

# **USC-SIPI REPORT #272**

## **Advanced Signal Processing for Hydroacoustic Data of Internal Launchway Flows**

**by**

**Dae C. Shin, George A. Tsihrintzis  
and Chrysostomos L. Nikias**

**October 1994**

**Signal and Image Processing Institute  
UNIVERSITY OF SOUTHERN CALIFORNIA  
Department of Electrical Engineering-Systems  
3740 McClintock Avenue, Room 404  
Los Angeles, CA 90089-2564 U.S.A.**

# Advanced Signal Processing for Hydroacoustic Data of Internal Launchway Flows\*

Dae C. Shin, George A. Tsihrintzis, and Chrysostomos L. Nikias

Signal & Image Processing Institute  
Department of Electrical Engineering - Systems  
University of Southern California  
Los Angeles, CA 90089-2564

## Abstract

In this report, we present the results from a number of signal processing approaches to analyzing hydroacoustic data measured in test internal launchway flows. The ultimate goal is to establish features of the data that allow their reliable classification into one of two categories, namely hydrophone data from a smooth section of the flow or from the section of the flow near a slot, as these are described in the report. We applied classification algorithms based on a variety of signal processing concepts, including modeling via second-order autoregressive equations, power spectrum- and polyspectra-based features, cepstral feature extraction based on third-order statistics, and modeling with symmetric, alpha-stable (S $\alpha$ S) distributions. In our analysis, we examined the transient flows (first half of each data set) separately from the steady-state flows (second half of each data set) and for each one we established reliable classification features. In particular, we found that the coefficients of a second-order autoregressive model as estimated via Gaussian maximum likelihood methods are features providing clear classification power between smooth and slot section section flow data. However, part of the classification power is lost if the autoregressive model estimates are obtained via Yule-Walker equations or their generalizations. It is noteworthy, that improvement was observed in the Yule-Walker approach when fractional, lower-order moments were employed. The power spectrum was another feature that allowed data classification, as was the phase of the bispectrum. In particular, it was observed that wide main lobes in the power spectrum or significant low frequencies in the phase of the bispectrum corresponded to flow data from the slot section, while narrow main lobes in the power spectrum or significant high frequencies in the phase of the bispectrum corresponded to flow data from the smooth section. On the other hand, the magnitude of the bispectrum did not provide clear classification power. Similarly, the cepstral parameters did not provide clear classification power either. As a conclusion, high classification power was provided by the Gaussian maximum likelihood estimates of the parameters of a second-order autoregressive model and the phase of the data bispectrum.

---

\*This work was supported by the Office of Naval Research under the ONR/SIPI Cooperative Research Program, contract N00014-91-J-4021.

# 1 Introduction

16 June 1994

MEMORANDUM

TO: George Tsihrintzis, USC / Max Nikias, USC  
FROM: J. Hillenbrand, Code 8322, Naval Undersea Warfare Center,  
Division Newport, Newport, RI

SUBJ: Slot Flow Noise Experiment

Scott Hassan (Code 8322) recently finished conducting slot flow noise experiments in the NUWC Transient Flow Loop Facility (see attached figure). Two test sections were used in the experiment: 1) a smooth test section smoothly transitioning from a 12 inch diameter pipe to a 6 inch diameter pipe and 2) a slot test section. The test sections were located downstream of the rectangular test section of the Flow Loop Facility. Schematics of both test sections are shown in the attached figures.

The flow is computer controlled. The computer controls the operation of a control valve located above the control room. Flow rate is measured by an electromagnetic flow meter located in the control room. Steady state flow conditions are obtained by operating the pump at some percentage of full pump speed and with the control valve held fully open. Two types of transient flow conditions are obtained by 1) operating the pump at full speed with the valve initially held closed; the flow rate is controlled through the transient valve positions and 2) holding the valve fully open; the flow rate is controlled by transient pump speeds. Transient data includes acceleration, acoustic and flow measurements.

Data sets for the two types of transient accelerations are available; valve controlled and motor controlled. The following lists the files that will be sent (all measurements were taken near the beginning of the 6 inch test section as noted in the figures):

**Smooth Test Section**

Binary Data Files, 3 ensembles each (total of 18 files), fs=32768  
1/sec

1) Valve controlled

- velocity: TRVVEL1.BIN, TRVVEL2.BIN, TRVVEL3.BIN
- hydrophone: TRVHYD1.BIN, TRVHYD2.BIN, TRVHYD3.BIN
- acceleration: TRVACC1.BIN, TRVACC2.BIN, TRVACC3.BIN

2) Motor controlled

- velocity: TRMVEL1.BIN, TRMVEL2.BIN, TRMVEL3.BIN
- hydrophone: TRMHYD1.BIN, TRMHYD2.BIN, TRMHYD3.BIN
- acceleration: TRMACC1.BIN, TRMACC2.BIN, TRMACC3.BIN



## Slot Test Section

Binary Data files, 3 ensembles each (total of 18 files), fs=32768  
1/sec

- 1) Valve controlled
  - velocity: TRSVVEL1.BIN, TRSVVEL2.BIN, TRSVVEL3.BIN
  - hydrophone: TRSVHYD1.BIN, TRSVHYD2.BIN, TRSVHYD3.BIN
  - acceleration: TRSVACC1.BIN, TRSVACC2.BIN, TRSVACC3.BIN
- 2) Motor controlled
  - velocity: TRSMVEL1.BIN, TRSMVEL2.BIN, TRSMVEL3.BIN
  - hydrophone: TRSMHYD1.BIN, TRSMHYD2.BIN, TRSMHYD3.BIN
  - acceleration: TRSMACC1.BIN, TRSMACC2.BIN, TRSMACC3.BIN

The binary data files were created to be read into MATLAB. The files were written by a PC (IEEE Little Endian format) using 32 bit floating point values.

The following are the first 10 numbers of TRVVEL1.BIN

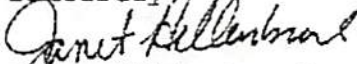
0.70246708393097  
0.82756388187408  
0.94303798675537  
0.91416949033737  
0.89492380619049  
0.78907257318497  
0.74095839262009  
0.88530087471008  
0.96228367090225  
0.92379230260849

Let me know when you have a computer account set up; then I will transfer them using FTP.

Thank you for your interest in analyzing the data. We would be interested in seeing any results that you get. I will be going on vacation June 27 through September 6, 1994. If you would like to obtain any more information or discuss the experiment during this time period; please contact the following people:

Scott Hassan, Code 8322 (401) 841-3131 x152  
email: Hassan@Code83.nuwc.npt.navy.mil  
or  
Jim Gutkowski, Code 8322 (401) 841-7210 x122

Sincerely,



Janet Hillenbrand  
(401) 841-7210 x162  
email: hillenbrand@Code83.nuwc.npt.navy.mil

cc: Dr. Kam Ng ONR 334

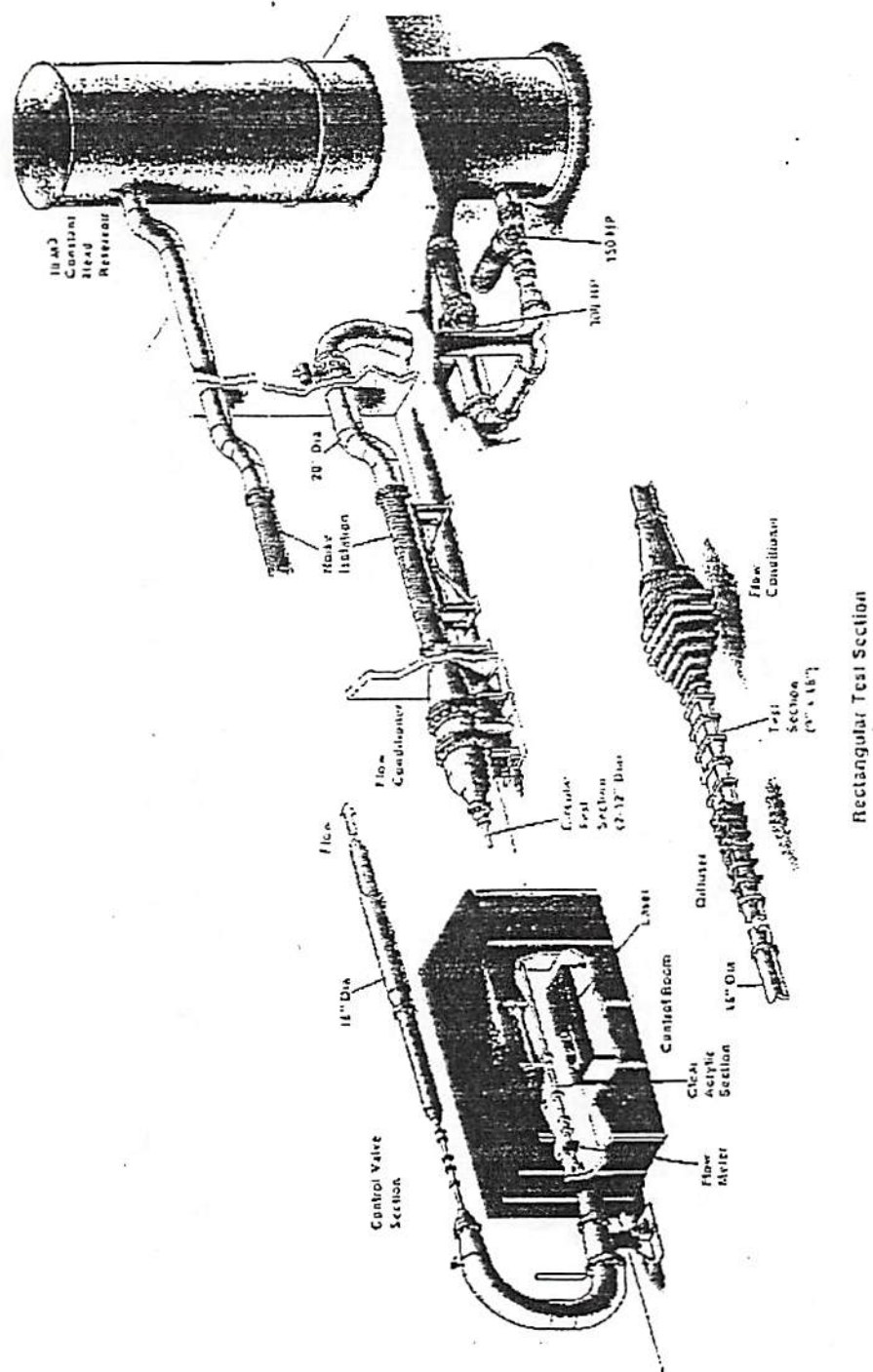
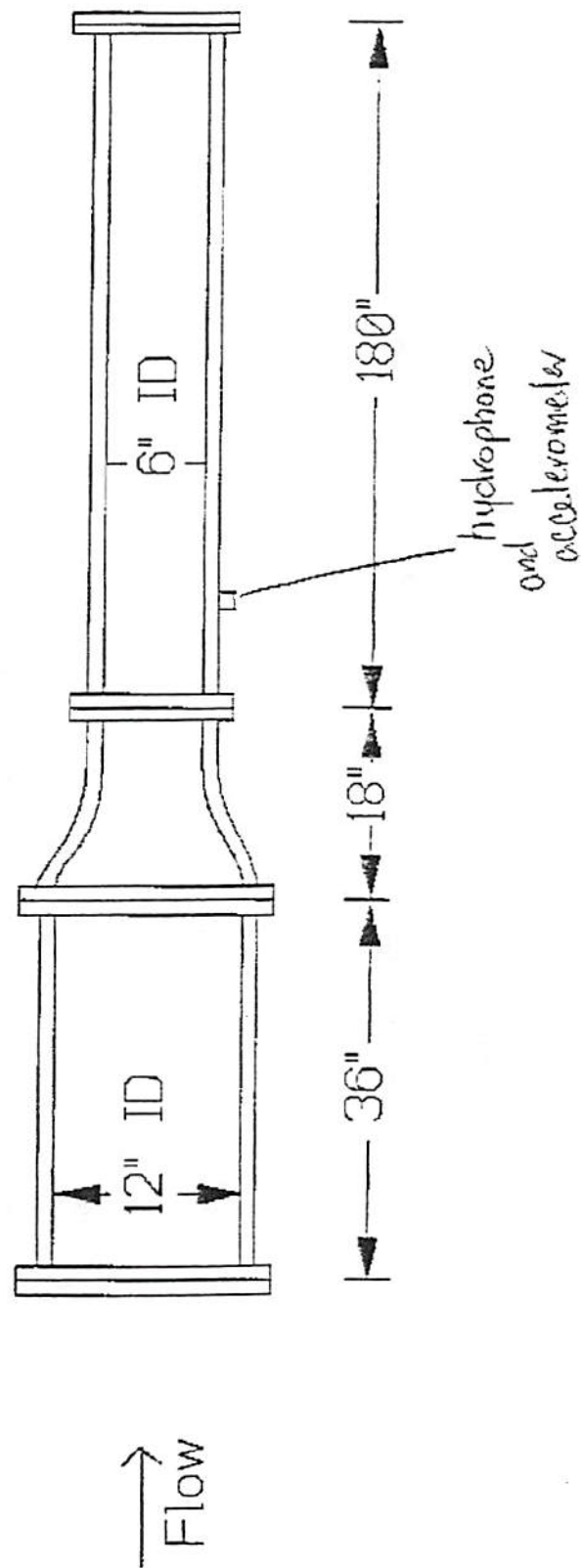


Figure 1. NUWC Transient Flow Loop Facility



Smooth Test Section

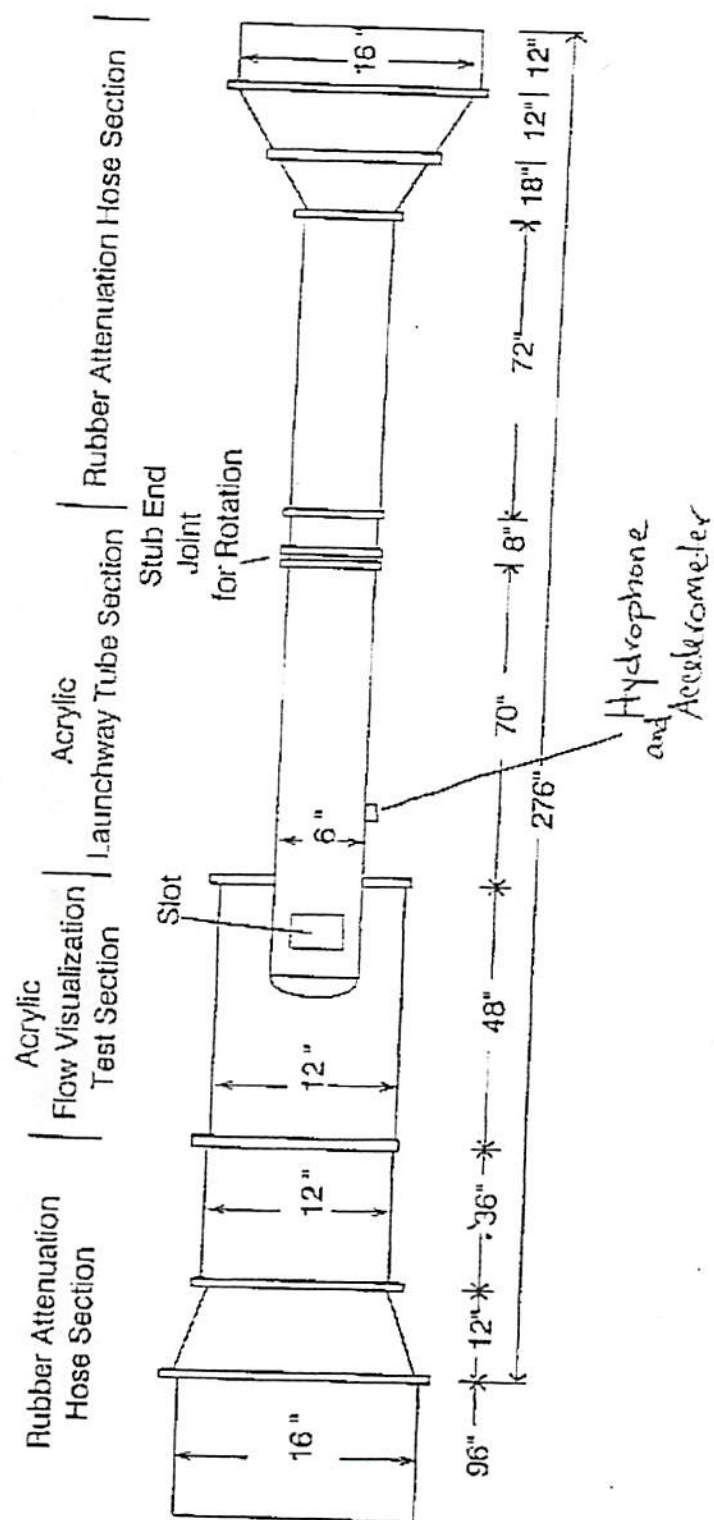


Figure 2. Slot Test Section



## 2 Description of Methods for Analyzing the Real Data Sets

In this section, we briefly describe the various methods used to classify the given test data (smooth and slot tests). Note that we let  $\{y(k)\}$  denote the given data sequences.

### 2.1 Second-Order AR Models

Assuming that we are given a set of measurement data  $\{y(k), k = 1, 2, \dots, N\}$ , we want to model it with an autoregressive (AR) model of order two:

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) + w(k) \quad (1)$$

where  $\{a_m, m = 1, 2\}$  are the AR parameters and  $\{w(k)\}$  is an i.i.d. random input process. Note that  $N \gg 2$ .

#### 2.1.1 AR Coefficient Estimation Using Gaussian ML methods

We can rewrite Eq. (1) as

$$w(k) = \sum_{m=0}^2 a_m y(k-m) \quad k = 3, 4, \dots, N \quad (2)$$

where  $a_0 = 1$ . When the order of an AR model is  $p$  and the distribution of the input sequence  $\{w(k)\}$  is  $f_w(w)$ , the likelihood function for estimation of the AR coefficients becomes

$$\begin{aligned} l(\mathbf{a}) &= f_w(w(p+1) | \mathbf{a}) f_w(w(p+2) | \mathbf{a}) \cdots f_w(w(N) | \mathbf{a}) \\ &= f_w\left(\sum_{m=0}^p a_m y(p+1-m)\right) f_w\left(\sum_{m=0}^p a_m y(p+2-m)\right) \cdots f_w\left(\sum_{m=0}^p a_m y(N-m)\right). \end{aligned} \quad (3)$$

Equivalently, the log-likelihood function becomes

$$L(\mathbf{a}) = \sum_{k=p+1}^N \ln f_w(w(k) | \mathbf{a}) = \sum_{k=p+1}^N \ln f\left(\sum_{m=0}^p a_m y(k-m)\right) \quad (4)$$

where  $\mathbf{a} = [1, a_1, a_2, \dots, a_p]^T$ . The ML estimate  $\hat{\mathbf{a}}$  of the parameters  $\mathbf{a}$  is the corresponding values at which the likelihood or the log-likelihood function reaches its maximum [Anderson, 1971; Brockwell and Davis, 1987].

Assuming that the input is a zero-mean, white Gaussian process with distribution

$$f_w(w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{w^2}{2\sigma^2}\right\} \quad (5)$$

where  $\sigma^2$  denotes the variance of the random process  $\{w(k)\}$ , we estimate the AR coefficients and the variance of the input random process using the Gaussian likelihood. Note that  $\sigma^2$  is independent of the AR parameters. Then, the log-likelihood function can be expressed as

$$\begin{aligned} L_g(\mathbf{a}, \sigma^2) &= -\sum_{k=3}^N \ln \sqrt{2\pi\sigma^2} - \sum_{k=3}^N \frac{\left(\sum_{m=0}^2 a_m y(k-m)\right)^2}{2\sigma^2} \\ &= -(N-2) \ln \sqrt{2\pi\sigma^2} - \sum_{k=3}^N \frac{(\mathbf{a}^T \mathbf{y}(k))^2}{2\sigma^2} \end{aligned} \quad (6)$$

where

$$\mathbf{a} = [1, a_1, a_2]^T \quad \text{and} \quad \mathbf{y}(k) = [y(k), y(k-1), y(k-2)]^T. \quad (7)$$

The ML estimate of the variance  $\sigma^2$  satisfies

$$\frac{\partial L_g(\mathbf{a}, \sigma^2)}{\partial \sigma^2} = -\frac{N-2}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{k=3}^N (\mathbf{a}^T \mathbf{y}(k))^2 = 0$$

and, thus, becomes

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{k=3}^N (\mathbf{a}^T \mathbf{y}(k))^2. \quad (8)$$

The ML estimate of the AR parameter vector  $\mathbf{a}$  satisfies

$$\frac{\partial L_g(\mathbf{a}, \sigma^2)}{\partial \mathbf{a}} = -\frac{\sum_{k=p+1}^N \mathbf{y}(k) \mathbf{y}(k)^T \hat{\mathbf{a}}}{2\sigma^2} = \mathbf{0}. \quad (9)$$

We can rewrite Eq. (9) in a vector-matrix form as

$$\mathbf{M}_y \cdot \hat{\mathbf{a}} = -\mathbf{V}_y \quad (10)$$

where  $\mathbf{M}_y$  and  $\mathbf{V}_y$  are the  $3 \times 2$  matrix and the  $3 \times 1$  vector, respectively, given by

$$\mathbf{M}_y = \begin{bmatrix} \sum_{k=3}^N y(k)y(k-1) & \sum_{k=3}^N y(k)y(k-2) \\ \sum_{k=p+1}^N y^2(k-1) & \sum_{k=3}^N y(k-1)y(k-2) \\ \sum_{k=p+1}^N y(k-1)y(k-2) & \sum_{k=3}^N y^2(k-2) \end{bmatrix} \quad \mathbf{V}_y = \begin{bmatrix} \sum_{k=3}^N y^2(k) \\ \sum_{k=3}^N y(k-1)y(k) \\ \sum_{k=3}^N y(k-2)y(k) \end{bmatrix}. \quad (11)$$

Let  $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2]^T$  be the least-squares solution to the overdetermined system of Eq. (10):

$$\hat{\mathbf{a}} = -\left(\mathbf{M}_y^T \mathbf{M}_y\right)^{-1} \mathbf{M}_y^T \mathbf{V}_y^T. \quad (12)$$

The lower rank approximation via singular value decomposition (SVD) may also be used in Eq. (12).

### 2.1.2 AR Coefficient Estimation Using the Normal Equations

An alternative method to estimate the AR coefficients (order  $p$ ) is to use the least-squares solution to the “normal” equations given by

$$\sum_{m=0}^p a_m c_2(\ell - m) = 0, \quad \ell > 0 \quad (13)$$

for a second-order statistics-based method, where  $c_2(\cdot)$  is the autocorrelation function of the measurement; and

$$\sum_{m=0}^p a_m c_3(\ell - m, \rho) = 0, \quad \ell > 0 \quad (14)$$

for a third-order statistics-based method, where  $c_3(\cdot, \cdot)$  is the third-order cumulant function of the measurement. Note that the AR identifiability condition requires that  $\rho = q_0 - p, \dots, q_0$ , for any  $q_0$  and  $\ell = i + 1, \dots, i + p$ , where  $i \geq 0$ .

### 2.1.3 Generalized Yule-Walker Equations

When it is assumed that the input sequence  $\{w(k)\}$  is i.i.d. symmetric  $\alpha$ -stable (S $\alpha$ S) random variables of characteristic exponent  $\alpha$  and dispersion  $\gamma$ , the stationary process  $\{y(k)\}$  in Eq. (1) is S $\alpha$ S [Shao and Nikias, 1993]. Note that  $\{y(k)\}$  and  $\{w(k + j)\}$  are independent for any  $j > 0$ . Taking the conditional expectation of both sides of (1) given  $\{y(m)\}$  result in, for  $k - 2 \leq m \leq k - 1$ ,

$$\mathbf{E}\{y(k)|y(m)\} = -a_1 \mathbf{E}\{y(k - 1)|y(m)\} - a_2 \mathbf{E}\{y(k - 2)|y(m)\} \quad (15)$$

since  $\mathbf{E}\{w(k)|y(m)\} = 0$  if  $\{w(k)\}$  and  $\{y(m)\}$  are independent. Since  $\{x(k)\}$  is stable and stationary, we have

$$\mathbf{E}\{y(k + \ell)|y(k)\} = \lambda(\ell) y(k) \quad (16)$$

where  $\lambda(0) = 1$  and  $\lambda(\ell)$  is the covariation coefficient of  $y(k + \ell)$  with  $y(k)$  defined by

$$\lambda(\ell) = \frac{\mathbf{E}\{y(k + \ell) y(k)^{<p-1>}\}}{\mathbf{E}\{|y(k)|^p\}} \quad (17)$$

for any  $1 \leq p < \alpha$ . In Eq. (17), the operator  $\langle \cdot \rangle$  is defined as

$$y^{\langle p \rangle} = |y| \cdot \text{sign}(y) \quad (18)$$

where  $\text{sign}(\cdot)$  is the signum function. We can therefore, obtain the following system of linear equations:

$$\mathbf{C}_\lambda \cdot \mathbf{a} = -\mathbf{p}_\lambda \quad (19)$$

where

$$\mathbf{C}_\lambda = \begin{bmatrix} \lambda(0) & \lambda(-1) \\ \lambda(1) & \lambda(0) \end{bmatrix}, \quad \mathbf{p}_\lambda = \begin{bmatrix} \lambda(1) \\ \lambda(2) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (20)$$

It is called the generalized Yule-Walker equation that is a direct generalization of the Gaussian Yule-Walker equation. The covariation coefficients can be estimated by several methods: fractional lower-order moment (FLOM), screened ratio (SCR), or least-squares (LS) estimators. However, it has been found that the generalized Yule-Walker equations using FLOM or LS estimates result in consistent estimates for the AR coefficients [Nikias and Shao, 1993; 1995] when the output sequences are not independent.

## 2.2 Power Spectrum and Polyspectrum Estimation

Spectral analysis is a signal processing method that characterizes the frequency components of a measured signal [Marple, 1987]. The power spectrum is defined as the Fourier transform of auto-correlation function and the bispectrum is defined as the Fourier transform of third-order cumulant function. The Fourier transforms of higher-order cumulant functions are called polyspectra.

### 2.2.1 Power Spectrum Estimation

Among a lot of methods for estimating the power spectrum of a process, we use the conventional Blackman-Tukey method described by

$$S(f) = F\{c_2(m) * W(m)\} \quad (21)$$

where  $S(f)$  denotes the power spectrum estimation,  $W(m)$  is a window function, and  $\{c_2(m)\}$  is a biased estimate of the auto-correlation function obtained by

$$c_2(m) = \frac{1}{N} \sum_k y(k) y(k+m). \quad (22)$$



### 2.2.2 Bispectrum Estimation

The bispectrum of a process is defined as the Fourier transform of its third-order cumulant (moment) function when the process is non-Gaussian and has a non-symmetric distribution. For bispectrum estimation, one of the direct and indirect estimators described in [Nikias and Petropulu, 1993] can be used. Here we describe the direct bispectrum estimator. When the sampling frequency is  $f_s$  and the required spacing between frequency samples in the bispectrum domain is  $\Delta = f_s/N_0$ , with  $N_0$  is the total number of frequency samples, the direct bispectrum estimator is given by

$$B_y(f_1, f_2) = \frac{1}{\Delta^2} Y(f_1) Y(f_2) Y^*(f_1 + f_2) \quad (23)$$

where  $\{Y(f_i), i = 1, 2\}$  is the Fourier transform of the process. If we have independent measurements of the process or use data segmentation technique, after estimating bispectrum for each measurement (segment) the final bispectrum is estimated by averaging:

$$B_y(f_1, f_2) = \frac{1}{\Delta^2} \sum_{\ell=1}^L Y^{(\ell)}(f_1) Y^{(\ell)}(f_2) Y^{(\ell)*}(f_1 + f_2) \quad (24)$$

where  $L$  is the number of blocks.

### 2.3 Cepstral Parameter Estimation Using Third-Order Statistics

Let us consider an autoregressive moving average (ARMA) sequence  $\{y(k)\}$  whose z-transform is give by

$$Y(z) = A z^{-r} I(z^{-1}) O(z) \quad (25)$$

where  $I(z^{-1})$  and  $O(z)$  denote the minimum and maximum phase components, respectively. Then the bicepstrum is defined as the inverse 2-D z-transform of the log-bispectrum [Nikias and Petropulu, 1993]. The z-transform of bicepstrum satisfies

$$\begin{aligned} Z\{b_h(m, n)\} &= \log |A^3| + \log I(z_1^{-1}) + \log I(z_2^{-1}) + \log I(z_1 \cdot z_2) \\ &\quad + \log O(z_1) + \log O(z_2) + \log O(z_1^{-1} \cdot z_2^{-1}) \end{aligned} \quad (26)$$

where  $\{b_h(m, n)\}$  is the bicepstrum. Then, the relationship between the bicepstrum and the complex cepstral parameters  $A^{(m)}$  and  $B^{(m)}$  is given in [Pan and Nikias, 1988; Nikias and Petropulu, 1993], where  $A^{(m)}$  and  $B^{(m)}$  are cepstral parameters containing information about the minimum

and maximum phases. Using partial differentiation with respect to  $z_1$  or  $z_2$ , we can obtain a direct relationship between the cepstral coefficients and the third-order statistics:

$$-n c_3(n, \ell) = \sum_{i=1}^{\infty} \left\{ A^{(i)} [c_3(n-i, \ell) - c_3(n+i, \ell+i)] + B^{(i)} [c_3(n-i, \ell-i) - c_3(n+i, \ell)] \right\}. \quad (27)$$

Using a finite sum and an overdetermined system of equations, we can obtain the cepstral parameters.

An alternative method is based on 2-D Fourier transform operation and is given by

$$n b_h(n, \ell) = F_2^{-1} \left\{ \frac{F_2\{n \cdot c_3(n, \ell)\}}{F_2\{c_3(n, \ell)\}} \right\} \quad (28)$$

where  $F_2\{\cdot\}$  and  $F_2^{-1}\{\cdot\}$  denote 2-D forward and inverse Fourier transforms.

## 2.4 Parameter Estimation of S $\alpha$ S Random Process

Let  $\{y(k), k = 1, 2, \dots, N\}$  be observed independent realizations of a S $\alpha$ S random variable  $Y$  of unknown characteristic exponent  $\alpha$ , location parameter  $\delta$ , and dispersion  $\gamma$ . We attempt to estimate the exact parameters of the S $\alpha$ S distribution of  $Y$  from the observed realizations. The estimation procedure we propose has a hierarchical rather than a simultaneous structure, as explained in the subsequent sections. More details on the rationale behind these estimators and on their performance can be found in [Tsihrintzis and Nikias, 1994].

### 2.4.1 Estimator of the Characteristic Exponent

For the estimation of the characteristic exponent  $\alpha$  of the probability density function (pdf), we propose the following algorithm. Consider a segmentation of the data into  $L$  nonoverlapping segments, each of length  $M = N/L$ :

$$\{y(1), y(2), \dots, y(N)\} = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(L)\}, \quad (29)$$

where  $\mathbf{y}(\ell) = \{y((\ell-1)N/L + 1), y((\ell-1)N/L + 2), \dots, y(\ell N/L)\}$ ,  $\ell = 1, 2, \dots, L$ . This segmentation is done arbitrarily for the time being and the reason for considering it will become apparent momentarily. Optimization of the segmentation is a topic of present and future research.

Let  $\bar{y}_\ell$  and  $\underline{y}_\ell$  be the maximum and the minimum of the data segment  $\mathbf{y}(\ell)$ . We then define

$$\bar{x}_\ell = \log \bar{y}_\ell \quad (30)$$

$$\underline{x}_\ell = -\log(-\underline{y}_\ell) \quad (31)$$

and the corresponding standard deviations

$$\bar{s} = \sqrt{\frac{1}{L-1} \sum_{\ell=1}^L (\bar{x}_\ell - \bar{x})^2} \quad (32)$$

$$\underline{s} = \sqrt{\frac{1}{L-1} \sum_{\ell=1}^L (\underline{x}_\ell - \underline{x})^2} \quad (33)$$

where

$$\bar{x} = \frac{1}{L} \sum_{\ell=1}^L \bar{x}_\ell \quad \underline{x} = \frac{1}{L} \sum_{\ell=1}^L \underline{x}_\ell. \quad (34)$$

With these definitions in mind, the estimate for the characteristic exponent  $\alpha$  of the S $\alpha$ S pdf takes the form

$$\hat{\alpha} = \frac{\pi}{2\sqrt{6}} \left( \frac{1}{\bar{s}} + \frac{1}{\underline{s}} \right). \quad (35)$$

#### 2.4.2 Estimator of the Location Parameter

For the estimation of the location parameter  $\delta$  of a S $\alpha$ S pdf, we propose the use of the sample median of the observations, i.e.

$$\hat{\delta} = \text{median} \{ y(1), y(2), \dots, y(N) \} \quad (36)$$

where the sample median is defined as follows: If the sample consists of an odd number  $N$  of observations, the median is defined as the center order statistic. If the sample consists of an even number  $N$  of observations, the median is defined as the average of the two center statistics. The sample median forms the maximum likelihood estimate of the location parameter of a Laplace (double exponential) distribution and, therefore, enjoys all the properties of maximum likelihood estimators in that case. Its performance as an estimator for the location parameter  $\delta$  of a S $\alpha$ S pdf is very robust.

### 2.4.3 Estimator of the Dispersion

For the estimation of the dispersion  $\gamma$  of a S $\alpha$ S pdf, we propose the following estimator which is based on the theory of fractional lower order moments of the pdf:

$$\hat{\gamma} = \left( \frac{\frac{1}{N} \sum_{k=1}^N |y(k) - \hat{\delta}|^p}{C(p, \hat{\alpha})} \right)^{\hat{\alpha}/p} \quad (37)$$

where  $C(p, \hat{\alpha})$  has been defined as

$$C(p, \hat{\alpha}) = \frac{1}{\cos(\frac{\pi}{2}p)} \frac{\Gamma(1 - p/\hat{\alpha})}{\Gamma(1 - p)} \quad (38)$$

and the choice of the order  $p$  ( $0 < p < \frac{\hat{\alpha}}{2}$ ) of the fractional moment is arbitrary.

As we can see, the dispersion estimator requires knowledge of the characteristic exponent and the location parameter of the S $\alpha$ S pdf. Thus, the dispersion estimate must be computed *after* estimates for the characteristic exponent and the location parameter have been obtained.



### 3 Experimental Results

We divided each data file into two parts and applied all the methods for data classification. The first part contains the data samples from 1 to 65,000 and the second one contains the data samples from 65,001 to 130,000. Note that the total number of data samples in each file is 131,072. Then, each data part was divided into 13 segments so that each segment contains 5,000 samples without overlapping except for estimation of the bispectrum and the cepstral parameters. When the estimates of the bispectrum and the cepstral parameters were obtained, each part was divided into several segments so that each segment contained 500 samples without data overlapping.

#### 3.1 Second-Order AR Models

##### 3.1.1 AR Coefficient Estimation Using the Gaussian ML Method

When the data pair is `trvhyd1` (smooth) and `trsvhyd1` (slot), the plots of the estimated AR coefficient clusters and the estimated variances obtained by the Gaussian ML method are shown in Fig. 1 and Fig. 2, respectively. For the smooth section test data, it is observed that the corresponding clusters are located at the upper-left corner of the figures. On the other hand, the clusters from the slot section test data are in the lower-right corner of the figures. The estimated variances from the smooth section test data begin to increase from segment number three, while those from the slot section test data noticeably increase from segment number six or seven. Also, note that the former variances increase slower and are less than the latter variances.

Fig. 3 and Fig. 4 illustrate the clusters of the estimated AR coefficients and the estimated variances for the data sets `trvhyd2` and `trsvhyd2`, respectively. Fig. 5 and Fig. 6 illustrate the corresponding plots for the data sets `trvhyd3` and `trsvhyd3`, respectively.

From the results shown in these figures, we can conclude that we can discriminate the smooth section test data from the slot section test data using clusters of the AR coefficients in the following way: if a cluster is located in the region where  $\hat{a}_1$  is around  $-2$  and  $\hat{a}_2$  is around unity; and the estimated variances begin to slowly increase from segment number three, then the data section is from the smooth section test. If a cluster is located in the region where  $\hat{a}_1 \geq -1.5$  and  $\hat{a}_2 \leq 0.5$ ; and the estimated variances begin to rapidly increase from segment number six, then the data section is from the slot section test.

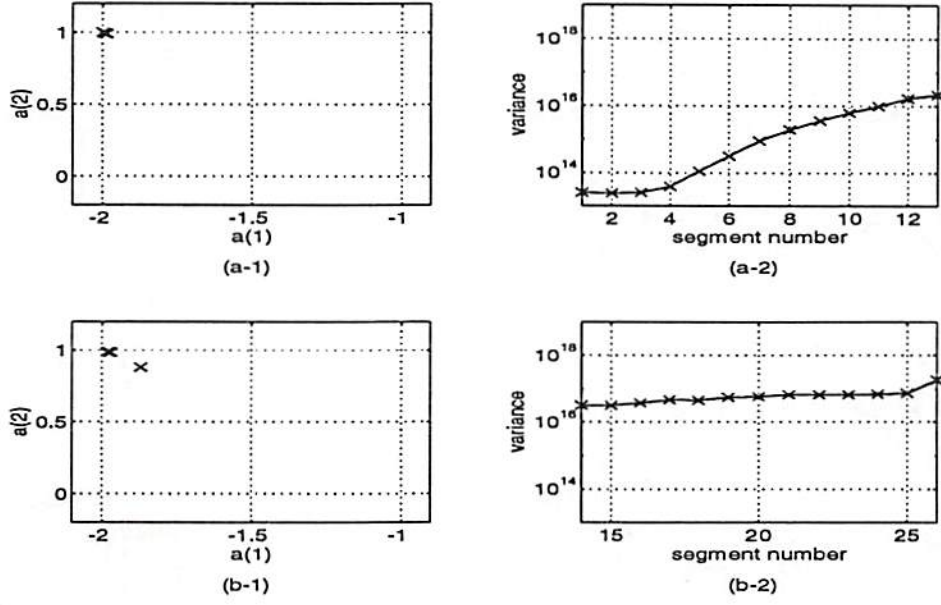


Figure 1: Results obtained by the Gaussian ML method. Using the first half of `trvhyd1`: (a-1) cluster of the estimated AR coefficients and (a-2) estimated variances; using the second half of `trvhyd1`: (b-1) cluster of the estimated AR coefficients and (b-2) estimated variances.

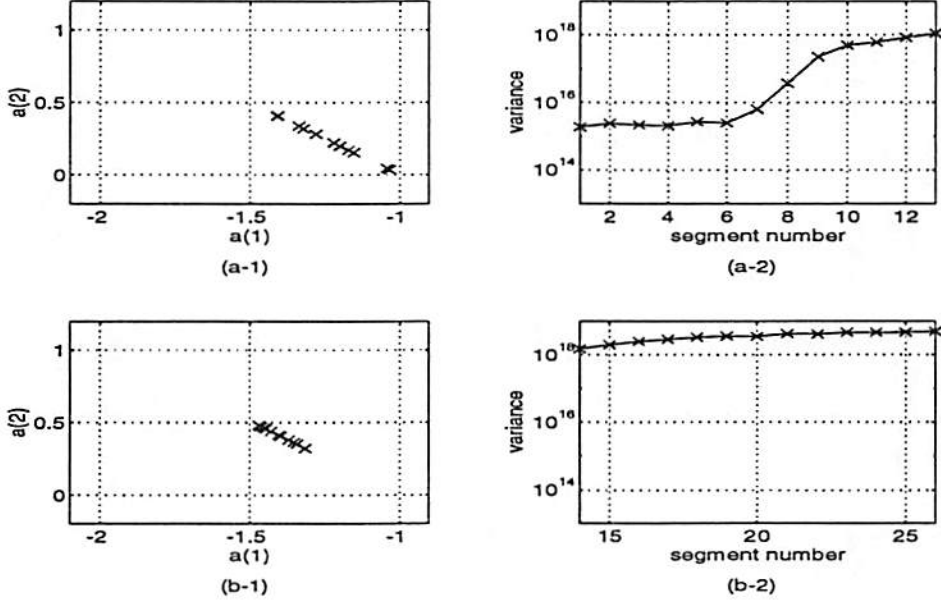


Figure 2: Results obtained by the Gaussian ML method. Using the first half of `trsvhyd1`: (a-1) cluster of the estimated AR coefficients and (a-2) estimated variances; using the second half of `trsvhyd1`: (b-1) cluster of the estimated AR coefficients and (b-2) estimated variances.

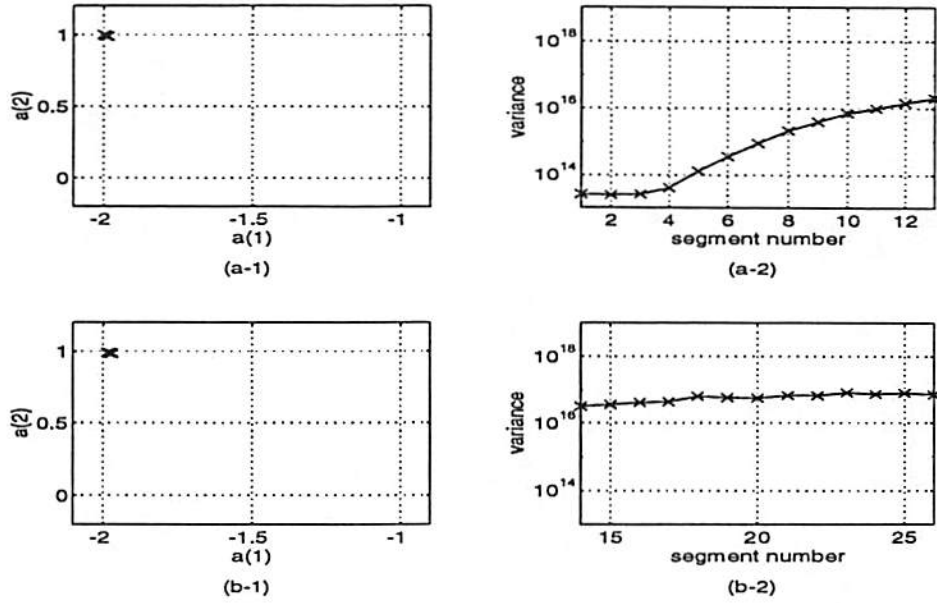


Figure 3: Results obtained by the Gaussian ML method. Using the first half of `trvhyd2`: (a-1) cluster of the estimated AR coefficients and (a-2) estimated variances; using the second half of `trvhyd2`: (b-1) cluster of the estimated AR coefficients and (b-2) estimated variances.

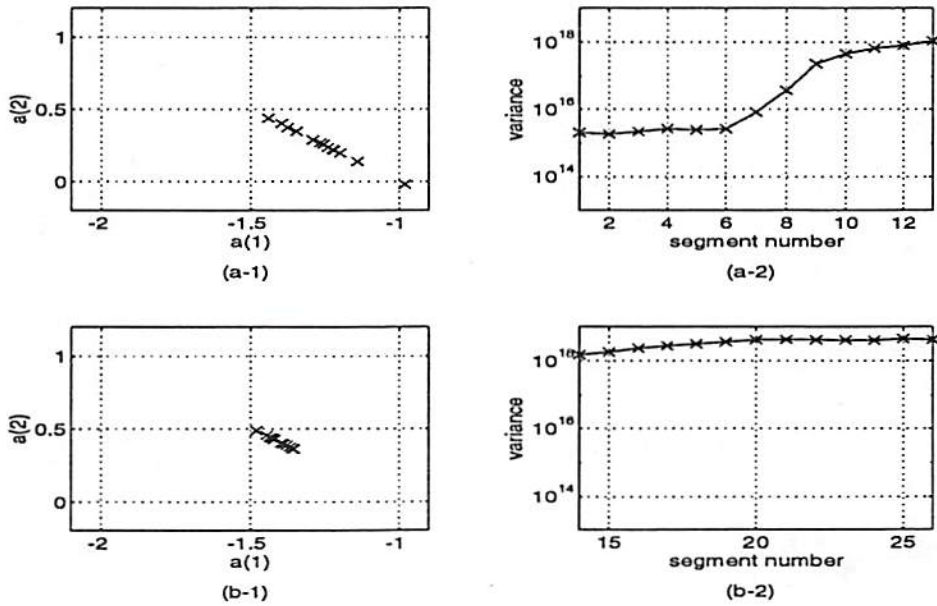


Figure 4: Results obtained by the Gaussian ML method. Using the first half of `trsvhyd2`: (a-1) cluster of the estimate the AR coefficients and (a-2) estimated variances; using the second half of `trsvhyd2`: (b-1) cluster of the estimate the AR coefficients and (b-2) estimated variances.



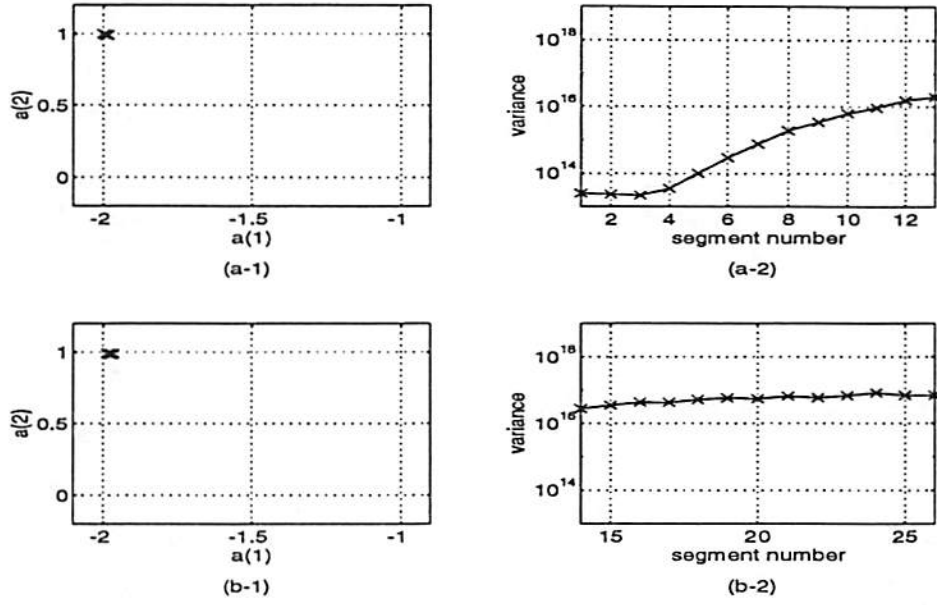


Figure 5: Results obtained by the Gaussian ML method. Using the first half of `trvhyd3`: (a-1) cluster of the estimate the AR coefficients and (a-2) estimated variances; using the second half of `trvhyd3`: (b-1) cluster of the estimate the AR coefficients and (b-2) estimated variances.

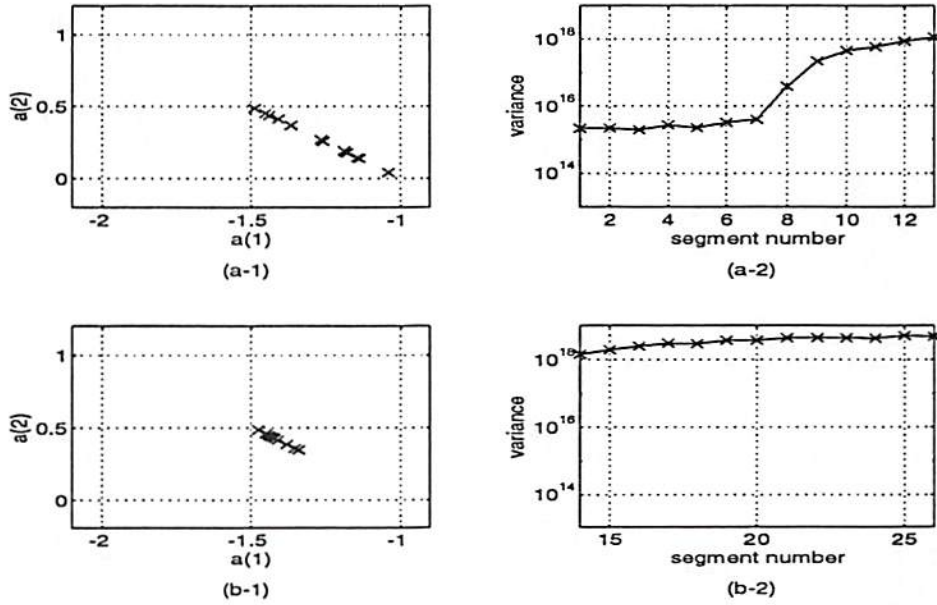


Figure 6: Results obtained by the Gaussian ML method. Using the first half of `trsvhyd3`: (a-1) cluster of the estimate the AR coefficients and (a-2) estimated variances; using the second half of `trsvhyd3`: (b-1) cluster of the estimate the AR coefficients and (b-2) estimated variances.



### 3.1.2 AR Coefficient Estimation Using the Normal Equations Based on Second-Order Statistics

Note that we formulated the normal equations based on second-order statistics using the maximum auto-correlation lag of 12. For the data pair **trvhyd1** and **trsvhyd1**, the cluster plots of the estimated AR coefficients obtained by the second-order statistics-based normal equations are shown in Fig. 7. For the smooth section test data, it is observed that the corresponding points are located more closely to each other than those obtained from the slot section test data. This phenomenon can also be observed in Fig. 8 for the data sets **trvhyd2** and **trsvhyd2** and in Fig. 9 for the data sets **trvhyd3** and **trsvhyd3**.

In addition, it is observed that the points are more tightly close to each other and clearly separated for the second half of the data samples than for the first half. However, when we have the first half of the data, it is not possible to clearly classify whether the data is from the smooth section test or from the slot section test.

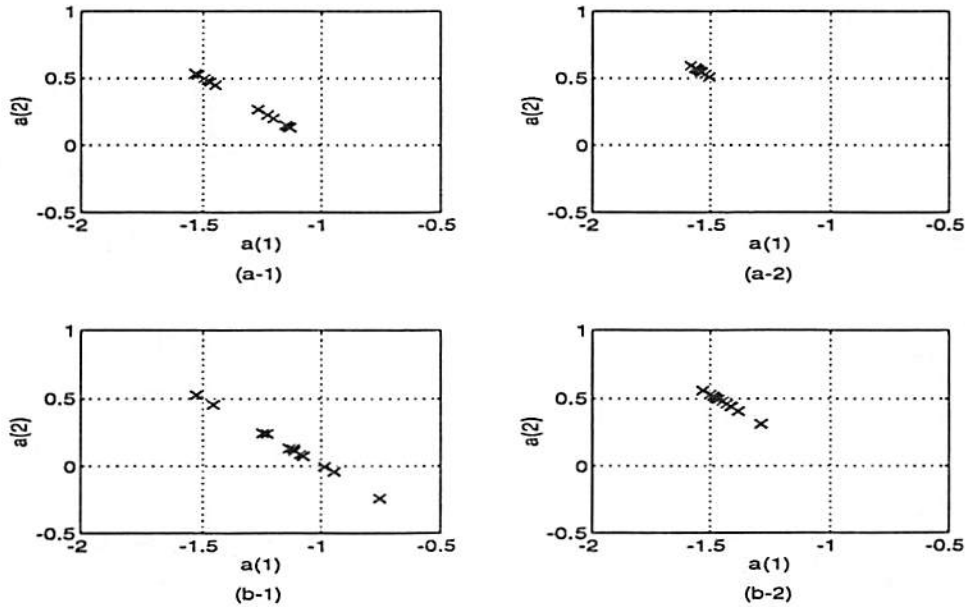


Figure 7: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on second-order statistics with the data: (a-1) first half of **trvhyd1**, (a-2) second half of **trvhyd1**; (b-1) first half of **trsvhyd1**, (b-2) second half of **trsvhyd1**.

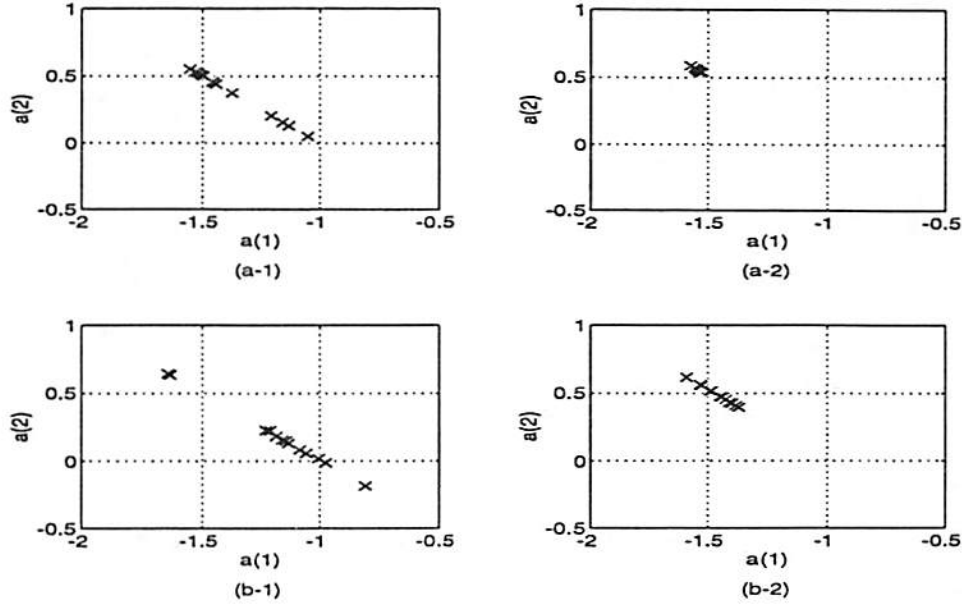


Figure 8: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on second-order statistics with the data: (a-1) first half of `trvhyd2`, (a-2) second half of `trvhyd2`; (b-1) first half of `trsvhyd2`, (b-2) second half of `trsvhyd2`.

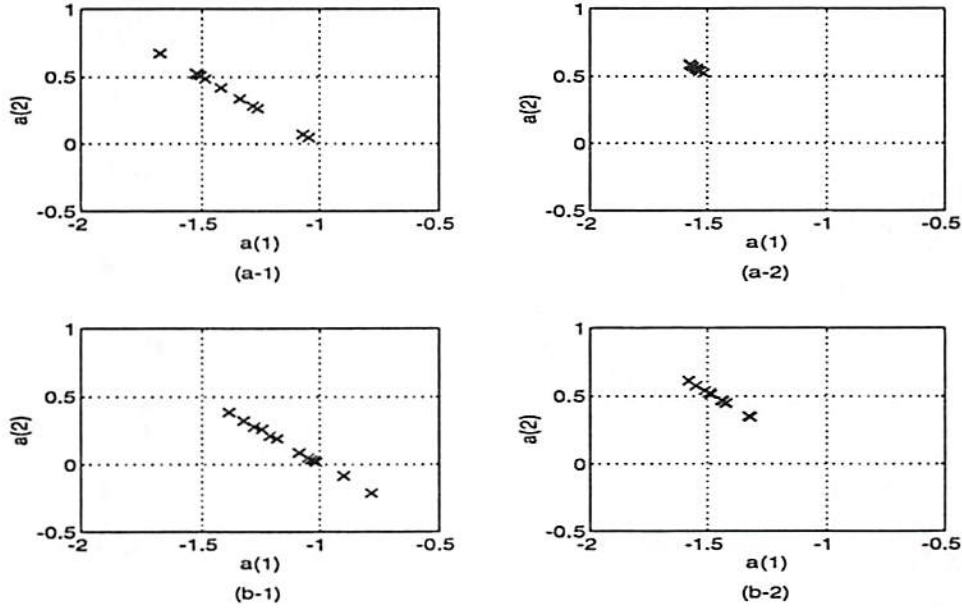


Figure 9: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on second-order statistics with the data: (a-1) first half of `trvhyd3`, (a-2) second half of `trvhyd3`; (b-1) first half of `trsvhyd3`, (b-2) second half of `trsvhyd3`.

### 3.1.3 AR Coefficient Estimation Using the Normal Equations Based on Third-Order Statistics

We formulated the normal equations based on third-order statistics using the maximum lag of 12. For the data pair *trvhyd1* and *trsvhyd1*, the cluster plots of the estimated AR coefficients obtained by the third-order statistics-based normal equations are shown in Fig. 10. And, Fig. 11 shows the corresponding plots for the data sets *trvhyd2* and *trsvhyd2*, while Fig. 12 corresponds to the data sets *trvhyd3* and *trsvhyd3*.

From the results shown in Fig. 10 – Fig. 12, we can conclude that it is not possible to classify the data sets by using the AR coefficient clusters obtained by the normal equations based on third-order statistics because the clusters with the smooth section test data are very similar to those with the slot section test data.

For an additional method, we fitted each cluster into a straight line in least-squares sense. However, we could not find any significant difference in the values of the slopes and the y-intercepts of the corresponding straight lines.

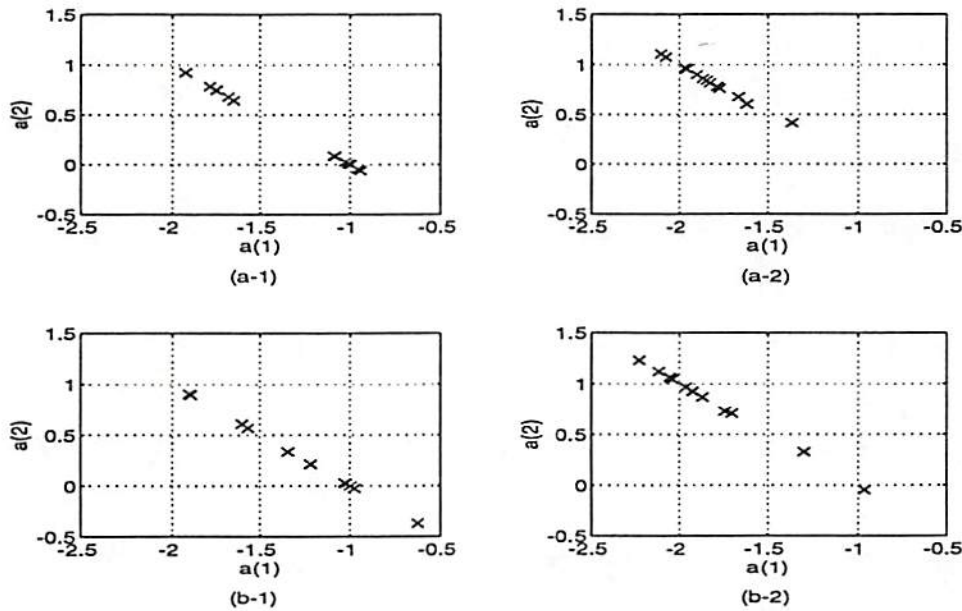


Figure 10: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on third-order statistics with the data: (a-1) first half of *trvhyd1*, (a-2) second half of *trvhyd1*; (b-1) first half of *trsvhyd1*, (b-2) second half of *trsvhyd1*.

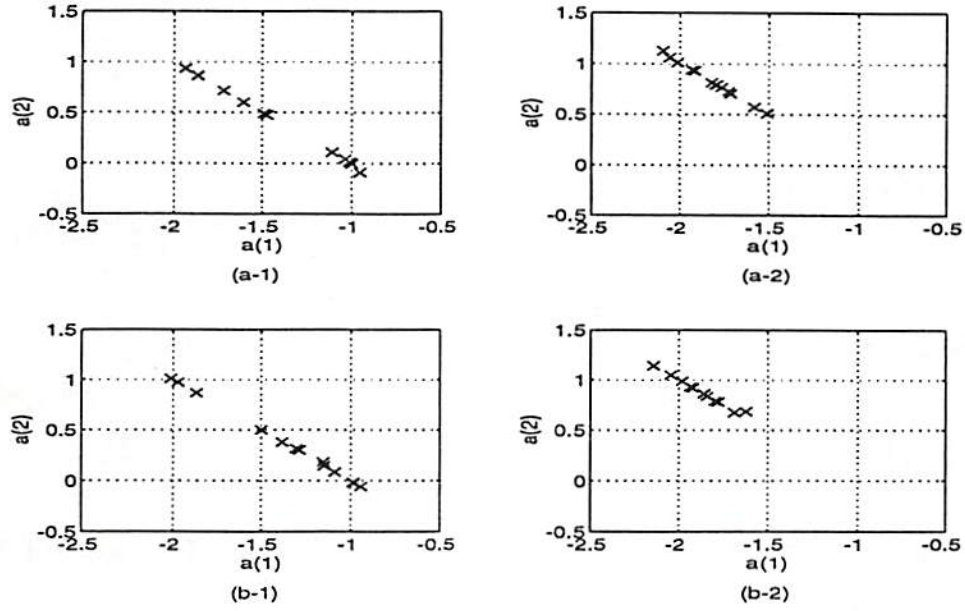


Figure 11: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on third-order statistics with the data: (a-1) first half of *trvhyd2*, (a-2) second half of *trvhyd2*; (b-1) first half of *trsvhyd2*, (b-2) second half of *trsvhyd2*.

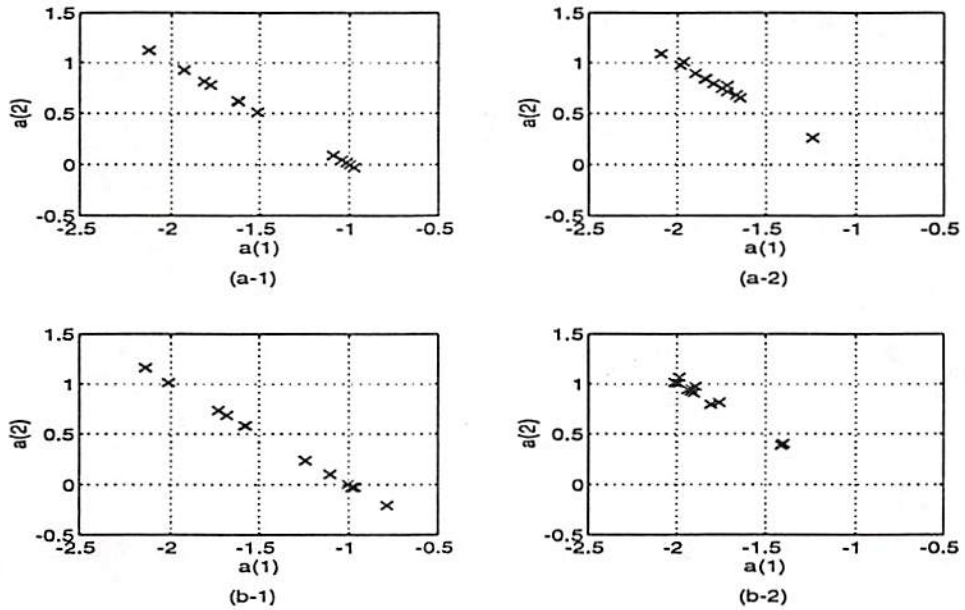


Figure 12: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by the normal equations based on third-order statistics with the data: (a-1) first half of *trvhyd3*, (a-2) second half of *trvhyd3*; (b-1) first half of *trsvhyd3*, (b-2) second half of *trsvhyd3*.



### 3.1.4 Generalized Yule-Walker Equations

**Results with  $p = 0.5$ :** With the three pairs of real data sets, we estimated the second-order AR coefficients from the generalized Yule-Walker equations using FLOM estimates of order  $p = 0.5$ . For the independent observations  $\{(x_n, y_n), n = 1, 2, \dots, N\}$ , the FLOM estimator is given by [Nikias and Shao, 1995]

$$\hat{\lambda}_{\text{FLOM}} = \frac{\sum_{n=1}^N x_n |y_n|^{p-1} \text{sign}(y_n)}{\sum_{n=1}^N |y_n|^p} \quad (39)$$

with  $1 \leq p < \alpha$ . After estimating the covariation coefficients using the FLOM estimator with  $p = 0.5$ , we solved the generalized Yule-Walker equations. The cluster plots of the estimated AR coefficients are shown in Fig. 13 – Fig. 15 for each pair of the data sets.

According to these figures, each figure using data sets from the smooth section test contains one or two points that clearly locate far away from the cluster. On the other hand, there is no such point in the figures using slot test section data sets. But, it seems not a good idea to use a classification method depending on one or two points rather than clusters, because the points may not appear if one has different parts or numbers of data.

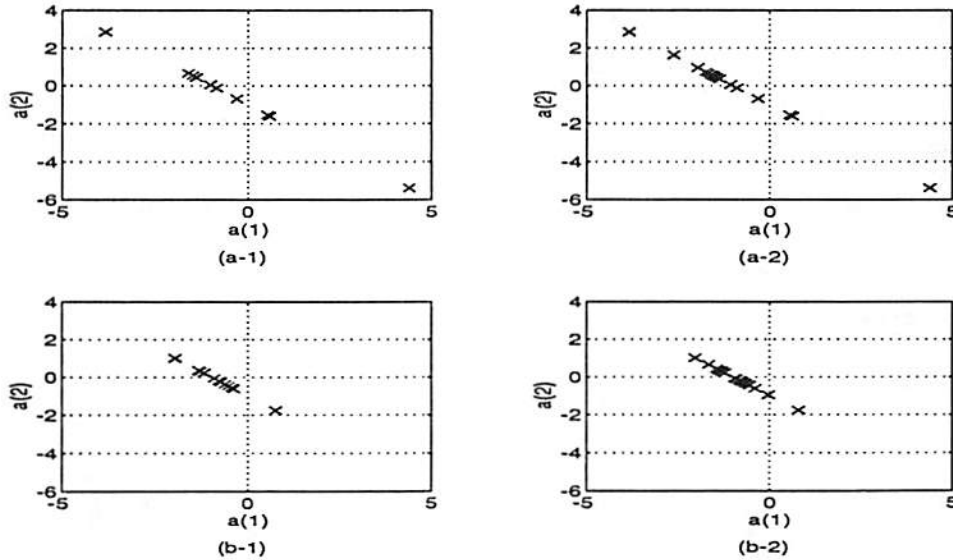


Figure 13: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 0.5$  and the data: (a-1) first half of *trvhyd1*, (a-2) second half of *trvhyd1*; (b-1) first half of *trsvhyd1*, (b-2) second half of *trsvhyd1*.

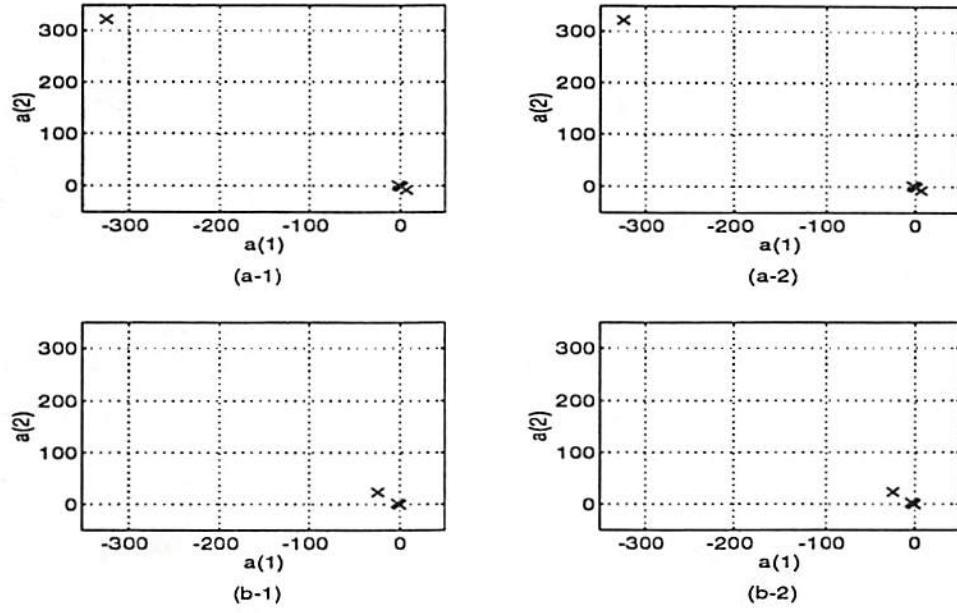


Figure 14: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 0.5$  and the data: (a-1) first half of trvhyd2, (a-2) second half of trvhyd2; (b-1) first half of trsvhyd2, (b-2) second half of trsvhyd2.

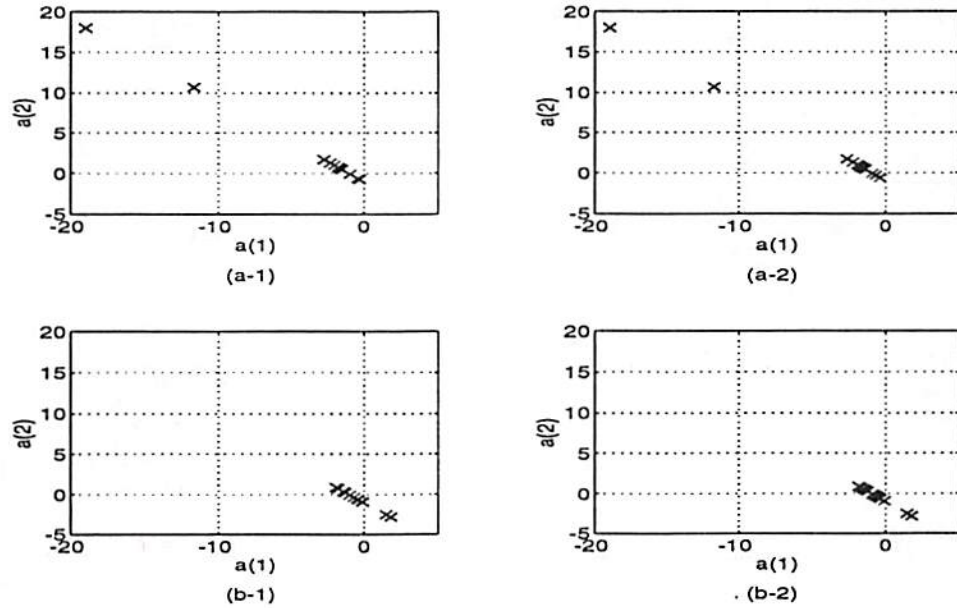


Figure 15: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 0.5$  and the data: (a-1) first half of trvhyd3, (a-2) second half of trvhyd3; (b-1) first half of trsvhyd3, (b-2) second half of trsvhyd3.

**Results with  $p = 1.0$ :** Fixing the value of  $p$  at 1.0, we estimated the covariation coefficients using the FLOM estimator and solved the generalized Yule-Walker equations. Note that setting  $p = 1.0$  make the FLOM estimator simpler than any other setting in the sense of computational complexity. Plots of clusters of the estimated AR coefficients are shown in Fig. 16 – Fig. 18 for each pair of the given data sets.

Fig. 16 and Fig. 18 illustrate two clearly separated clusters. All the points from the smooth test section data cluster in the region where  $a_1 < -1.5$  and  $a_2 > 0.5$ . On the other hand, all the points from the slot test section data cluster in the region where  $a_1 > -1.5$  and  $a_2 < 0.5$ . In Fig. 17 (b-1) and (b-2), we can observe one point located around  $(-1.6, 0.6)$  that is far away from its corresponding cluster and is closely located to the clusters for the smooth section test data. It is clear, however, that the main clusters are well separated.

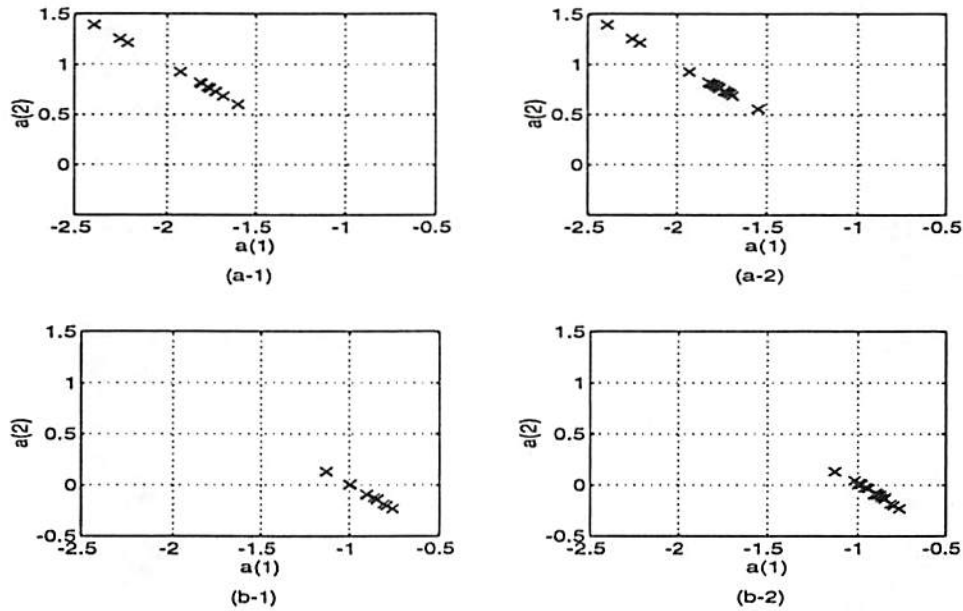


Figure 16: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 1.0$  and the data: (a-1) first half of trvhyd1, (a-2) second half of trvhyd1; (b-1) first half of trsvhyd1, (b-2) second half of trsvhyd1.

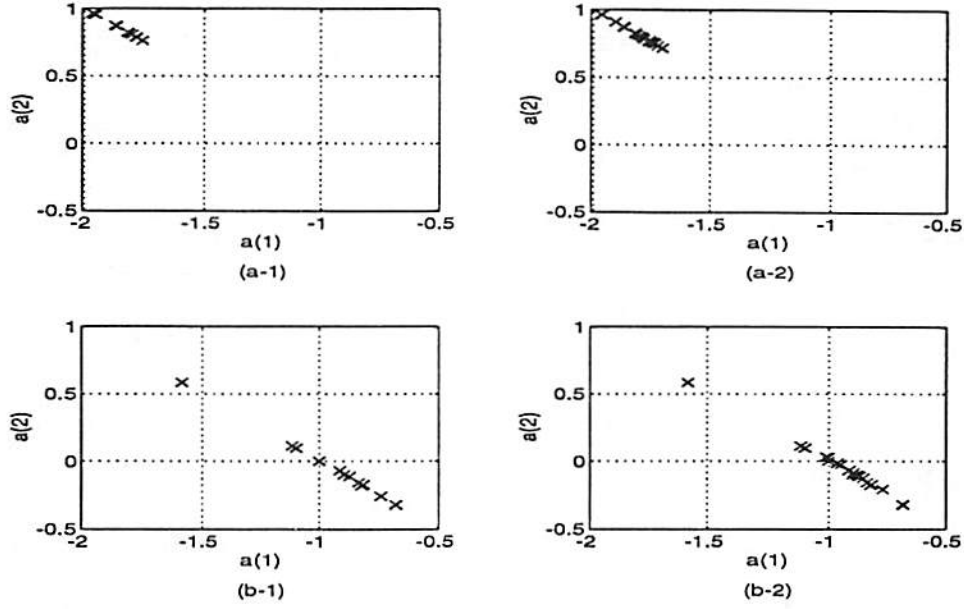


Figure 17: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 1.0$  and the data: (a-1) first half of `trvhyd2`, (a-2) second half of `trvhyd2`; (b-1) first half of `trsvhyd2`, (b-2) second half of `trsvhyd2`.

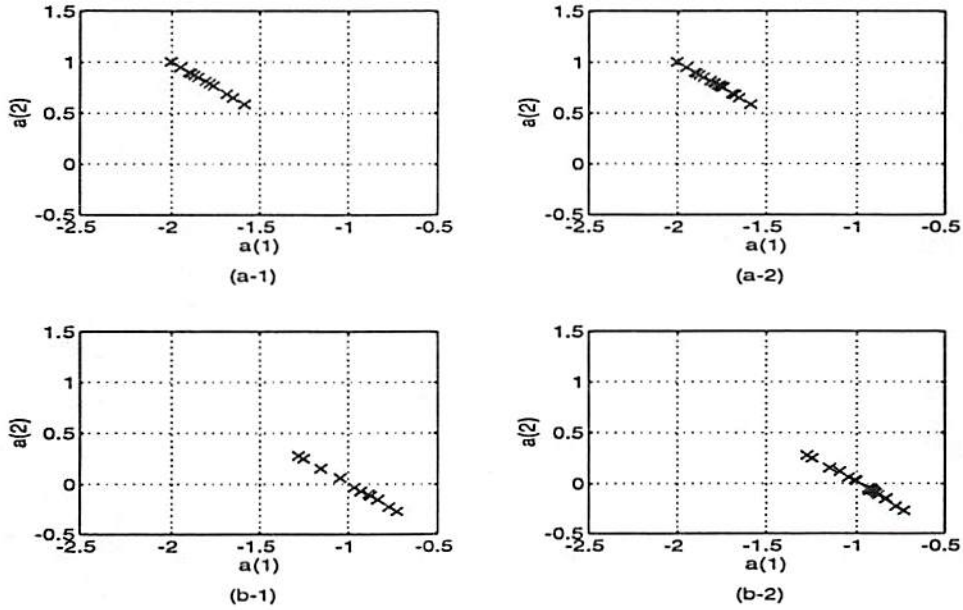


Figure 18: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 1.0$  and the data: (a-1) first half of `trvhyd3`, (a-2) second half of `trvhyd3`; (b-1) first half of `trsvhyd3`, (b-2) second half of `trsvhyd3`.



**Results with  $p = 1.5$ :** Next we fixed the value of  $p$  at 1.5 and estimated the covariation coefficients using the FLOM estimator and solved the generalized Yule–Walker equations. Plots of the clusters of the estimated AR coefficients are shown in Fig. 19 – Fig. 21 for each pair of the real data sets.

From these results, we can observe clearly improved performance. The clusters from between the smooth section test data and the slot section test data are clearly separated in all the figures including Fig. 20 obtained from the pair of data sets `trvhyd2` and `trsvhyd2` which provide one outsider point when  $p = 1.0$ . All the points from the smooth test section data are located in the region where  $a_1 < -1.5$  and  $a_2 > 0.5$ . On the other hand, all the points from the slot test section data are located in the region where  $a_1 > -1.5$  and  $a_2 < 0.5$ .

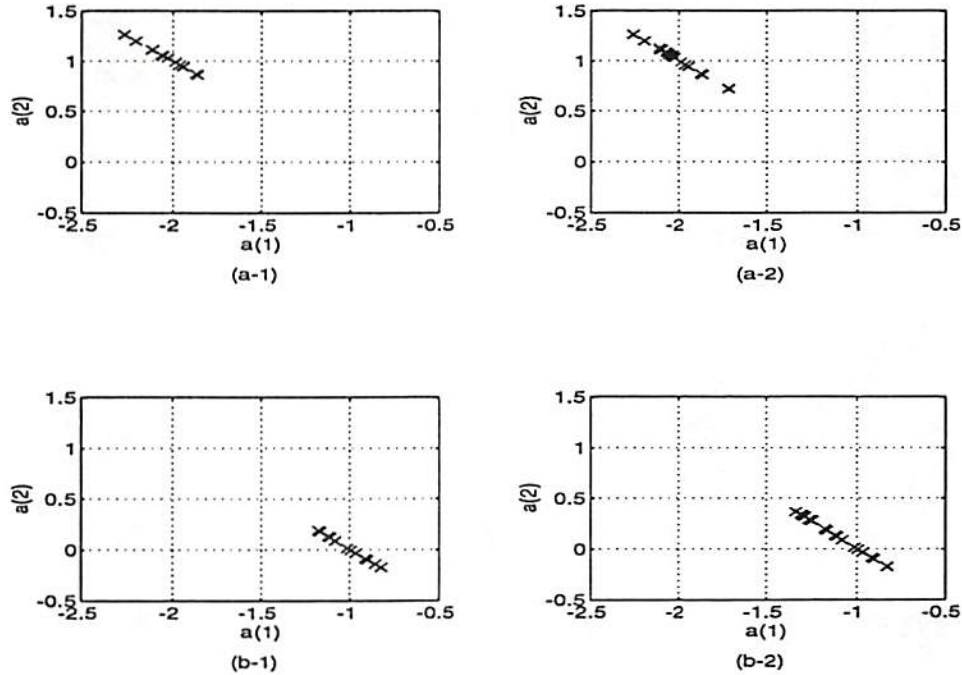


Figure 19: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule–Walker equation with  $p = 1.5$  and the data: (a-1) first half of `trvhyd1`, (a-2) second half of `trvhyd1`; (b-1) first half of `trsvhyd1`, (b-2) second half of `trsvhyd1`.

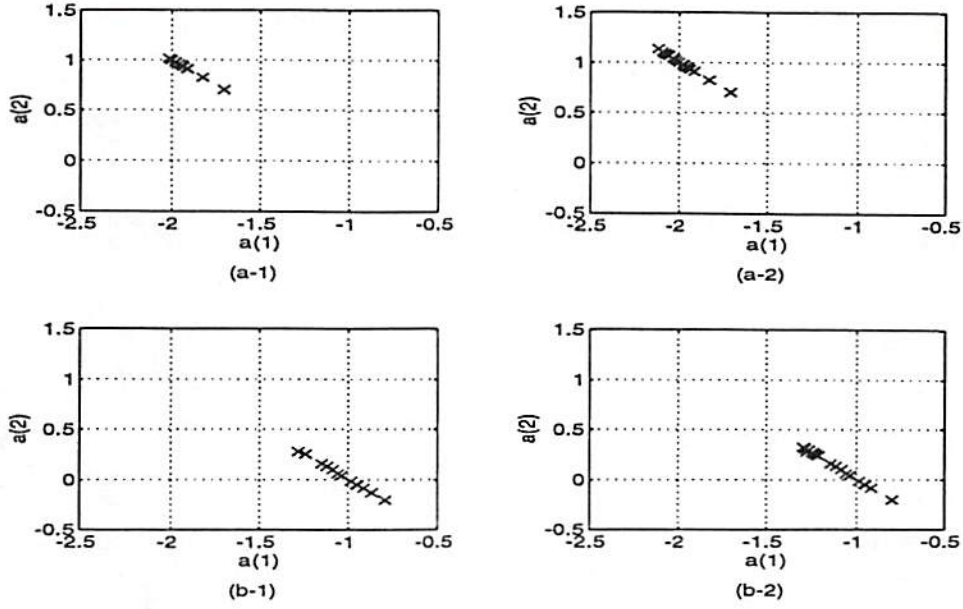


Figure 20: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 1.5$  and the data: (a-1) first half of trvhyd2, (a-2) second half of trvhyd2; (b-1) first half of trsvhyd2, (b-2) second half of trsvhyd2.

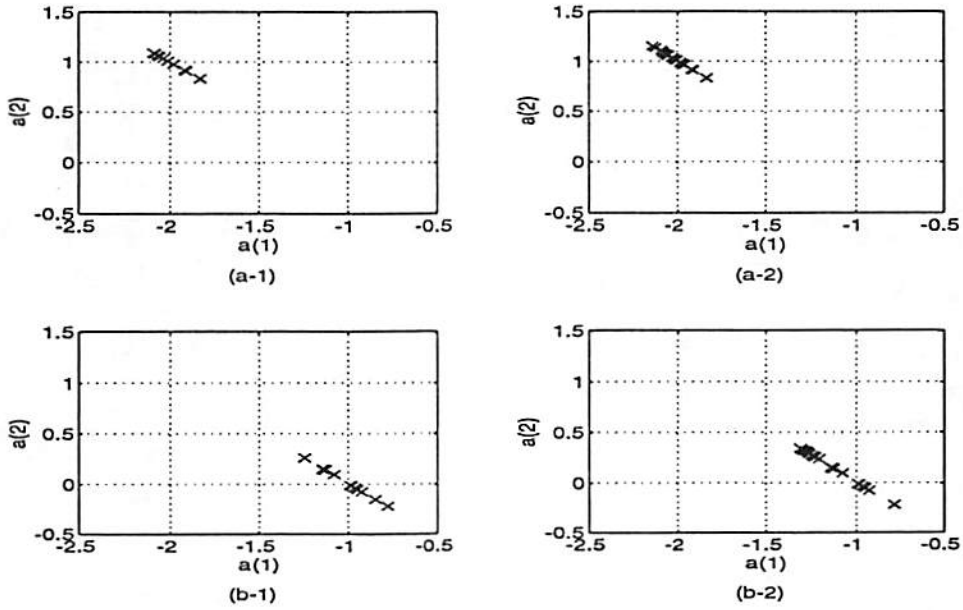


Figure 21: Clusters of the estimated AR coefficients  $\hat{a}_1$  and  $\hat{a}_2$  obtained by generalized Yule-Walker equation with  $p = 1.5$  and the data: (a-1) first half of trvhyd3, (a-2) second half of trvhyd3; (b-1) first half of trsvhyd3, (b-2) second half of trsvhyd3.

## 3.2 Power Spectrum and Bispectrum

In this section, we present the classification method using the estimated power spectra and bispectra of the given data sets.

### 3.2.1 Power Spectrum Estimation

When the Blackman-Tukey power spectrum estimates are obtained, the maximum auto-correlation lag was fixed at 64. Note that all the plotted power spectra were obtained by averaging power spectrum estimates of each segment. Fig. 22 illustrates the power spectra of the first and second parts of the data sets *trvhyd1* and *trsvhyd1*. As observed in the figure, there is no big difference between the power spectra of the first parts of the smooth and the slot section test data. On the other hand, a clear difference can be observed in the power spectra of their second parts of the data samples. The power spectrum of the second part of the smooth section test data

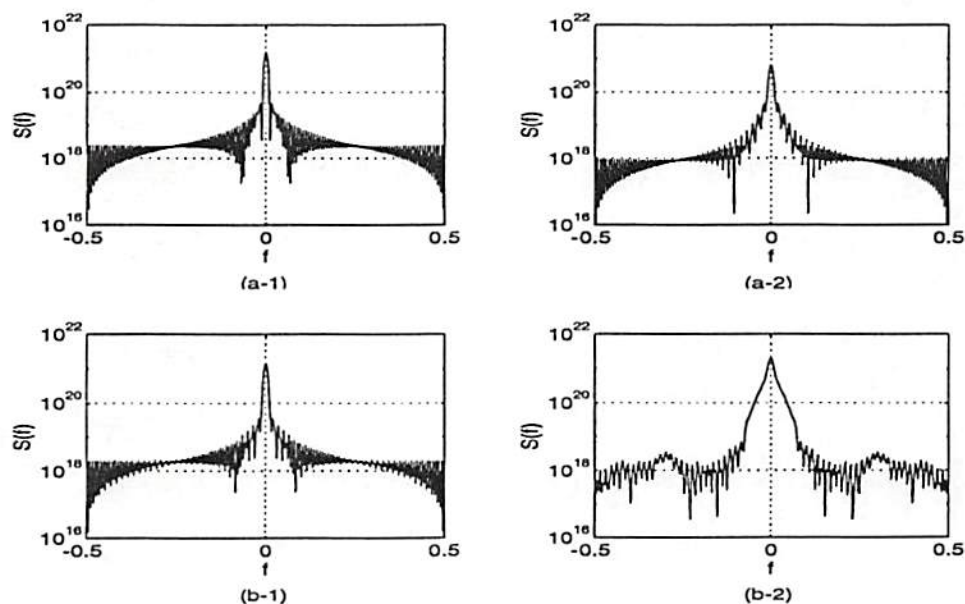


Figure 22: Estimated power spectra of (a-1) first half of *trvhyd1*, (a-2) second half of *trvhyd1*; (b-1) first half of *trsvhyd1*, (b-2) second half of *trsvhyd1*.

is similar to that of its first part except for increasing sidelobes (see Fig. 22 (a-2)). On the other hand, the power spectrum of the second part of the slot section test data has a smoothed main lobe whose bandwidth is quite a large. Thus, we can classify the second part of data samples (steady state data) using its power spectrum.

This phenomenon can also be observed in Fig. 23 obtained from the data sets **trvhyd2** and **trsvhyd2** and in Fig. 24 obtained from the data sets **trvhyd3** and **trsvhyd3**. From the results, we can conclude that the power spectrum estimation technique can be used to classify the data obtained from steady state in the following way: if estimates of the power spectrum have a wide and smoothed main lobe, then the data is obtained from the slot section test. If they have a narrow main lobe, then the data is from the smooth section test.

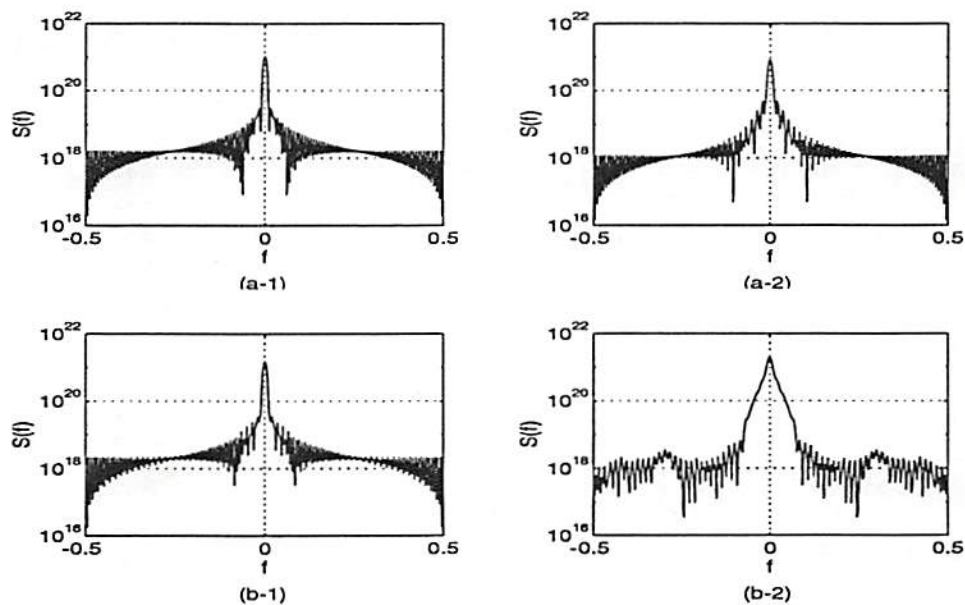


Figure 23: Estimated power spectra of (a-1) first half of **trvhyd2**, (a-2) second half of **trvhyd2**; (b-1) first half of **trsvhyd2**, (b-2) second half of **trsvhyd2**.



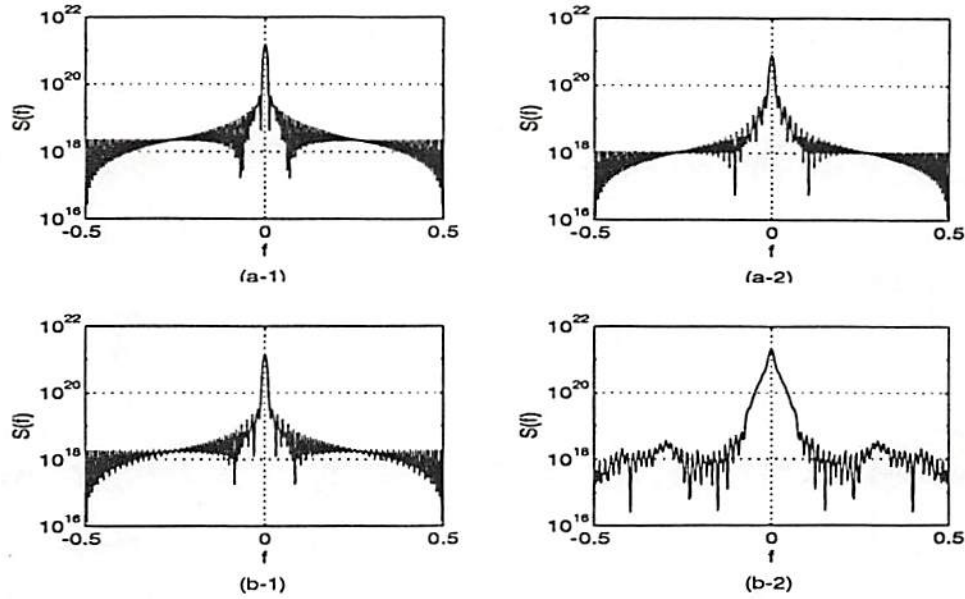


Figure 24: Estimated power spectra of (a-1) first half of *trvhyd3*, (a-2) second half of *trvhyd3*; (b-1) first half of *trsvhyd3*, (b-2) second half of *trsvhyd3*.

### 3.2.2 Bispectrum Estimation

When the bispectrum of each data part was estimated, each part of the data sets was segmented so that each segment contained 500 data samples. After estimating the bispectrum of each segment using the direct method described in the previous section, the final bispectrum estimates were obtained by averaging over all the bispectra of the individual segments.

Fig. 25 illustrates the magnitudes and phases of the bispectra using the first and second parts of the smooth section test data *trvhyd1*. Fig. 26 shows those of the slot section test data *trsvhyd1*. Fig. 27 and Fig. 28 show those of the data sets *trvhyd2* and *trsvhyd2*, respectively. And, Fig. 29 and Fig. 30 show those of the data sets *trvhyd3* and *trsvhyd3*, respectively.

First let us compare the magnitudes of the estimated bispectra. Comparing them, we can observe that the magnitudes of the bispectra of the first parts of the smooth section test data has several dominant and sharp peaks. On the other hand, those of the corresponding slot section test data have more smoothed tops and no dominantly sharp peak. Although the magnitudes of the bispectra of the second parts of the smooth section test data is also different from those of the corresponding slot section test data, the difference is not clear enough to specify a certain rule

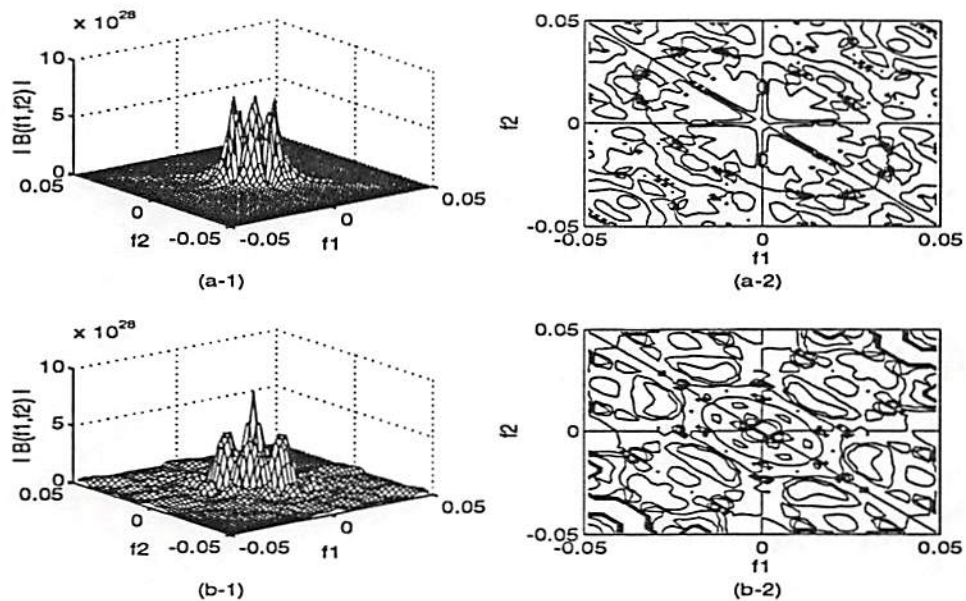


Figure 25: Estimated Bispectra of the first half of *trvhyd1*: (a-1) magnitude and (a-2) phase; estimated Bispectra of the second half of *trvhyd1*: (b-1) magnitude and (b-2) phase.

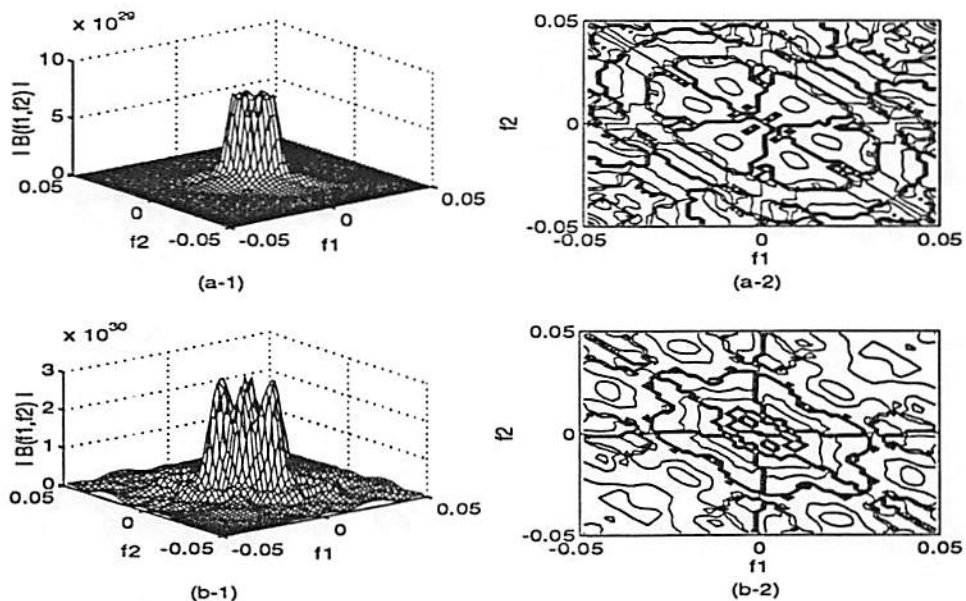


Figure 26: Estimated Bispectra of the first half of *trsvhyd1*: (a-1) magnitude and (a-2) phase; estimated bispectra of the second half of *trsvhyd1*: (b-1) magnitude and (b-2) phase.

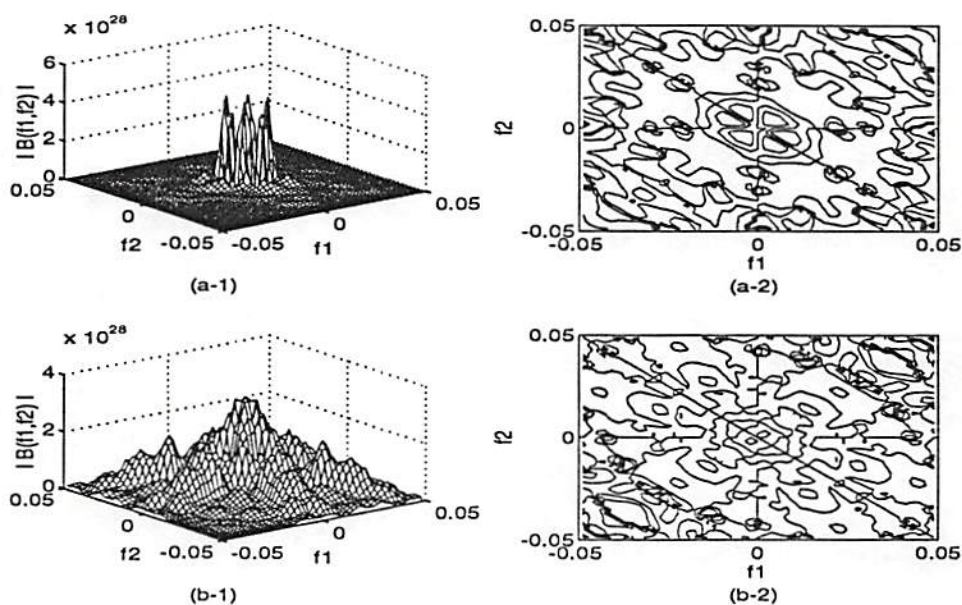


Figure 27: Estimated Bispectra of the first half of *trvhyd2*: (a-1) magnitude and (a-2) phase; estimated bispectra of the second half of *trvhyd2*: (b-1) magnitude and (b-2) phase.

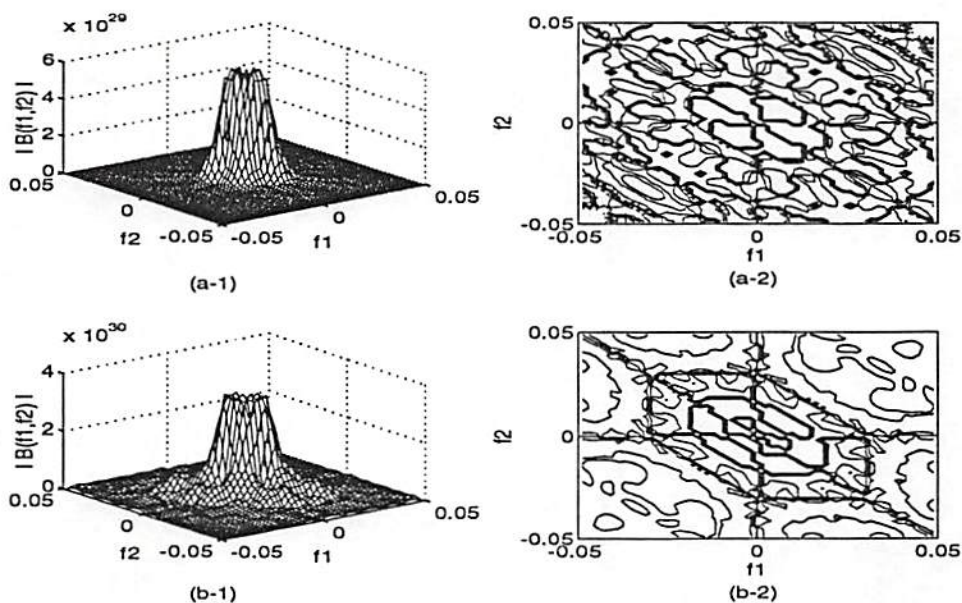


Figure 28: Estimated Bispectra of the first half of *trsvhyd2*: (a-1) magnitude and (a-2) phase; estimated bispectra of the second half of *trsvhyd2*: (b-1) magnitude and (b-2) phase.



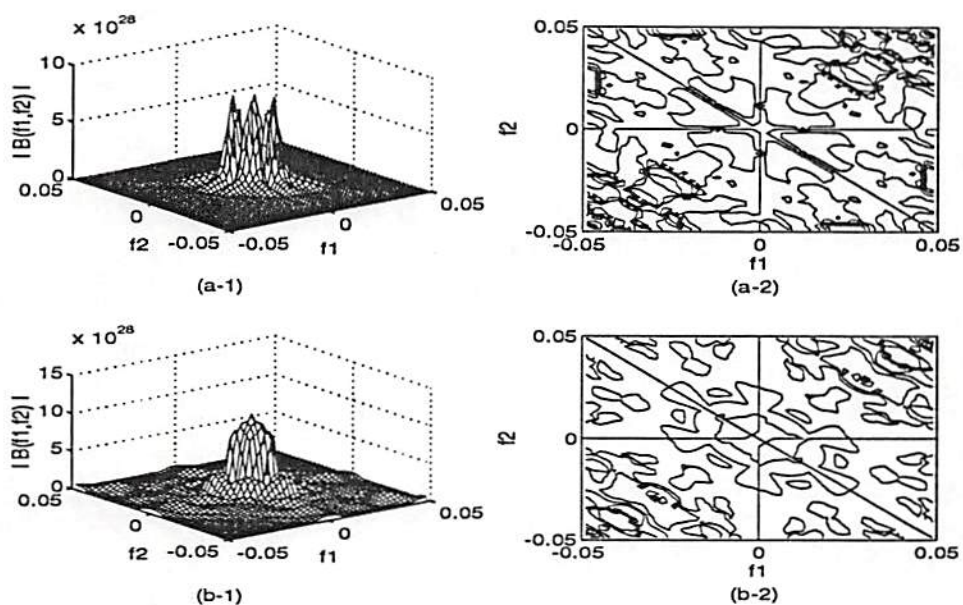


Figure 29: Estimated Bispectra of the first half of trvhyd3: (a-1) magnitude and (a-2) phase; estimated bispectra of the second half of trvhyd3: (b-1) magnitude and (b-2) phase.

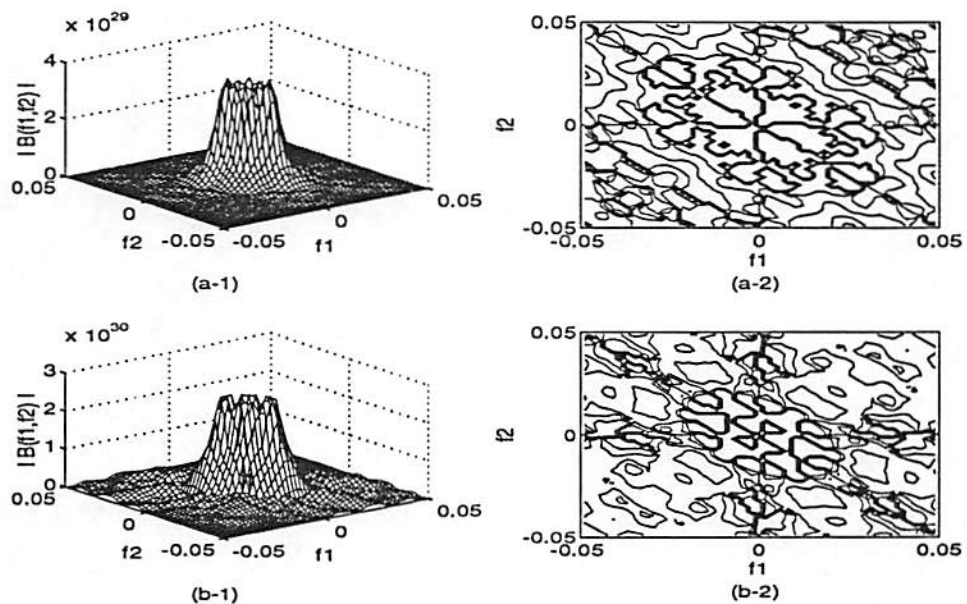


Figure 30: Estimated Bispectra of the first half of trsvhyd3: (a-1) magnitude and (a-2) phase; estimated bispectra of the second half of trsvhyd3: (b-1) magnitude and (b-2) phase.



for classification. There are no common characteristics in the bispectra obtained by the three smooth section test data sets. Each bispectrum has quite different shape. From the results, we can conclude that the magnitude of the estimated bispectrum can be used for data classification if the first part of the data samples (transient data) is given. If its magnitude has several dominant and sharp peaks, the data comes from the smooth section test. Otherwise, it comes from the slot section test.

Instead of the magnitudes of the bispectra, let us now observe their phases. It is not difficult to distinguish the phases of the estimated bispectra of the smooth section test data from those of the slot section test data. In all the phase contour plots obtained from the slot section test data, we can observe thick contour lines locating quite densely in the middle (around origin) of the plots. On the other hand, the corresponding thick lines in all the phase contour plots obtained from the smooth section test data are sparse and can be found around the boundary of the contour plots. Therefore, we can use the thick lines in the phase contour plots for classification. For specifying the thick lines, we need more study on the phase contour plots.

### 3.3 Complex Cepstral Parameter Estimation

Using the direct relationship between the cepstral parameters and the third-order statistics described in the previous section, we estimated the cepstral parameters. Note that  $A(m)$  and  $B(m)$  contain the minimum phase and maximum phase information, respectively. When the cepstral parameters of each data part were estimated, each part was segmented so that each segment contained 500 data samples and the length of each cepstral parameter was fixed at 30.

Fig. 31 illustrates the minimum and maximum phase cepstral parameters and their cluster plots for the first and second parts of the smooth section test data `trvhyd1`. Fig. 32 shows those of the slot section test data `trsvhyd1`. Fig. 33 and Fig. 34 show those of the test data sets `trvhyd2` and `trsvhyd2`, respectively; and Fig. 35 and Fig. 36 show those of the test data sets `trvhyd3` and `trsvhyd3`, respectively.

From these plots using cepstral parameters, we can conclude that it is not possible to specify a certain rule for classification.

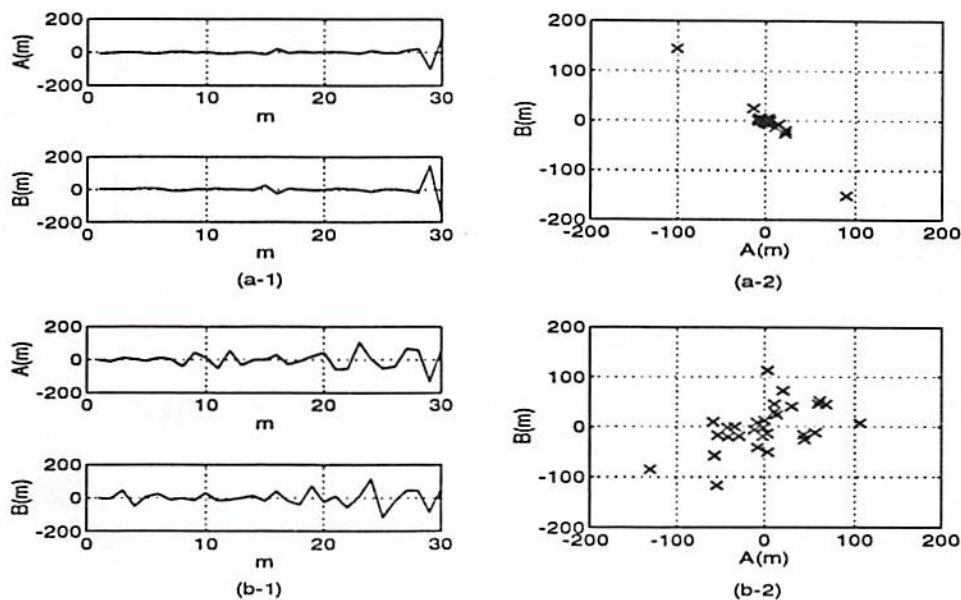


Figure 31: Using the first half of *trvhyd1*, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of *trvhyd1*, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.

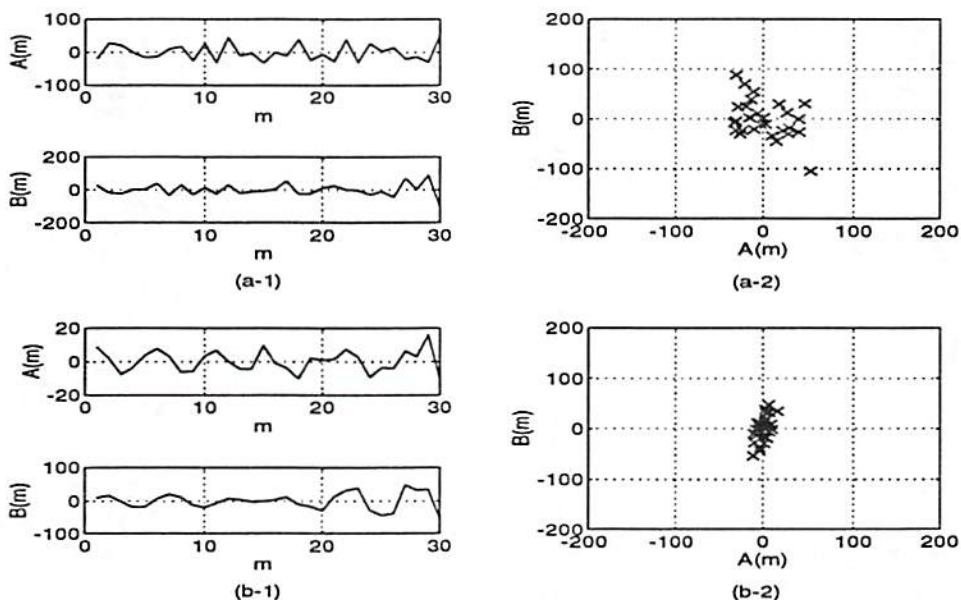


Figure 32: Using the first half of *trsvhyd1*, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of *trsvhyd1*, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.

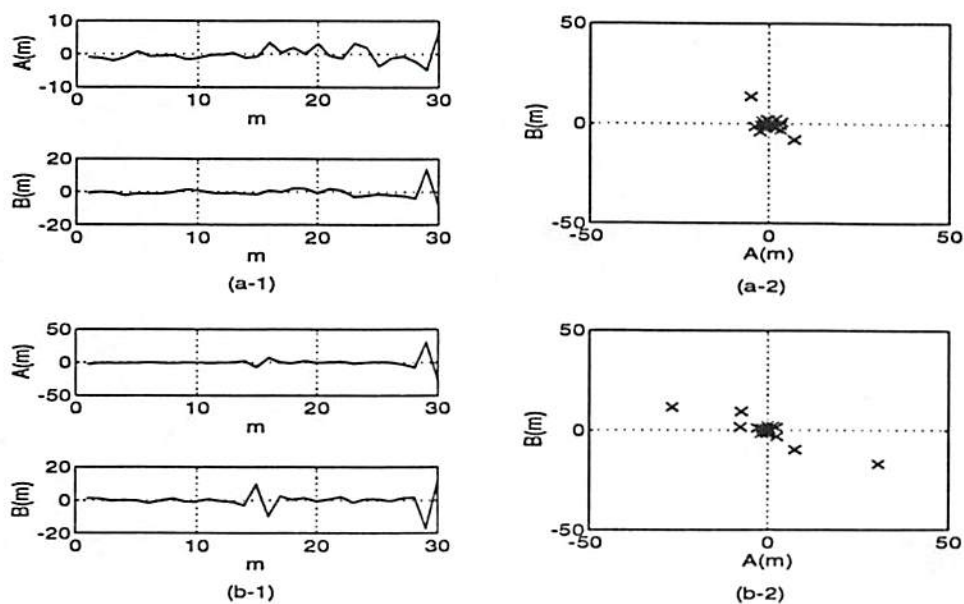


Figure 33: Using the first half of `trvhyd2`, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of `trvhyd2`, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.

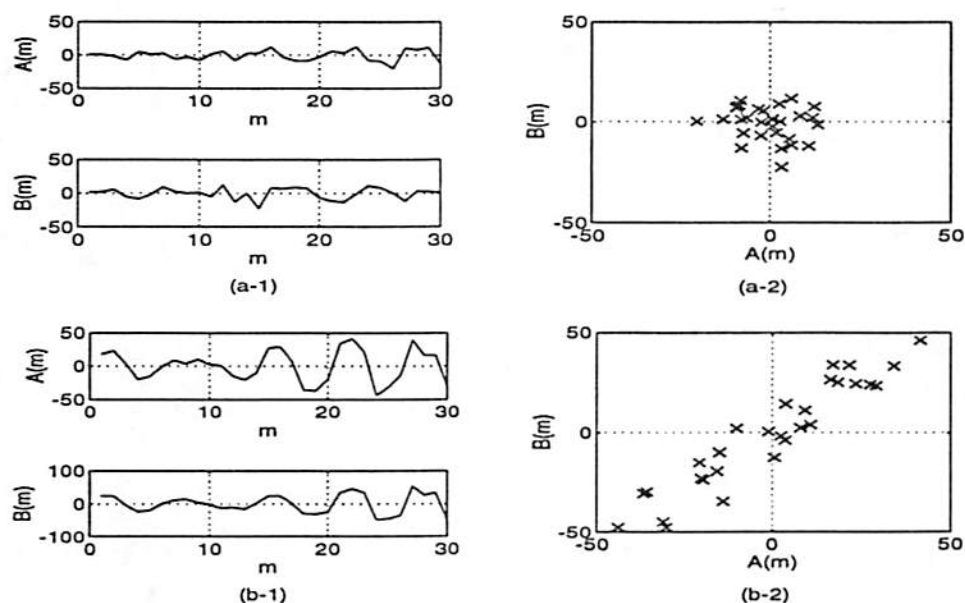


Figure 34: Using the first half of `trsvhyd2`, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of `trsvhyd2`, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.



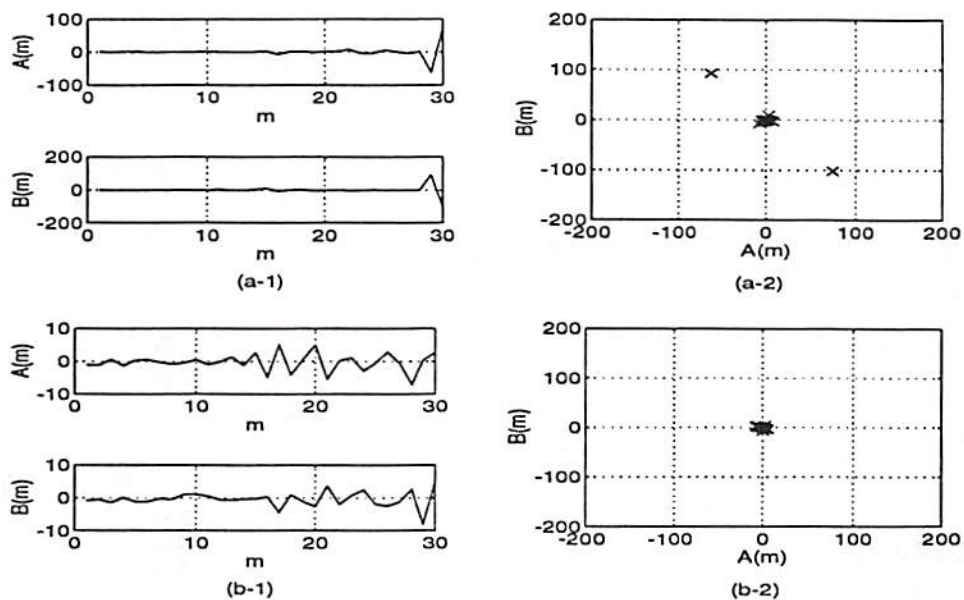


Figure 35: Using the first half of `trvhyd3`, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of `trvhyd3`, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.

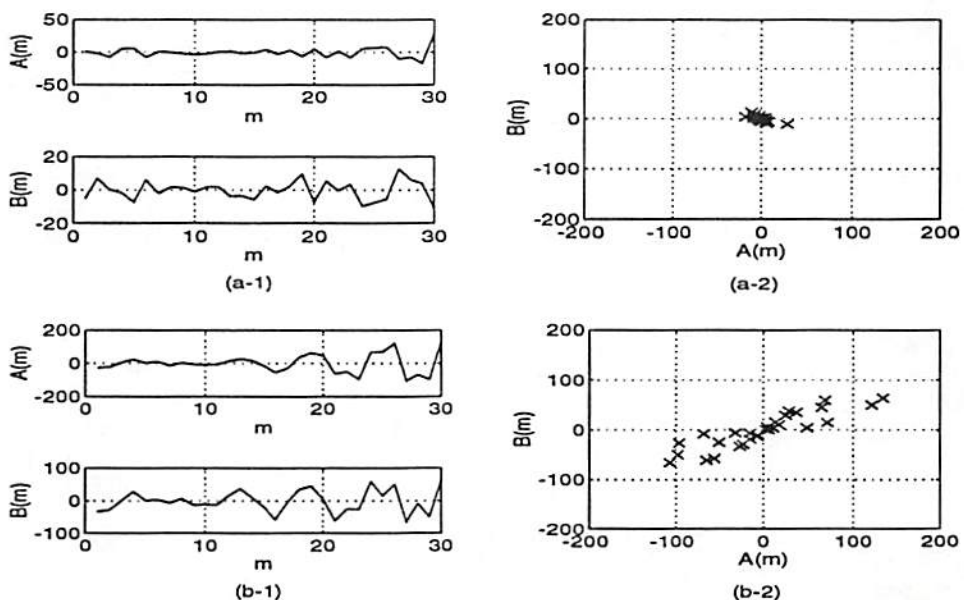


Figure 36: Using the first half of `trsvhyd3`, (a-1) cepstral parameters  $A(m)$  and  $B(m)$  and (a-2) their cluster; using the second half of `trsvhyd3`, (b-1) cepstral parameters  $A(m)$  and  $B(m)$  and (b-2) their cluster.



### 3.4 S $\alpha$ S Random Process Parameter Estimation

We applied the algorithms described in the section 2.4 on all the sets of real data. We could expect, however, that the results would be very poor due to the clear lack of outliers and the presence of strong dependence in the data.

As an example, we present here the estimates  $\hat{\alpha}$  for the data sets `trvhyd1` and `trsvhyd1`. Dividing each part of these data sets into  $L = 65$  segments of length  $M = 1,000$  data sample each, we estimated the value of the characteristic exponent  $\alpha$  as in Table 1. These values are highly unreliable and, therefore, we did not analyze the data with S $\alpha$ S models any further.

Table 1: Estimated Value of  $\alpha$  for the data sets `trvhyd1` and `trsvhyd1`

	first half of data	second half of data
<code>trvhyd1</code>	$\hat{\alpha} = 1.25$	$\hat{\alpha} = 2.0575$
<code>trsvhyd1</code>	$\hat{\alpha} = \text{NaN}$	$\hat{\alpha} = 4.56$

## 4 Concluding Remarks

In this report, we present the results from a number of signal processing approaches to analyzing hydroacoustic data measured in test internal launchway flows. The ultimate goal is to establish features of the data that allow their reliable classification into one of two categories, namely hydrophone data from a smooth section of the flow or from the section of the flow near a slot, as these are described in the report. We applied classification algorithms based on a variety of signal processing concepts, including modeling via second-order autoregressive equations, power spectrum- and polyspectra-based features, cepstral feature extraction based on third-order statistics, and modeling with symmetric, alpha-stable distributions. In our analysis, we examined the transient flows (first half of each data set) separately from the steady-state flows (second half of each data set) and for each one we established reliable classification features. In particular, we found that the coefficients of a second-order autoregressive model as estimated via Gaussian maximum likelihood methods are features providing clear classification power between smooth and slot section flow data. However, part of the classification power is lost if the autoregressive model estimates are obtained via Yule-Walker equations or their generalizations. It is noteworthy, that improvement was observed in the Yule-Walker approach when fractional, lower-order moments were employed. The power spectrum was another feature that allowed data classification, as was the phase of the bispectrum. In particular, it was observed that wide main lobes in the power spectrum or significant low frequencies in the phase of the bispectrum corresponded to flow data from the slot section, while narrow main lobes in the power spectrum or significant high frequencies in the phase of the bispectrum corresponded to flow data from the smooth section. On the other hand, the magnitude of the bispectrum did not provide clear classification power. Similarly, the cepstral parameters did not provide clear classification power either. As a conclusion, high classification power was provided by the Gaussian maximum likelihood estimates of the parameters of a second-order autoregressive model and the phase of the data bispectrum.

The analysis of the data using symmetric, alpha-stable distributions has not been complete. We studied the data ignoring the apparent high dependencies among them and applying techniques that are appropriate for independent data. A more appropriate approach would be to analyze the data using alpha-stable autoregressive models. This approach will be followed in the

Table 2: Classification Methods and Their Classification Power

SECOND-ORDER AR MODELING	CLASSIFICATION PERFORMANCE
Gaussian ML	High for both transient and steady-state data
Normal eq. based on second-order statistics	High for steady-state data only
Normal eq. based on third-order statistics	Low
Normal eq. based on lower-order statistics	High for both transient and steady-state data
POWER SPECTRUM	High for steady-state data only
MAGNITUDE OF THE BISPECTRUM	Low
PHASE OF THE BISPECTRUM	High
COMPLEX CEPSTRAL PARAMETERS	Low
I.I.D. ALPHA-STABLE MODELS	Low

future and the results of our findings will be reported then.



## References

- [1] T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York, 1971.
- [2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.
- [3] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1987.
- [4] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, PTR Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [5] C. L. Nikias and M. R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework," *Proc. IEEE*, vol. 75, no. 7, pp. 869-891, July 1987.
- [6] C. L. Nikias and M. Shao, *Signal Processing with  $\alpha$ -Stable Distributions and Applications*, John Wiley & Sons, Inc., New York, 1995.
- [7] R. Pan and C. L. Nikias, "The Complex Cepstrum of Higher Order Cumulants and Non-minimum Phase System Identification," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-36, no. 2, pp. 186-205, Feb. 1988.
- [8] M. Shao and C. L. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986-1010, July 1993.
- [9] G. A. Tsihrintzis and C. L. Nikias, "Fast Estimation of the Parameters of Alpha-Stable Impulsive Interference," submitted to *IEEE Trans. Signal Process.*, Aug. 1994; also available as USC-SIPI Technical Report # 270, Sept. 1994.