lt. Col. Hugh L. Southall, USAF	22b. TELEPHONE *(Include Area Code)* (617) 981-2330	22c. OFFICE SYMBOL ESD/TML

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.