NASA CR-145090

# IMPROVEMENTS IN AIRCRAFT EXTRACTION PROGRAMS

By

A. V. Balakrishnan and Richard E. Maine

EXTRACTION PROGRAMS (Optimization Software,
Inc.)  96 p HC A05/MF A01          CSCL 01C

                                                   Unclas
                                      G3/05   56944

Prepared under Contract No. NASI-13824

By

OPTIMIZATION SOFTWARE, INC.
Los Angeles, CA  90024

for

# NASA

National Aeronautics and
Space Administration

DEC 1976
RECEIVED
NASA STI FACILITY
INPUT BRANCH

## ABSTRACT

Flight data from an F-8 corsair and a Cessna 172 have been analyzed to demonstrate specific improvements in the LRC parameter extraction computer program. The Cramer-Rao bounds (diagonal terms in the dispersion matrix) have been shown to provide a satisfactory relative measure of goodness of parameter estimates. It cannot be used as an absolute measure due to an inherent uncertainty within a multiplicative factor, traced in turn to the uncertainty in the 'noise' bandwidth in the statistical theory of parameter estimation. The measure is also derived on an entirely non-statistical basis, yielding thereby also an interpretation of the significance of off-diagonal (correlation) terms in the dispersion matrix. The distinction between coefficients as 'linear' and 'non-linear' is shown to be important in its implication to a recommended order of parameter iteration. Techniques of improving convergence generally, have also been developed, and tested out on flight data. In particular, an easily implemented modification incorporating a gradient search is shown to improve initial estimates and thus remove a common cause for lack of convergence. A close scrutiny of the 'maximum-likelihood' theory (which provides the basis for current extraction algorithms) indicating its limitations is an important by-product of this study. A technique of 'pooling' has been developed with demonstrated improvement in processing multiple maneuvers under similar flight conditions. A variety of questions that arise in interpreting computer results are also explicitly answered in the light of the theory developed.

# CONTENTS

## FIGURES

## FIGURES (cont'd)

iv

# TABLES

## LIST OF SYMBOLS

| | |
|---|---|
| A | (Stability) square matrix |
| $a_Y$ | lateral acceleration (g's) |
| B | control matrix |
| $B_i$ | bandwidth of ith noise component |
| C,D,E | observation matrices |
| $d_i$ | noise variance |
| $\hat{d}_i$ | estimated variance, ith component |
| $d_{ij}$ | components of the matrix $S(\theta)^{-1}$ |
| $e_k$ | orthonormal vectors |
| G | square root of diagonal spectral density matrix |
| $G(\theta)$ | gradient of cost functional |
| g | accleration due to gravity (ft/sec$^2$) |
| $g_i$ | square root of noise spectral density of ith component |
| $H_n$ | Hessian matrix |
| $I_X$ | moment of inertia about roll axis (slug-ft$^2$) |
| $I_{XZ}$ | cross product of inertia (slug-ft$^2$) |
| $I_Z$ | moment of inertia about yaw axis (slug-ft$^2$) |

## SYMBOLS (Cont'd)

$K$      default matrix (square)

$L$      weighting matrix

$L^*$      transpose of L

$L_o$      roll equation bias (deg/sec$^2$)

$L_p$      derivative of roll acceleration with respect to roll rate (/sec)

$L_r$      derivative of roll acceleration with respect to yaw rate (/sec)

$L_\beta$      derivative of roll acceleration with respect to sideslip angle (/sec$^2$)

$L_{\delta_a}$      derivative of roll acceleration with respect to aileron (/sec$^2$)

$L_{\delta_r}$      derivative of roll acceleration with respect to rudder (/sec$^2$)

$N_o$      yaw equation bias (deg/sec$^2$)

$N_i^j(t)$      ith noise component in jth run

$N_p$      derivative of yaw acceleration with respect to roll rate (/sec)

$N_r$      derivative of yaw acceleration with respect to yaw rate (/sec)

$N_\beta$      derivative of yaw acceleration with respect to sideslip (/sec$^2$)

$N_{\delta_a}$      derivative of yaw acceleration with respect to aileron (/sec$^2$)

$N_{\delta_r}$      derivative of yaw acceleration with respect to rudder (/sec$^2$)

## SYMBOLS (Cont'd)

$n_i(.)$      noise in the ith component of the observation

PSD      power spectral density

p      roll rate (deg/sec)

$p_i(f)$      spectral density as a function of frequency of the ith noise component

$q(\theta)$      cost functional

R      rotational transformation matrix

$R_i(0)$      covariance of ith noise component

r      yaw rate (deg/sec)

$S(\theta)$      sensitivity matrix

$S(\theta)_{jk}$      jkth entry in $S(\theta)$

$S(\theta)^{-1}$      dispersion matrix: (Cramer Rao Bound)

T      time history length - always in seconds

$T_j$      time history length of jth run (sec)

U      control vector

$V(t)$      response vector less noise

v      velocity (ft/sec)

## SYMBOLS (Cont'd)

| | |
|---|---|
| W(t) | Wiener process |
| X(t) | state vector |
| XAY | longitudinal distance from center of gravity to lateral accelerometer (ft) |
| XB | longitudinal distance from center of gravity to vane (angle) sensor |
| Y(t) | sensor measurement vector |
| $Y_i(t)$ | ith component of Y(.) |
| $\hat{Y}_i(t)$ | calculated ith component of Y(.) |
| $Y_o$ | sideslip equation bias (deg/sec) |
| $Y_\beta$ | derivative of sideslip rate with respect to sideslip angle (/sec) |
| $Y_{\delta_a}$ | derivative of sideslip rate with respect to aileron (/sec) |
| $Y_{\delta_r}$ | derivative of sideslip rate with respect to rudder (/sec) |
| Z(t) | residual - difference between observed and calculated response |
| z | perturbation vector |
| ZAY | vertical distance from center of gravity to lateral accelerometer (ft) |
| ZB | vertical distance from center of gravity to vane (angle) sensor |
| $\alpha_i$ | components of $\Theta$, the parameter vector |
| $\beta$ | sideslip angle (deg) |
| $\Lambda(t)$ | integral of observation vector |

# SYMBOLS (Cont'd)

$\gamma_r$          coefficient determining step size

$\Delta t$          sampling interval

$\delta_a$          aileron deflection (deg)

$\delta_r$          rudder deflection (deg)

$\Theta$          parameter vector

$\Theta_s$          starting parameter vector

$\hat{\Theta}_T$          estimate of $\Theta$ based on data of length T

$\phi$          bank angle (deg)

$\phi_o$          bank equation bias (deg/sec)

# 1. SUMMARY

Flight data from an F-8 corsair and a Cessna 172 have bee. analyzed to demonstrate specific improvements in the LRC parameter extraction computer program. The Cramer-Rao bounds (diagonal terms in the dispersion matrix) have been shown to provide a satisfactory relative measure of goodness of parameter estimates. It cannot be used as an absolute measure due to an inherent uncertainty within a multiplicative factor, traced in turn to the uncertainty in the 'noise' bandwidth in the statistical theory of parameter estimation. The measure is also derived on an entirely non-statistical basis, yielding thereby also an interpretation of the significance of off-diagonal (correlation) terms in the dispersion matrix. The distinction between coefficients as 'linear' and 'non-linear' is shown to be important in its implication to a recommended order of parameter iteration. Techniques of improving convergence generally, have also been developed, and tested out on flight data. In particular, an easily implemented modification incorporating a gradient search is shown to improve initial estimates and thus remove a common cause for lack of convergence. A close scrutiny of the 'maximum-likelihood' theory (which provides the basis for current extraction algorithms) indicating its limitations is an important by-product of this study. A technique of 'pooling' has been developed with demonstrated improvement in processing multiple maneuvers under similar flight conditions. A variety of questions that arise in interpreting computer results are also explicitly answered in the light of the theory developed.

## 2. INTRODUCTION

Parameter extraction from flight data has been recognized to be important for many purposes, for instance:

i) for comparison with wind-tunnel da ı

ii) for analytical and simulator studies for flight and handling qualities, and

iii) for application to adaptive control [1].

While various techniques for parameter extraction have been in usr for some time, it was not until the latter half of the 1960's that digital computer processing based on the modified Newton-Raphson algorithm made its advent [2], followed by accelerated activity along similar lines in the early seventies [3,4]. We are now entering what may be called the second phase of this effort, where we move from the many studies indicating feasibility of the technique to the implementation of the program on a routine day-to-day basis with minimal need for supervision by a specialist. Before this an be accomplished, many factors have to be ironed out; the most important question being the development of a calculable, satisfactory measure of the goodness - the reliability - of the extracted parameters, and its interpretation. A second and concommitant consideration is the development of a computer program that has the built-in ability to handle cases where the 'normal' algorithm fails to produce acceptable estimates. In addition, this would make it possible to determine if the data is so poor as not to warrant further processing and thus save the time and effort of computation.

This report is an attempt to improve the current Langley Research

Center (LRC) parameter extraction program. The study focuses on a number of questions arising from the use to date of that program and provides answers within the framework of two alternate theories: one statistical and other based on sensitivity. We begin in section 3 with the aircraft models - the equations of motion and the type of aircraft analyzed. The distinction between two types of parameters 'linear' and 'non-linear' is an important consideration in the basic computational procedure for parameter extraction, described in section 4, where the procedure per se is divorced from the rationale for using it. The performance of the procedure is then tested on flight data provided by LRC, using the Optimization Software (OSW) parameter extraction program. The scatter observed in the estimates obtained by maneuvers under similar flight conditions leads us to the primary question of the goodness of the estimates, and is examined in the next two sections. The statistical theory is developed in section 5, culminating in the Cramer-Rao (CR) bounds providing a statistical measure of uncertainty. The error due to the usually accepted practice of taking the (two-sided) bandwidth of the noise as equal to the sampling rate is explained here for the first time. Section 6 describes a non-statistical measure in terms of the largest possible variation in the estimates for a fixed percent change in the cost functional.

The question of how to take care of currently experienced difficulties in parameter extraction centering on convergence problems is taken up in section 7. A technique for pooling flight data obtained at identical flight conditions yielding estimates better than those from the individual runs is developed in section 8 and its performance evaluated. Questions arising in the use to date of parameter extraction programs are answered

3

in section 9, based on the theory developed in the previous sections. Concluding remarks in section 10 summarize some of the specific suggestions for possible program improvement.

## 3. THE AIRCRAFT MODEL

a.  Equation of Motion

We begin with the model – the linearized equations of airplane motion, lateral mode only – in state space form, and associated sensor measurements. Thus the state vector written as a column is

$$X = col \: [\beta, p, r, \phi]$$

The control vector is

$$U = col \: (\delta_a, \delta_r, 0, 0, 1)$$

The vector of sensor measurements is denoted Y:

$$Y = col. \: [\beta, p, r, \phi, a_Y]$$

The continuous time dynamic equations relating these quantities are

$$R \: \dot{X} = AX + BU$$

$$Y = CX + DU + E\dot{X} + N$$

where

$$
A = \begin{bmatrix}
Y_\beta & \sin(\alpha) & -\cos(\alpha) & \cos(\phi) \cos(\theta) \frac{g}{V} \\
L_\beta & L_p & L_r & 0 \\
N_\beta & N_p & N_r & 0 \\
0 & 1 & \cos(\phi) \tan(\theta) & 0
\end{bmatrix}
$$

5

$$B = \begin{bmatrix} Y_{\delta_a} & Y_{\delta_r} & 0 & 0 & \dot{V}_0 \\ L_{\delta_a} & L_{\delta_r} & 0 & 0 & L_0 \\ N_{\delta_a} & N_{\delta_r} & 0 & 0 & N_0 \\ 0 & 0 & 0 & 0 & \dot{\phi}_0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \dfrac{-I_{XZ}}{I_X} & 0 \\ 0 & \dfrac{-I_{XZ}}{I_Z} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & \dfrac{-ZB}{v} & \dfrac{XB}{v} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \dfrac{v}{g}Y_\beta & 0 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \dfrac{v}{g}Y_{\delta_a} & \dfrac{v}{g}Y_{\delta_r} & 0 & 0 & a_{Y\text{ BIAS}} \end{bmatrix}$$

6

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \dfrac{-ZAY}{57.3\ g} & \dfrac{XAY}{57.3\ g} & 0 \end{bmatrix}$$

and where N is the 'noise' and is the one really vague quantity in this description. As we shall see, it provides us with the 'rationale' for the procedure for our estimation technique but quantitative interpretation is beset with uncertainties. Generally, it is taken as white Gaussian or Gaussian of large band compared to airplane response but the precise description is a crucial point to which we shall return in section 5. It is assumed that the noise is independent from sensor to sensor.

The parameters to be identified are the various derivatives indicated by a subscript. The 'stability' derivatives are in the matrix A, the terms $\sin\alpha$, $\cos\alpha$, $\frac{g}{V}\cos\phi\ \cos\theta$, $\cos\phi\ \tan\theta$ are taken to be known constants. (In reality of course there may be a variation in time but this is taken to be small enough to be neglected. Actually there is no difficulty in accounting for the dependence on time if it is known.) The 'control' derivatives in the B matrix are in the first two columns. The parameters shown in the last column are 'bias' parameters and in themselves therefore have no physical significance and can, in particular, change even under identical flight conditions. A bias term also occurs in the matrix D and is described as "YBIAS" - this accounts for the average of the $(\sin \phi)$ $(\cos \theta)$ g/v term as well as any instrument bias in the $a_Y$ measurement.

7

b. Types of Aircraft Analyzed

All data presented in this study are confined to two

aircraft: A Cessna 172 and an F-8.

The Cessna 172 flight tests were conducted at NASA LRC. The Cessna

is a light high wing single engine general aviation air plane; relevant

physical characteristics are indicated below.


## CESSNA PHYSICAL CHARACTERISTICS

| | |
|---|---|
| Wing Area | 174 ft$^2$ |
| Span | 36.2 ft |
| Chord | 4.87 ft |
| Weight | 2200 lbs |
| $I_X$ | 872 slug-ft$^2$ |
| $I_Z$ | 1701 slug-ft$^2$ |
| $I_{XZ}$ | 14 slug-ft$^2$ |


In the data supplied, angle of attack and side-slip measurements were

corrected for instrument position by LRC. Accelerometer offset from the

C.G. was initially stated to be negligible. In the course of the OSW data

processing however, significant correlation between $a_y$ and $\dot{p}$ was detected

visually, indicating lateral accelerometer offset (above or below C.G.).
location
The accelerometer was then taken as an additional unknown parameter to be

determined by the OSW parameter extraction program. This yielded an

estimate of approximately 0.9 ft below C.G., agreeing well with the

measured position. Any parameter that can be measured independently and

accurately, should of course be modeled as known. Thus, the measured

position of the accelerometer should be included, rather than neglected

8

as unimportant. The accelerometer location did in fact have a significant effect on several of the minor coefficients and the r.m.s. residual fit on $a_y$ was approximately halved by correcting for the instrument location.

The second aircraft was a modified Navy F-8 Corsair 2. This particular aircraft was fitted with a super-critical wing for evaluation and was flown at NASA-Dryden. Physical characteristics are given below.

### F-8 PHYSICAL DATA

| | |
|---|---|
| Wing Area | 25.5 $M^2$ |
| Span | 13.14 M |
| Chord | 2.08 M |
| Weight | 10500 Newton |
| $I_X$ | 20500 Kg-$M^2$ |
| $I_Z$ | 140000 Kg-$M^2$ |
| $I_{XZ}$ | 4500 Kg-$M^2$ |

The angle of attack vane in this aircraft had a significant error (about 1°) that was not corrected in the data. Therefore, the $\sin\alpha$ term in the A-matrix was allowed to be an extra unknown.

Data from both aircraft was transmitted on a PCM link. For the F-8 the sampling rate was 25 samples a second. For the Cessna, the rate was only 10 samples a second, low enough to cause concern, but proved adequate nevertheless because of the good quality of the data (good resolution, low noise and accuracy of the linear model).

## 4. MAXIMUM LIKELIHOOD ESTIMATION PROCEDURE

### a. The Cost Functional

The 'maximum likelihood' (it is not strictly maximum likelihood in statistical terms as we shall indicate in section 5 below) technique of estimating the parameters is to minimize the expression:

$$\sum_{i-1}^{n} \text{Log } d_i + \sum_{i=1}^{n} \frac{1}{T} \int_0^T (Y_i(t) - \hat{Y}_i(\theta;t))^2 / d_i \; dt \qquad (4.1)$$

where

the subindex i denotes ith component of the observation vector, n being the size of the vector, 5 in the present case. $\hat{Y}(\theta,t)$ is the calculated observation vector using approximate parameter values and the known input and is thus a function of the parameter vector $\theta$.

$\{d_i\}$, i = 1,..n are non-negative constants, T is the available time-history.

The minimization is with respect to the parameter set

col. $[\theta, d_i, ..d_n]$

Pending further examination (sec. 5) we may accept minimizing (4.1) as a 'good thing' to do. Thus the second term in (4.1) is a 'mean-square error', 'weighted' by $\{d_i\}$. Moreover at the 'true value' of the parameters in $\theta$, the cost functional is a minimum. We can see this as follows. At a minimum, the gradient (the first partial derivatives) with respect to all the parameters must vanish. Thus we must have

$$d_i = \frac{1}{T} \int_0^T (Y_i(t) - \hat{Y}_i(\theta,t))^2 \; dt, \quad i = 1,..n \qquad (4.2)$$

and

10

$$\frac{\partial q(\theta)}{\partial \theta_j} = 0 = \frac{1}{T} \sum_{i=1}^{n} \frac{1}{\hat{d}_i} \int_0^T \left( \frac{\partial}{\partial \theta_j} \hat{Y}_i(\theta,t) \right) (Y_i(t) - \hat{Y}_i(\theta,t)) \, dt \qquad (4.3)$$

Because of (4.2), the minimization proceeds in two successive series of steps: setting

$$\hat{d}_i = \frac{1}{T} \int_0^T (Y_i(t) - \hat{Y}_i(\theta,t)^2 \, dt \qquad (4.4)$$

and then minimizing the 'cost functional':

$$q(\theta) = \frac{1}{T} \sum_{i=1}^{n} \int_0^T \frac{(Y_i(t) - \hat{Y}_i(\theta,\tau))^2 \, dt}{\hat{d}_i} \qquad (4.5)$$

with respect to $\theta$, keeping $\hat{d}_i$ fixed. Then calculate a new $\hat{d}_i$ at the minimum and repeat the minimization of (4.5). Note that (4.3) has the minimal (ideal) zero value at the true value of the parameters $\theta$. Of course we do not expect to see 'exact' zero. Note also that in this way at the 'minimum' (4.5) reduces to the value:

$$n,$$

again assuming we do not run across 'exact' zeros for $\hat{d}_i$. We call

$\hat{d}_i$ = the mean square fit error corresponding to the ith

measurement (4.6)

and

$$\sum_1^n \hat{d}_i = \text{total mean square fit error} \qquad (4.7)$$

Similarly we shall call

$$Z(t) = Y(t) - \hat{Y}(\hat{\theta},t) \quad 0 < t < T \qquad (4.8)$$

the residual, where $\hat{\theta}$ is the final parameter estimate.

11

The column vector with components

$$\frac{\partial q}{\partial \alpha_j} = \frac{1}{T} \sum_1^n \frac{1}{d_i} \int_0^T (\frac{\partial}{\partial \alpha_j} \hat{Y}_i (\theta,t))(\hat{Y}_i(t) - Y_i(\theta,t)) \; dt \qquad (4.9)$$

where $\alpha_i$ are the components of $\theta$, will be called the gradient (denoted $G(\theta)$) of $q(\theta)$, and the matrix $S(\theta)$ with components

$$S(\theta) = \{S(\theta)_{jk}\}$$

$$S(\theta)_{jk} = \frac{1}{T} \sum_1^n \frac{1}{d_i} \int_0^T \frac{\partial}{\partial \alpha_j} \hat{Y}_i (\theta,t) \frac{\partial}{\partial \alpha_k} \hat{Y}_i (\theta,t) \; dt \qquad (4.10)$$

will be called the 'sensitivity' matrix. It is recognized as the part of the Hessian of $q(\theta)$ which is independent of the data, and is moreover non-negative definite.

b.   Recommended Minimization Procedure

Since there are many ways of minimizing (4.5), we now describe a 'recommended' procedure. [It is more than a recommendation; it will be closely tied in with the theory in sections 5 and 6.] When we have no 'good' initial values for $\{d_i\}$, we take them all to be the same - not, in other words, favoring one measurement over another. Next we fix the parameters in the A matrix at their nominal starting values and minimize $q(\theta)$ with respect to the unknown parameters in B and D - we shall refer to the latter parameters as 'linear' parameters since they enter 'linearly' in $\hat{Y} (\theta,t)$. Moreover, derivatives of $q(\theta)$ with respect to these parameters of order higher than two vanish identically. As a result, the minimizing parameters are determined by:

$$\Theta_1 = \Theta_s - S(\Theta_s)^{-1} G(\Theta_s) \tag{4.11}$$

where $\Theta_s$ is the starting value of $\Theta$ and we simply set the derivatives with respect to the 'non-linear' parameter to zero in $G(\Theta)$ and $S(\Theta)$. This is because we have, by virtue of the fact that derivatives of order higher than two are zero:

$$G(\Theta_1) = G(\Theta_s) + S(\Theta_s)(\Theta_1 - \Theta_s) = 0, \text{ by } (4.11)$$

We wish to note now that if we rewrite (4.11) as

$$S(\Theta_s)[\Theta_1 - \Theta_s] = -G(\Theta_s) \tag{4.12}$$

then (4.12) which is equivalent to (4.11) has a solution even if $S(\Theta_s)$ is singular, and this solution, even though not unique, will still yield the minimizing $\Theta_1$. This can be seen as follows. Suppose for some non-zero vector h we have that

$$S(\Theta_s)h = 0 \tag{4.13}$$

where the components of h are zero corresponding to the parameters in the matrix A. Because the derivatives of order higher than two are zero, we have the Taylor expansion:

$$q(\Theta_s + h) = q(\Theta_s) + [G(\Theta_s),h] + \frac{1}{2}[S(\Theta_s)h,h] \tag{4.14}$$

where we have used the notation

$$[a,b] = \text{Tr } ab^* = \text{Tr } a^*b$$

Since $q(\Theta)$ is non-negative and

$$[S(\Theta_s)h,h] = 0,$$

it follows from (4.14) that

13

$$[G(\theta_s), h] = 0$$

also and that

$$q(\theta_s + kh) = q(\theta_s) \qquad (4.15)$$

for any scalar multiplier k. Let $\{\lambda_i\}$ denote the non-zero eigenvalues of $S(\theta_s)$ and $\{e_i\}$ the corresponding orthonormalized eigenvectors, in the space of parameter vectors whose components are zero corresponding to the parameters in the matrix A. Then $G(\theta_s)$ can be expressed in terms of the $\{e_i\}$ as

$$G(\theta_s) = \sum [e_i, G(\theta_s)] \, e_i$$

and hence

$$S(\theta_s) \, [\theta_1 - \theta_s] = -G(\theta_s)$$

has the solution

$$\theta_1 - \theta_s = - \sum \frac{[e_i, G(\theta_s)]}{\lambda_i} \, e_i \qquad (4.16)$$

This solution is not unique since we can add any h satisfying (4.13) but then the corresponding value of q(.) is the same by (4.15). Hence (4.12) has a solution and any technique used to 'solve' (4.12) should yield an acceptable $\theta_1$.

After this first iteration step, all parameters in $\theta$ are allowed to change and algorithm (4.12) is used. Since we are no longer confined to linear parameters, $S(\theta_n)$ is no longer the Hessian. Hence (4.12) is referred to as the 'modified' Newton-Raphson technique (Taylor and

and Iliff [2]). An element of the actual Hessian is

$$S(\Theta)_{j,k} - \left\{ \frac{2}{T} \int_0^T \sum_{i=1}^n \frac{1}{\hat{d}_i} \left( \frac{\partial^2}{\partial \alpha_j \, \partial \alpha_k} \hat{Y}_i (\Theta,t) \right) (Y_i(t) - \hat{Y}_i(\Theta,t)) \, dt \right\}$$

(4.17)

and the correction term is small if the residual is small. The mo^+ compelling reason for using only the first term is that the calcul of 2nd partial derivatives of $\hat{Y}(\Theta,t)$ can be quite tedious normally, and the improvement obtained would not be significant. This is because the N-R technique is itself efficient only "close" to the true parameter values, so that the residual must be small to begin with.

Note also that when allowing all parameters to change, the gradient is not necessarily zero at each step, because the 3rd and higher order derivatives are no longer zero. In fact the cost functional $q(\Theta)$ may not be monotone decreasing with each iteration if $S(\Theta_n)$ has eigenvalues close to zero. 'Normally' however these difficulties will not appear. How to deal with them when they do occur will be discussed further down below under improving convergence in section 7.

Illustration. The normal situation is illustrated by Cessna data: Run 11. Here the quantities $\{a_i\}$ were initially set at $d_1^{-1} = 9.37$, $d_2^{-1} = $ $= 3.55$, $d_3^{-1} = 3.64$, $d_4^{-1} = 4.28$, $d_5^{-1} = 8300$. The starting parameter values were:

| | | | |
|---|---|---|---|
| $C_Y = -.006$ | $C_{n_p} = -.1$ | $C_{Y_{\delta_a}} = 0$ | $C_{Y_{\delta_r}} = .001$ |
| $C_{\ell_\beta} = -.001$ | $C_{\ell_r} = .1$ | $C_{\ell_{\delta_a}} = -.0013$ | $C_{\ell_{\delta_r}} = .0002$ |
| $C_{N_\beta} = .00032$ | $C_{n_r} = -.1$ | $C_{N_{\delta_a}} = -.00005$ | $C_{N_{\delta_r}} = .0002$ |
| $C_{\ell_p} = -.5$ | | | |

The behavior of q(θ) by iteration is plotted in Figure 1. The cost functional at iteration 1 was obtained by varying 'linear parameters' only from the 'starting values' at iteration 0. At the suceeding iteration, all parameters were changed at each iteration using the modified N.R. algorithm. It is to be noted that the most significant reduction in the cost functional took place in the first iteration and that the functional levelled off by about the third. Tables 1 show the corresponding parameter values at each iteration, Table 2 the gradient, and Table 3 the r.m.s. residual of each sensor measurement.

The gradients of the non-linear terms were not computed for the first iteration, as only the linear terms were needed there. Note tne dramatic decrease in the gradient of the linear terms from iteration 0 (starting values) to iteration 1. This decrease is about 4-1/2 orders of magnitude. These numbers should theoretically be 0, but this decrease is well within expected numerical accuracy in solving a 8 dimensional system on the IBM in single precision (the 5 bias terms are not listed). Overall, this case exhibits ideal convergence characteristics.

c. Results: Cessna 172 Flight Data

Using the procedure outlined above, several runs of the Cessna 172 data were processed. The results obtained for 4 cases (2 aileron and 2 rudder) are summarized in Table 4. The starting values of $\{d_i\}$ (same as before) as well as the parameters were the same in all cases. The bias parameters are not tabulated.

Since the flight conditions were close enough to be identical, we would expect that the extracted parameter values should be the same.

Unfortunately this is not borne out by the data, even though the fit errors on all the runs are about the same and acceptably low enough. The most striking discrepancies are highlighted in Table 5.

Representative time history plots - actual and calculated - for a rudder run (case 10) and an aileron run (case 11) are shown in Figures 2 and 3 respectively. The fits are generally good, the only evident abnormality being due to the accelerometer location problem referred to earlier. This correction did not however affect the parameter discrepancies observed.

## 5.  ERROR BOUNDS: STATISTICAL THEORY

a.   Continuous Time Model

In this section we shall indicate a 'statistical' approach to the basic question of how can we tell how good our determination of the parameters is? The basis of this theory rests on the assumption that the limiting error in the observation Y(t) is 'random noise'. We begin with the 'most idealized' case: where we assume that the noise is temperature-limited black-body radiation - that the noise is 'Gaussian white'. Unfortunately, current fad in Stochastic Process theory requires us to be more pedantic as follows: we say

$$A(t) = \int_0^t Y(s)ds = \int_0^t V(s)ds + G\,W(t)$$

where W(t) is a full-rank Wiener Process and

$$G = \text{Diag } [g_1,..g_n]$$

$g_i^2$ corresponds to the spectral density of the ith noise component, or,

$$E \cdot [GW(t)(GW(t))^*] = \text{diag. } (g_1^2,..g_n^2) = G^2$$

where

$E \cdot$ stands for Expected Value, and

$$V(t) = Cx + Du + ER^{-1}(Ax + Bu)$$

In this formalism, we assume that the $\{g_i\}$ are known.  Then the logarithm of the likelihood functional becomes

$$= -\tfrac{1}{2}\int_0^T [G^{-2}\,\hat{Y}(\theta,t), \hat{Y}(\theta,t)]\,dt$$

$$+\int_0^T [G^{-2}\,\hat{Y}(\theta,t), d_i(t)] \qquad\qquad (5.1)$$

Note that (5.1) cannot be expressed in the usual 'least squares' form, (cf 4.5)). Nevertheless, the gradient of (5.1) is the same, since the term that is missing is of the form

$$\int_0^T [G^{-2} d\Lambda(t), d\Lambda(t)]$$

which of itself is meaningless in the Wiener process formalism. But (5.1) is a true likelihood functional and hence maximizing it would yield a 'maximum-likelihood' estimator. At the maximum the gradient of (5.1) vanishes:

$$G(\theta) = \left\{ \int_0^T \left[ G^{-2} \frac{\partial}{\partial \alpha_j} \hat{Y}(\theta,t), \hat{Y}(\theta,t)dt - d\Lambda(t) \right] \right\} = 0 \tag{5.2}$$

To solve (5.2) we use the modified N-R algorithm:

$$\theta_{n+1} = \theta_n - S(\theta_n)^{-1} G(\theta_n) \tag{5.3}$$

where

$$S(\theta_n) = \left\{ \int_0^T \left[ G^{-2} \frac{\partial}{\partial \alpha_j} \hat{Y}(\theta,t), \frac{\partial}{\partial \alpha_k} \hat{Y}(\theta,t) \right] dt \right\} \tag{5.4}$$

We can prove: (asymptotic consistency theorem)

Theorem 5.1  Suppose

$$\lim_{T \to \infty} \frac{1}{T} S(\theta) = \bar{S} \tag{5.5}$$

is positive definite in an open parameter set N containing the unknown point $\theta_0$. Then there exists a non-zero neighborhood at $\theta_0$ in which (5.2) has a root, for all T sufficiently large. Let $\hat{\theta}_T$ denote the root. Then

$$E[||\hat{\theta}_T - \theta_0||^2] \to 0 \tag{5.6}$$

and in fact

19

$$T \; E(||\hat{\theta}_T - \theta_0||^2) \to \text{Tr.} \, (\bar{S})^{-1}$$

$$T \, E||\, (\theta_T) - \theta_0|| \to 0 \qquad\qquad (5.7)$$

<u>Proof</u>  For a proof see Balakrishnan [5].

The main calculation is that

$$0 = G(\theta_0) + S(\theta_0) \, (\hat{\theta}_T - \theta_0) + \text{higher order terms} \qquad\qquad (5.8)$$

We also have the C-R bound for unbiased estimators:

$$\text{Var.} \; \hat{\theta}_T \geqslant S(\theta_0)^{-1}$$

In fact for large enough T we can use the approximation:

$$\hat{\theta}_T = \theta_0 - S(\theta_0)^{-1} \, G(\theta_0) + \cdots$$

where the second term is Gaussian with mean zero and variance

$$S(\theta_0)^{-1} \approx \frac{(\bar{S})^{-1}}{T}$$

Since $\theta_0$ is unknown, one usually calculates

$$S(\hat{\theta}_T)^{-1}$$

as an estimate of the variance.

<u>Limitations</u>

The main drawback in the above is the fact that the spectral density matrix

$$\text{diag} \, (g_1^2, \ldots g_n^2)$$

is assumed known. But this drawback is only in the calculation of the C-R bound. We can show that we may use any non-zero diagonal matrix in place of $\{g_i^2\}$ and still obtain asymptotic unbiasedness and asymptotic consistency of the parameters $\theta$.

## Band-Limited Noise Approach

We may make the more reasonable assumption that $N(t)$ is band-limited Gaussian, with bandwidth large compared to that of $\hat{Y}(\theta,t)$. Moreover we may then also consider the case where the noise power is unknown as well. We shall show we can then estimate the noise power as well but that uncertainty will again arise in the C-R bound due to lack of precise knowledge of the bandwidth.

We invoke the cost functional:

$$\sum_{i=1}^{n} \text{Log } d_i + \sum_{i=1}^{n} \frac{1}{T} \int_0^T \frac{\left(\hat{Y}_i(\theta,t) - Y_i(t)\right)^2}{d_i} \, dt \qquad (5.9)$$

which we minimize with respect to $\{d_i\}$ and $\theta$. As before (see sec. 4a) we take:

$$\hat{d}_i = \frac{1}{T} \int_0^T \left(\hat{Y}_i(\theta,t) - Y_i(t)\right)^2 \, dt \qquad (5.10)$$

and minimize

$$\sum_{i=1}^{n} \frac{1}{T} \int_0^T \frac{\left(\hat{Y}_i(\theta,t) - Y_i(t)\right)^2}{\hat{d}_i} \, dt \qquad (5.11)$$

Now (5.11) is not the log likelihood functional, even apart from the fact $\hat{d}_i$ is only an estimate for $d_i$. Nevertheless it can yield us an asymptotically unbiased and asymptotically consistent estimate for $0$. Denoting the gradient with respect to $\theta$ of (5.11) by

$$G(\theta) = \left\{ \sum_1^n \frac{1}{T} \int_0^T \left( \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\theta,t) \right) \frac{\left( \hat{Y}_i(\theta,t) - Y_i(t) \right) dt}{\hat{d}_i} \right\}$$

we have the Taylor Expansion, denoting the root by $\hat{\theta}_T$:

$$0 = G(\theta_o) + S(\theta_o) [\hat{\theta}_T - \theta_o] + \cdots \tag{5.12}$$

where $S(\theta)$ is the matrix:

$$S(\theta) = \left\{ \sum_1^n \frac{1}{T\hat{d}_i} \int_0^T \left( \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\theta,t) \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\theta,t) \right) dt \right\} \tag{5.13}$$

we have that

$$\hat{\theta}_T = \theta_o - S(\theta_o)^{-1} G(\theta_o) \tag{5.14}$$

where

$$G(\theta_o) = \left\{ \frac{1}{T} \sum_1^n \int_0^T \frac{\left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\theta_o,t) \right) n_i(t) dt}{\hat{d}_i} \right\} \tag{5.15}$$

$n_i(t)$ being the noise in the ith observation. Now

$$E[\hat{d}_i] = \frac{1}{T} \int_0^T E\left[ \left( \hat{Y}_i(\theta_o,t) - Y_i(t) \right)^2 \right] dt$$

$$= \frac{1}{T} \int_0^T E[n_i(t)^2] dt$$

$$= R_i(0)$$

where $R_i(t)$ is the covariance function of the noise process $n_i(t)$. Hence $\hat{d}_i$ is an unbiased estimate of the noise-variance. Replacing $\hat{d}_i$ by its expectation in (5.15), we have for the variance of $G(\theta_o)$

$$\left\{ \sum_1^n \overline{\frac{1}{R_i(0)^2}} \frac{1}{T^2} \int_{-\infty}^{\infty} \left( \int_0^T \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\Theta_o,t) \, e^{2\pi i f t} \, dt \right) \right.$$

$$\left. \left( \int_0^T \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\Theta_o,s) \, e^{-2\pi i f s} ds \right) P_i(f) df \right\} \qquad (5.16)$$

where $P_i(f)$ is the spectral density of the noise $n_i(t)$. If we take

$$P_i(f) = g_i^2 \quad \text{for } -B_i < f < B_i$$

$$= 0 \quad \text{otherwise}$$

and assume $B_i$ large, we can reduce (5.16) to:

$$\left\{ \sum_1^n \frac{1}{T^2} \frac{g_i^2}{R_i(0)^2} \left( \int_0^T \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\Theta_o,t) \, \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\Theta_o,t) \, dt \right) \right\}$$

where we note that

$$\frac{g_i^2}{R_i(0)} = \frac{1}{2B_i}$$

Hence within the approximation of replacing

$$\hat{d}_i \text{ by } E[\hat{d}_i] = R_i(0)$$

we have that

$$\text{var.}[\hat{\Theta}_T - \Theta_c]$$

$$= S(\Theta_o)^{-1} \left\{ \sum_{i=1}^n \frac{1}{T^2} \frac{1}{2B_i} \frac{1}{R_i(0)} \int_0^T \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\Theta_o,t) \, \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\Theta_o,t) dt \right) \right\} S(\Theta_o)^{-1}$$

which if we use the reasonable approximation

23

$$\frac{1}{2B_i} = \frac{1}{2B} \text{ for all } i,$$

we set

$$\text{var. } [\hat{\theta}_T - \theta_o]$$

$$= \frac{1}{2BT} S(\theta_o)^{-1}$$

$$= \left\{ \sum_1^n \frac{1}{g_i^2} \int_0^T \frac{\partial}{\partial \alpha_j} \hat{Y}(\theta,t) \; \frac{\partial}{\partial \alpha_k} \hat{Y}(\theta_o,t) \; dt \right\}^{-1} \qquad (5.17)$$

Note that again the use of this formula requires knowledge of the bandwidth:

$$g_i^2 = \hat{d}_i/2B \qquad (5.18)$$

Within the band-limited noise assumption we can also show then the variance of the estimate $\hat{d}_i$

$$= E ( \; [\hat{d}_i] - R_i(0))^2$$

$$\leq \frac{R_i(0)^2}{BT} \qquad (5.19)$$

b.   Discrete Time Theory

In the discrete-time theory, the data is sampled periodically at 2B samples per second, and it is assumed that the noise samples are independent. We have then the discrete version of (4.1)

$$\sum_1^n \log d_i + \sum_{i=1}^n \left( \frac{1}{N} \sum_{k=1}^N \frac{(\hat{Y}_i(\theta,k\Delta t) - Y_i(k\Delta t))^2}{d_i} \right) \qquad (5.20)$$

where

$$\Delta t = 1/2B$$

Taking the gradient yields

$$\hat{d}_i = \frac{1}{N} \sum_{k=1}^{N} (\hat{Y}_i(\theta, k\Delta t) - Y_i(k\Delta t))^2 \qquad (5.21)$$

and the parameters in $\theta$ are determined so as to minimize

$$\sum_{i=1}^{n} \left( \frac{1}{N} \sum_{k=1}^{N} \frac{(\hat{Y}_i(\theta, k\Delta t) - Y_i(k\Delta t))^2}{\hat{d}_i} \right) \qquad (5.22)$$

or use the algorithm

$$\theta_{n+1} = \theta_n - S(\theta_n)^{-1} G(\theta_n)$$

where

$$G(\theta_n) = \left\{ \sum_{1}^{n} \left( \frac{1}{N} \sum_{1}^{N} \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\theta, k\Delta t) \right) (\hat{Y}_i(\theta, k\Delta t) - Y_i(k\Delta t)) \right) \right\}$$

$$S(\theta_n) = \left\{ \sum_{1}^{n} \frac{1}{\hat{d}_i} \frac{1}{N} \sum_{1}^{N} \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\theta, k\Delta t) \right) \left( \frac{\partial}{\partial \alpha_m} \hat{Y}_i(\theta, k\Delta t) \right) \right\}$$

The estimate is again asymptotically unbiased and consistent. The variance of the estimate is

$$\left\{ \sum_{1}^{n} \frac{1}{\hat{d}_i} \sum_{1}^{N} \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\theta_o, k\Delta t) \right) \left( \frac{\partial}{\partial \alpha_m} \hat{Y}_i(\theta_o, k\Delta t) \right) \right\}^{-1} \qquad (5.23)$$

The difficulty with the need to know spectral density is somewhat hidden in the 'discrete-time' analysis: here the sampling rate is

$$2B$$

samples/sec and (5.13) is expressed as:

25

$$S(\theta) = \sum_{1}^{n} \frac{1}{\hat{d}_i} \sum_{1}^{N} \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\Theta_o, k\,\Delta t) \right) \left( \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\Theta_o, k\Delta t) \right)$$

which is

$$S(\theta) = \sum_{1}^{n} \sum_{1}^{N} \left( \frac{\partial}{\partial \alpha_j} \hat{Y}_i(\Theta_o, k\,\Delta t) \right) \left( \frac{\partial}{\partial \alpha_k} \hat{Y}_i(\Theta_o, k\Delta t) \right) \frac{1}{(2B)g_i^2}$$

But
$$\Delta t = \frac{1}{2B} \; ,$$

and hence this (assuming the sampling rate is adequate) yields the continuous time integral in (5.13). The assumption is made in this that

$$(g_i^2) = d_i/2B$$

and the error in this is that (2B) is overestimated so that $g_i^2$ is underestimated with the result that the C-R bound is also underestimated by a sizeable factor. Thus, the customary 50 samples/sec yields a bandwidth of 25 Hertz which in comparison with the actual observed bandwidth of the residual can be as much as ten times the actual. Any discrete-time (or sampled) theory requires that the noise samples be independent sample to sample which is less likely the higher the sampling rate [or assumed noise bandwidth].

c.  Illustration

The F-8 data furnishes a good example of the importance of the role played by the noise bandwidth in the calculation of the C-R bound. The four runs - 4, 5, 20 and 21 - are at essentially the same flight conditions listed in Table 6.

To provide a check on the calculations made with actual data, a simulated test was run first using the converged values from case 21 as the 'true' values, and the same control input as in case 21. Four different cases were computed with independent noise samples with the same total power as in run 21 but with (the one-sided) bandwidths set precisely to be $1/2\Delta t$, $\Delta t$ being the sampling interval. The OSW extraction program was then used to estimate the coefficients using each of the four simulation runs, yielding four independent estimates for each coefficient. The sample standard deviation $\hat{\sigma}$ calculated based on these four samples is then compared with the calculated C-R bound averaged over the four cases - the bounds were in fact very nearly the same on the four cases. The comparison is indicated in Table 7. The last column of this table lists the ratio of $\hat{\sigma}$ to the C-R bound (standard deviation). [Control derivatives are not included since only two estimates were available for these.] We see that the ratio is very close to unity, the agreement being excellent considering the small sample size.

The same comparison was then carried out using the F-8 flight data. The results are given in Table 8. The C-R bounds were calculated on the basis of a (one-sided) bandwidth of $1/2\Delta t$, just as in the simulated case, following current practice. In striking contrast to the simulated data, the ratio of $\hat{\sigma}$ to C-R bound (s.d) is now roughly of the order of 10.

27

This discrepancy is explained by our theory as being due to the actual noise bandwidth being much smaller than the arbitrary and incorrect specification of the bandwidth as $1/2\Delta t$.

Figures 4, 5, 6, 7, and 8 show the p.s.d. of the residuals for the F-8 flight data, based on which one may assess the true (one-sided) bandwidth at about 1 Hertz in contrast to $1/2\Delta t$ which is 12-1/2 Hertz.

Table 9 shows the C-R bounds calculated for the Cessna runs 9, 10, 11 and 12 based on a noise bandwidth $1/2\Delta t$. For psd's on the Cessna data see section 8.

## 6. ERROR BOUNDS: NON-STATISTICAL THEORY

It is possible to develop an interpretation of the accuracy of the extracted parameter values without invoking any statistical notions, and based entirely on the minimization of the functional (4.1). The functional itself can be interpreted without involving any notion of noise. Thus the second term of the cost functional is recognized as an "output fit error", normalized by the weighting parameter $\{d_i\}$. The latter can reflect our relative degree of confidence in each of the different instruments.

The uncertainty in the estimates $\hat{\theta}$, $\hat{d}_i$ may be evaluated in the following way: how much can we change $\hat{\theta}$, $\hat{d}_i$ keeping the cost functional within a fixed percent? In other words (looking at $\hat{\theta}$ for the time being), how 'large' can we make z:

$$\hat{\theta} \rightarrow \hat{\theta} + z$$

keeping: (f(.) denoting the cost functional used)

$$|f(\hat{\theta} + z) - f(\hat{\theta})| \leqslant c\ f(\hat{\theta})$$

where c is a fixed fraction (say 1%). The question how 'large' depends on the measure we wish to choose. In general, it may be measured by the square of a linear weighted sum, more compactly expressed:

$$[Lz, Lz] = ||Lz||^2$$

where $L$ is a given rectangular matrix. Moreover, since the changes are small, we may approximate the cost functional by retaining only the linear and quadratic terms:

$$f(\hat{\theta} + z) = f(\hat{\theta}) + [G(\hat{\theta}), z] + 1/2\ [S(\hat{\theta})z, z]$$

since $d_i$ is fixed (for the time being) $f(\theta) = q(\theta)$ and $f(\hat{\theta}) = n$; $G(\hat{\theta}) = 0$ so that

$$\frac{f(\hat{\theta} + z) - f(\hat{\theta})}{f(\hat{\theta})} = \frac{1}{2n}\ [S(\hat{\theta})z, z]$$

29

Hence we have the problem of maximizing

$$||Lz||^2$$

subject to

$$\frac{[S(\hat{\theta})z,z]}{2n} = c$$

This problem is readily solved by using Lagrange multiplier $\lambda$ and maximizing

$$[Lz,Lz] - \lambda \, [S(\hat{\theta})z,z] = [(L^*L - \lambda \, S(\hat{\theta}))z,z]$$

or, the optimal z will satisfy:

$$L^*Lz = \lambda \, S(\hat{\theta})z$$

and hence the answer to our problem is:

$$\max ||Lz||^2 = 2 \, nc \text{ (largest eigenvalue of: } S(\hat{\theta})^{-1} L^*L)$$

$$= 2 \, nc \text{ (smallest eigenvalue of: } (L^*L)^{-1} S(\hat{\theta})) \tag{6.1}$$

according as either $S(\theta)$ is non-singular, or $(L^*L)$ is non-singular. Also we may replace in (6.1):

$$S(\hat{\theta})^{-1} L^*L \quad \text{by} \quad \sqrt{L^*L} \, S(\hat{\theta})^{-1} \, \sqrt{L^*L}$$

$$(L^*L)^{-1} S(\hat{\theta}) \quad \text{by} \quad (\sqrt{L^*L})^{-1} \, S(\hat{\theta}) \, (\sqrt{L^*L})^{-1}$$

We shall now consider various special cases of L:

## Case 1

z is required to be of the form $ke_i$ where $e_i$ is a coordinate vector and k is a scalar. In other words we perturb one particular component

30

variable only.

$$\max ||Lz||^2 = 2n \ c \Big/ s_{ii} \quad \text{where } S(\hat{\theta}) = \{s_{ij}\}$$

Thus the smaller the diagonal elements in the 'sensitivity' matrix $S(\hat{\theta})$, the larger the uncertainty. The correlation in the $S(\hat{\theta})$ matrix plays no role.

But it does, as soon as we consider

### Case 2

$$||Lz||^2 = [e_i,z]^2$$

where $e_i$ is a coordinate vector with 1 for the ith component and zero elsewhere. Although the measure of uncertainty is based only on the ith component of $z$, all components of $z$ are allowed to change. In this case the answer is

$$\max ||Lz||^2 = (2n) \ c \ (d_{ii}) \qquad\qquad (6.2)$$

where

$$\{d_{ij}\} = (S(\hat{\theta}))^{-1}$$

Note that now the answer is different from case 1 as soon as $S(\hat{\theta})$ is not diagonal. In fact we note the elementary inequality that

$$d_{ii} \geqslant 1/s_{ii}$$

and equality holds for all i, if and only if $S(\hat{\theta})$ is diagonal. Note the connection with the statistical variance measure – we have in fact

31

the C-R bound but with an arbitrary constant of proportionality. For a similar approach, see Klein [6] and even earlier Shinbrot [7].

Let us next consider the case in which correlation in the $S(\theta)^{-1}$ matrix plays a role. Thus take

Case 3

$$\|Lz\|^2 = [e_i,z]^2 + [e_j,z]^2 \quad i \neq j$$

Let $\gamma$ be the i-jth correlation coefficient:

$$\gamma = d_{ij} \,\Big|\, \sqrt{d_{ii}\,d_{jj}}$$

Then for

$$\gamma = 0: \ \max \|Lz\|^2 = 2n\, c \ \max.\ (d_{ii},d_{jj})$$

$$\gamma = 1: \ \max \|Lz\|^2 = 2n\, c\ (d_{ii} + d_{jj})$$

It is interesting to note that the z that achieves the maximum (for $\gamma = \pm 1$) has the form

$$z_i = k_i$$

$$z_j = \pm k_i\ \sqrt{d_{jj}/d_{ii}}$$

In other words we may choose $z_i$ arbitrarily provided we also make $z_j$ a fixed multiple of $z_i$. Such a possible linear relationship has been noted in         reference [8]. Correlation makes the uncertainty worse.

Finally we may consider the case where we weight the $z_i$ conversely with respect to the $\hat{\theta}$ values $\{\alpha_i\}$:

Case 4

$$\|Lz\|^2 = \sum \frac{[z,e_i]^2}{\alpha_i^2}\ , \quad \hat{\theta} = \{\alpha_i\}$$

In this case

$$\max \ ||Lz||^2 = 2nc \quad \text{Largest eigenvalue of } \alpha \ S(\hat{\theta})^{-1}\alpha$$

$$= 2nc \ / \text{Smallest eigenvalue of } \alpha^{-1} \ S(\hat{\theta}) \ \alpha^{-1}$$

where

$$\alpha \ = \text{diag. } [1/\alpha_1, \ldots 1/\alpha_n].$$

Table 10 shows the results for the Cessna run 9 with a 1% cost change. The second and third columns show the calculation for cases 1 and 2; the maximal uncertainties are shown as a percent of the calculated $\hat{\theta}$ value. The final column shows the calculation for case 4; the values of the corresponding optimal z vector are indicated as a percent of the estimated parameter values. These values would appear to be rather high for only a 1% change in cost, and indicates in particular that run 9 was of poor quality. From C-R bounds given in Table 9, we see that the aileron runs are generally better than the rudder runs.

We may also study the uncertainty in the estimated $d_i$ values. The minimal value of (4.1) is given by:

$$\sum_1^n \ \log \ \overline{z_i^2} + n$$

where

$$\overline{z_i^2} = \frac{1}{T} \int_0^T \ (\hat{Y}_i(\hat{\theta},t) - Y_i(t)^2 \ dt$$

And accounting for both $\hat{\theta}$ and $\hat{d}_i$, we can write up to second order terms:

33

$$f(\hat{\theta} + z, \hat{d}_i + \gamma_i) - f(\hat{\theta}, \hat{d}_i) = \frac{[S(\theta)z, z]}{2} + \sum_1^n \frac{\gamma_i^2}{\hat{d}_i^2}$$

Hence

$$\max \gamma_i^2 = (2n\ c\ )\ \hat{d}_i^2$$

and this result is consistent with $\hat{d}_i$ being unbiased with variance proportional to $d_i^2$. The uncertainty in the $\hat{d}_i$ is proportional to $\hat{d}_i$.

## 7. CONVERGENCE IMPROVEMENT TECHNIQUES

The basic computational problem is the minimization of $q(\Theta)$ defined by (4.5). Let $\Theta_s$ denote the starting value of the parameters, including both the 'linear' and the 'non-linear', and let $\Theta_1$ denote the end of the first iteration using the N-R algorithm when only the linear terms are allowed to vary. The full non-linear terms are allowed to vary next using the N-R algorithm predicated on the assumption that the Hessian given by (4.17) is positive definite and that $\Theta_1$ is "sufficiently close" to the true [or minimizing] value $\Theta_T$. In fact, we know that $||\Theta_{n+1} - \hat{\Theta}_T|| \leq k \, ||\Theta_n - \hat{\Theta}_T||^2$ where k is a constant involving third derivatives of $q(.)$. It is difficult, if not impossible, to determine the closeness in any calculable quantitative way. On the other hand, if the closeness condition is not satisfied, the cost functional $q(\Theta)$ need not be monotone decreasing - in other words, $q(\Theta)$ oscillates and we experience lack of convergence.

In practice we can determine whether $\Theta_1$ is 'close enough' by running the program using the N-R algorithm. If there is lack of convergence and this cannot be traced to other sources, we should suspect that the starting values were not close enough. In other words, when non-convergence is due to poor starting values, the trouble can be with the N-R algorithm rather than any intrinsic defect of the maximum likelihood formulation.

The N-R algorithm should always be used on the 'linear' coefficients as already noted. Implementation of this is relatively easy. After computation of the gradient and the matrix S, a short routine can be inserted to reset the 'non-linear' terms in the gradient to zero. It

should also set the off diagonal elements in S corresponding to these terms also to zero, and the diagonal terms to 1. This is not very efficient in that it computes the non-linear terms and then ignores them. The computational effort is not large; however, since it applies only to the first iteration. On the other hand, it has the advantage that it is an easy, compact addition to any program, reducing to a one-line call to a short subroutine. More efficient implementation is also possible if the program is well modularized.

If the initial iteration of determining the linear terms is not adequate for convergence, we may use one of two methods to improve starting values: the  a) gradient method, or  b) the a'priori weighting method.

a)    The Gradient Technique

In the initial stages we may substitute the gradient technique in place of the N-R technique. In the gradient iteration we proceed as follows:

$$\theta_{n+1} = \theta_n + \gamma_n \; G(\theta_n)$$

where $\gamma_n$ is a numerical coefficient determining the 'step size'. The step size is not unique and can be chosen in many ways. In the 'steepest descent' version $\gamma_n$ is determined by taking the value of $\gamma_n$ corresponding to

$$\min_{\gamma} \; q(\theta_n + \gamma \; G(\theta_n))$$

If we omit terms higher than those of second degree in the expansion of $q(\theta)$, this would yield

36

$$\gamma = -\frac{[G(\theta_n), G(\theta_n)]}{[H_n G(\theta_n), G(\theta_n)]}$$

where $H_n$ is the Hessian at $\theta = \theta_n$. Since the determination of $H_n$ involves second derivatives, we may replace $H_n$ by $S(\theta_n)$, defined by (4.11). Alternately, fixed step sizes may be used (such as 0.6); or more time-consuming 'search' procedures may be employed.

## Performance

The use of the gradient for initial improvement will be illustrated by the Cessna 172 flight data. As starting values we use the following non-dimensional matrices, rounded off from PA-30 data:

$$A_N \quad \begin{matrix} -.006 & 0 & 0 & 0 \\ -.001 & -.5 & .1 & 0 \\ .0015 & -.1 & -.1 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

$$B_N \quad \begin{matrix} 0 & .001 & 0 & 0 & 0 \\ -.0013 & .0002 & 0 & 0 & 0 \\ -.00004 & .0002 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

The starting $d_i$ values were:

150

7

600

25

50000

None of the four cases 9, 10, 11 or 12 converged from these values at the first attempt using the N-R algorithm. Case 11 was then singled out for detailed study.

For the first iteration, only the 'linear' coefficients were determined as explained above. The cost function decreased from 13750 to 5511.

Then the gradient iteration (explained above) was used on all the coefficients. The cost functional values were: 5511, 4867, 4387, 3937, 3316, 2875, 2638, 2526, 2426, 2371, 2331. The coefficient values (of the 'non-linear' ones) at the final iteration were:

$$C_{Y_\beta} = -.005992$$
$$C_{\ell_\beta} = -.0006647$$
$$C_{N_\beta} = +.0006061$$

The other non-linear coefficients showed little change.

At this point we switched to the N-R technique (determining only the linear terms at the first iteration) and the corresponding cost functionals by iteration were: 4318, 1336 (linear only), 2634, 746, 199.6, 174, 174. The final coefficient values (non-linear) were:

$A_N$

| | | | |
|---|---|---|---|
| -.004650 | 0 | 0 | 0 |
| -.0006259 | -.2376 | .00758 | |
| .0003358 | -.05535 | -.07167 | 0 |
| 0 | 0 | 0 | 0 |

L.    'A Priori Weighting'

In this technique we modify the functional $q(\theta)$ by adding a positive definite quadratic form:

$$q'(\theta) = q(\theta) + [K(\theta - \theta_1), \theta - \theta_1]$$

where $K$ is a diagonal "default" matrix with positive entries corresponding to the non-linear parameters and zero otherwise. We may interpret this as assigning an a priori Gaussian density to the non-linear parameters and taking the 'unconditional' log likelihood function. The effect is basically to keep the search for the minimum in a chosen region. It has also the effect of making the Hessian positive definite for suitably chosen K.

After a few iterations using $K$, one then starts all over from the parameter values reached setting K to zero thereafter.

To demonstrate how this technique actually works out, we use the same Cessna case as above. A default matrix K was used: diagonal with non-zero entries as below:

| $C_{Y_\beta}$ | $C_{\ell_\beta}$ | $C_{n_\beta}$ | $C_{\ell_p}$ | $C_{n_p}$ | $C_{\ell_\delta}$ | $C_{n_\delta}$ |
|---|---|---|---|---|---|---|
| 1300 | .15 | 15 | 500 | 800 | 5 | 800 |

A multiplicative factor was left open. Runs were made with factors 1, 10, 100 and 1000 with no resulting convergence. Taking the factor as 100000, the following set of successive cost functionals obtained:
13800, 5600, 5500000, 143000, 29300, 24300, 8900.
The monotone behavior after the second was encouraging. In terms of the corresponding parameter values, $C_{n_\beta}$ overshot to negative values and

then regained slowly over the monotonic decreasing portion. The corresponding actual $C_{n_\beta}$ values were

.0015    .0015       -.000805

              -.0006847      -.0007170    -.0002629

The multiplying factor was then increased to $10^7$. The corresponding cost functional values became

13800    7800    62000    8160    34200    1347    417

The noticeable feature here is the smallness of the final cost functional, significantly lower than the starting value. The coefficient value at the last iteration indicated that $C_{n_\beta}$ was the only non-linear coefficient changing significantly from the starting value. Hence, the starting value of $C_{n_\beta}$ was changed from .0015 to .00032, the final value obtained. The cost functional then behaved satisfactorily:

3225    433    449    200    174    174    174

Moreover the coefficients obtained were identical to those obtained with the gradient technique.

It can be seen that the a priori weighting technique is very ad hoc in nature and is subject to the criticism that it is "massaging the data to get the answer you want."

## 8. POOLING TECHNIQUE

### a. Theory

When we have a set of runs all at or near the same flight conditions, we should, from the statistical noise theory point of view, 'pool' them in the following way: (as <u>opposed</u> to averaging the parameter estimates obtained independently from each run). Thus let us number the time-histories $Y^i(t)$, $i = 1,..m$ where the initial time is normalized to zero in each case so that we have m observation (vectors):

$$Y^i(t), \ 0 \leqslant t \leqslant T_i, \ i = 1,...m.$$

The main assumption is that the noise in each of the runs is statistically independent from run to run – this assumption is statistified if there is a time difference of a few seconds between the end of one and the beginning of the next – the runs are in practice obtained 'sequentially' in time anyway. The cost functional to be minimized is

$$\sum_{i=1}^{n} \log d_i$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \sum_{1}^{m} T_j \right)^{-1} \int_0^{T_j} \frac{\left( \hat{Y}_i^{\,j}(\theta,t) - Y_i^{\,j}(t) \right)^2}{d_i} \, dt \qquad (8.1)$$

where $\hat{Y}_i^{\,j}(\theta,t)$ is the calculated response for fixed parameters.

The form of (8.1) is derived from the fact the conditional probability:

$$P \, [Y^1, Y^2,.. Y^m \, | \theta]$$

$$= P \, [Y^1 | \theta].. \ P \, [Y^m | \theta]$$

which is in turn based on the independence of the noise processes from run to run. The noise variances from run to run are of course taken to be the same. The bias terms and the initial conditions are allowed to be dependent on the run. Only the aircraft stability and control derivatives are fixed for all the runs.

We have:

$$\hat{d}_i = \frac{\sum\limits_{1}^{m} \int_{0}^{T} (\hat{Y}_i^{\;j}(\theta,t) - \hat{Y}_i^{\;j}(t))^2 \, dt}{\sum\limits_{1}^{m} T_j} \tag{8.2}$$

and we minimize for fixed $\{\hat{d}_i\}$,

$$q(\theta) = \sum\limits_{1}^{n} . \sum\limits_{1}^{m} \frac{1}{\sum\limits_{1}^{m} T_j} \int_{0}^{T_j} \left( (\hat{Y}_i^{\;j}(\theta,t) - \hat{Y}_i^{\;j}(t))^2 \Big/ \hat{d}_i \right) dt \tag{8.3}$$

with respect to all the other unknown parameters. We have again the modified N-R algorithm:

$$\theta_{n+1} = \theta_n - S(\theta_n)^{-1} G(\theta_n)$$

where

$\theta$ is of course a vector of the form:

$\theta = \text{col} \; [\alpha_i^j, \alpha_k] \quad j = 1,..m$

$\qquad\qquad\qquad\qquad i = 1,..p, \text{ say}$

$\qquad\qquad\qquad\qquad k = 1, ...r, \text{ say}$

The double-indexed parameters are the bias terms and initial conditions which enter linearly and are allowed to vary from run to run. The gradient $G(\Theta)$ is the column vector of partial derivatives:

$$G(\Theta) = \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{\sum_{1}^{m} T_j} \int_{0}^{T_j} \left( \frac{\partial}{\partial \alpha} \hat{Y}_i^j(\Theta,t) \right) \cdot \frac{(\hat{Y}_i^j(\Theta,t) - \hat{Y}_i^j(t)) \, dt}{\hat{d}_i} \right\} \quad (8.4)$$

where $\alpha$ denotes the generic parameter. Similarly $S(\Theta)$ is the matrix

$$S(\Theta) = \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{\sum_{1}^{m} T_j} \int_{0}^{T_j} \frac{\frac{\partial}{\partial \alpha} \hat{Y}_i^j(\Theta,t) \, \frac{\partial}{\partial \beta} \hat{Y}_i^j(\Theta,t) \, dt}{\hat{d}_i} \right\} \quad (8.5)$$

where $\alpha$, $\beta$ stand for the parameters.

The main question that remains is the C-R bound for the aircraft parameters. We have the expansion, $\Theta_o$ denoting the true total parameter vector set:

$$\hat{\Theta} - \Theta_o = - S(\Theta_o)^{-1} \, G(\Theta_o) \quad (8.6)$$

from which we calculate that

$$E\left( (\hat{\Theta} - \Theta_o)\,(\hat{\Theta} - \Theta_o)^* \right)$$

$$= S(\Theta_o)^{-1} \left\{ \sum_{1}^{n} \sum_{1}^{m} \frac{(g_i)^2}{\sum_{1}^{m} T_j^2} \int_{0}^{T_j} \frac{\left( \frac{\partial}{\partial \alpha} \hat{Y}_i^j(\Theta_o,t) \right)\left( \frac{\partial}{\partial \beta} \hat{Y}_i^j(\Theta_o,t) \right) \, dt}{(d_i)^2} \right\} \cdot$$

$$S(\Theta_o)^{-1} \quad (8.7)$$

43

where

$$(g_i)^2 = \text{spectral density of } N_i^{\,j}(t) = d_i/2B_i$$

This is a little complicated and can be simplified further if we assume that the bandwidths are the same:

$$B_i = B$$

and of course the variances are the same:

$$E\,[N_i^{\,j}(t)^2] = d_i$$

In that case

$$E\,[\hat{d}_i] = d_i$$

Hence replacing $\hat{d}_i$ by $d_i$ in the formula (8.5) for $S(\theta)$, the expression (8.7) simplifies to

$$S(\theta_o)^{-1}\left\{\sum_1^n \sum_1^m \frac{g_i^{\,2}}{\left(\sum_1^m T_j\right)^2} \int_0^{T_j} \frac{\frac{\partial}{\partial\alpha}\hat{Y}_i^{\,j}(\theta_o,t)\,\frac{\partial}{\partial\beta}\hat{Y}_i^{\,j}(\theta_o,t)dt}{d_i^{\,2}}\right\} S(\theta_o)^{-1}$$

$$= \left\{\sum_1^n \sum_1^m \frac{1}{g_i^{\,2}} \int_0^{T_j} \frac{\partial}{\partial\alpha}\hat{Y}_i^{\,j}(\theta_o,t)\,\frac{\partial}{\partial\beta}\hat{Y}_i^{\,j}(\theta_o,t)dt\right\}^{-1} \qquad (8.8)$$

where we emphasize that

$$(2B)\,g_i^{\,2} = d_i$$

The "pooled" C.R. bound is thus the inverse of

$$\left\{ \sum_{1}^{m} \left( \sum_{1}^{n} \frac{1}{g_i^2} \int_0^{T_j} \frac{\partial}{\partial \alpha} \hat{Y}_i^j(\theta_o, t) \; \frac{\partial}{\partial \beta} \hat{Y}_i^j(\theta_o, t) dt \right) \right\}$$

which is also the inverse of

$$= \sum_{1}^{m} S_j \qquad\qquad (8.9)$$

where $S_j$ are the individual sensitivity matrices. Note that if we took as the composite the average of the m determinations, the variance would be

$$\frac{1}{m^2} \sum_{1}^{m} (S_j)^{-1} \qquad\qquad (8.10)$$

and of course this variance would be larger:

$$\frac{1}{m^2} \sum_{1}^{m} S_j^{-1} \geqslant \left( \sum_{1}^{m} S_j \right)^{-1} \qquad\qquad (8.11)$$

the improvement being bigger, the farther apart the individual matrices $S_j$ are. Note also that there is the advantage that (8.9) will tend to be more non-singular than the individual matrices.

b.  Performance

Performance of this technique was tested on the Cessna 172 flight data. The rudder input cases (9 and 10) yield poorer results than the aileron cases (11 and 12). Since the flight conditions are very close, it seems natural to pool the rudder and aileron cases. Thus, run 9 was pooled with run 11 (and designated 9-11) and run 10 was pooled with run 12 (designated 10-12). The estimates of the coefficients and the C-R bounds are presented

45

in Table 11. The fits are shown in Figures 9-12 and the residuals in Figures 13-16. We note first that there is a significant reduction in the scatter of the estimates from the individual runs. The agreement between the two pooled runs is also quite good. The C-R bounds are slightly better than for the individual aileron runs and significantly better than the individual rudder runs. These bounds were calculated on the basis of a (one-sided) bandwidth of $1/2\Delta t$ (5 Hertz).

Figures 17-21 show the power spectral density of the residuals. It is based on the first 512 time point of the combined residuals from all four cases. Based on these plots, a bandwidth of around 1 Hertz or less would be reasonable. This would indicate that the C-R bounds should be 2 or 3 times larger at least. The resulting error estimates are consistent with the estimate scatter.

## 9. QUESTIONS AND ANSWERS

In this section we shall provide explicit answers to the questions raised by LRC in the Statement of Work. The theory leading to the answers has been presented in sections 4 through 8 in this report and will be drawn upon as needed.

Question 1 - Does non-uniqueness of derivatives always occur with high correlation?

Answer - With $S(\theta)$ as defined in this report, one may refer to high correlation in $S(\theta)$ or, as LRC suggests, in its inverse - the dispersion matrix. Uncertainty in derivative extraction can be interpreted either with the statistical theory (section 4) or the non-statistical theory (section 5). In either case we have shown that the diagonal elements in the dispersion matrix provide a direct measure of this uncertainty and is higher, the higher the correlation in the $S(\theta)$ matrix. Also, correlation in $S(\theta)^{-1}$ has an additional effect of increasing this uncertainty. In this sense, high correlation is indicative of non-uniqueness of derivatives; see under case (3) of section 6.

Question 2 - Can we tell which parameters are 'observable'? a) What statistics or what part of the program can we interrogate to find out? b) Why won't the variances tell us if a parameter is not observable?

Answer - The observability of a parameter is measured by the corresponding diagonal entry in the dispersion matrix $S(\theta)^{-1}$. The interpretation of this can be statistical or non-statistical. However, there is uncertainty in this within a multiplicative factor owing to the uncertainty in the noise bandwidth (see section 4). Hence, an 'absolute'

47

(as opposed to relative) quantitative use of the variance can be misleading (and may be inconsistent with observed scatter).

Question 3 - When two parameters are highly correlated, we can vary either parameter over a wide range, provided we compensate by changing the other parameter. Since we can vary the parameters in this manner, why do we converge on one value for each parameter, and still have indicated low variance for each parameter?

Answer - By 'parameter correlation' what is meant here is the correlation in the dispersion matrix $S^{-1}$. We have shown in section 5 that when such correlation occurs we may change either parameter linearly with respect to the other with little or no change in the cost functional being minimized. However, how much the parameter can be changed in this manner is still determined by the sum of the corresponding variances, as indicated in section 5.

Question 4 - Under what conditions do correlations of parameters occur? Is it caused by correlations of states? Is there some other reason?

Answer - Again, by 'correlation of parameters' is meant the correlation in the $S^{-1}$ matrix. In the statistical theory, this correlation is actually the correlation in the error covariance and not in the parameters themselves. The parameters are not conceived as random variables. In any event, the correlation in the dispersion matrix is not due to correlation in the states, and does not have any direct interpretation other than as indicating 'stiffness' in the dispersion matrix (large eigenvalue spread). On the other hand, correlation in the matrix $S(\theta)$ would indicate closeness

48

to singularity and hence largeness of the C-R bounds. Also, the 'two-by-two' correlation by itself need not be large and yet sub-determinants may be zero or close to zero. Hence, singling out two-by-two correlations for attention does not appear to be of much direct relevance.

Question 5 - What does 'correlation' between parameters and/or states really mean?

Answer - Correlation between states is apparently interpreted as in the following type of situation :

$$p = k\beta$$

(the time history $p(t)$ is a constant multiple of $\beta(t)$). More generally, if $x(t)$ represents the state, then we may consider

$$[v,x(t)] = 0 \qquad 0 < t < T$$

for some non-zero vector v. Presumably this is satisfied at the true parameter values. But this has nothing to do with the dispersion matrix since the latter is determined by partial derivatives with respect to the parameters. In other words, it does not follow, for example, that the sensitivity matrix $S(\Theta)$ is singular.

Question 6 - If correlation between $C_{n_\beta}$ and $C_{n_r}$ is 1:1, will an incremental change in $C_{n_\beta}$ produce the same effect as an identical incremental change in $C_{n_r}$ ? If correlation between two derivatives is less than 1.0, what does it mean?

Answer - Continuing to interpret correlation as again that in the dispersion matrix, it is clear first of all that correlation of $\pm 1$ cannot occur, since otherwise the matrix would be singular. On the other hand, if the correlation is merely "close to one" then the implication is only that the matrix is 'stiff'. If the correlation in the S matrix is taken, however, occurrence of exactly $\pm 1$ would mean that it has a zero eigenvalue and hence if $\hat{\theta}_T$ is the optimal estate,

$$S(\hat{\theta}_T) = S(\hat{\theta}_T + \lambda e)$$

where e is the eigenvector corresponding to the zero eigenvalue and hence also

$$q(\hat{\theta}_T) = q(\hat{\theta}_T + \lambda e)$$

up to the second order approximation. Thus, if the 'correlation between $C_{n_\beta}$ and $C_{n_r}$ is $+1$' is interpreted to mean that the corresponding correlation in the S matrix is 1, then the eigenvector e has zero entries except corresponding to the $C_{n_\beta}$, $C_{n_r}$ places where it is $\frac{1}{\sqrt{2}}$, $-\frac{1}{\sqrt{2}}$ and we may keep the cost functional the same by proportionately changing $C_{n_\beta}$ and $C_{n_r}$. This statement continues to be approximately true if the correlation is sufficiently close to $\pm 1$.

Question 7 - Is it possible to specify flight-test techniques that will minimize correlations and maximize observability? Is there any analytical basis for determining best surface to use, and best control-input time history to minimize correlations? Will control to minimize correlations also maximize sensitivity parameters, or will minimizing correlations also minimize state sensitivity to the parameters.

Answer - It is quite possible to require that the input u(.) be such as to make the correlations in the S matrix or the dispersion matrix to be small. However, the only real analytical basis for determining the basic surface to use will be to require that the trace of the dispersion matrix be minimized. The optimal solution corresponding to this criterion will have a smaller trace than the case where the correlations in the S matrix are zero.

Minimizing correlations in the dispersion matrix will mean very little if the trace is unaffected - see section 5, case 3.

## 10. CONCLUDING REMARKS

By way of conclusion, some of the specific recommendations for improving current parameter extraction programs will now be itemized.

1.   Using approximate noise variances where available or otherwise making them all the same, the first iteration should vary only the 'linear' coefficients. Using the residuals as estimates for the noise variances, all the parameters are allowed to vary from then on, until convergence is obtained. The new set of residuals is then used to repeat the above procedure until the residuals stabilize.

2.   If convergence difficulties arise - or even otherwise routinely - after the 'linear' first iteration is completed, it is recommended that the gradient technique described be employed until the gradient stabilizes and then the switch to the N-R algorithm be made. The a priori weighting technique is too subjective and ad hoc and is not recommended.

3.   The proper measure of uncertainty or observability of the parameters is provided by the diagonal terms in the dispersion matrix. However, there is some danger in using this as an "absolute" measure rather than a 'relative' measure because it will always contain an uncertain multiplicative factor.

4.   The p.s.d's of the residuals may be used to estimate actual noise bandwidth.

5.   Where multiple maneuvers at identical or similar flight conditions are available, the 'pooling technique' should be used in contrast to averaging the estimates from the individual maneuvers. It is particularly helpful to pair the aileron-input data with the rudder-input

52

data so as to improve the estimates since the latter generally turn out to be worse. [A study of this phenomenon, verifying whether the aileron input alwayr yields better results than the rudder input and if so, what the reasons are, should be of value - and would shed much light on the 'optimal input' problem.]

6. Caution is necessary in using the dispersion matrix at the end of the first 'linear' iteration as a measure of the data since the matrix may well be non-singular and acceptable at the starting values and yet singular at the true values.

7. Minimizing correlations in the dispersion matrix is of little value - minimizing its trace is more meaningful.

# LIST OF REFERENCES

1.  Kurzhals, P. R., "NASA Advanced Control Technology: An Overview," NASA Conference on "Advanced Control Technology and Its Potential for Future Transport Aircraft," Los Angeles, July 1974.

2.  Taylor, L. W. and Iliff, K. W., "A Modified Newton-Raphson Method for Determining Stability Derivatives from Flight Data," Conference on Computing Methods in Optimization Problems, San Remo, Italy, September 9-13, 1968.

3.  "Parameter Estimation Techniques and Applications in Aircraft Flight Testing," NASA Symposium, April 24-25, 1973, NASA TN-D-7647.

4.  Suit, W. T., "Aerodynamic Parameters of the Navion Airplane Extracted from Flight Data," NASA TN-D-6643, 1972.

5.  Balakrishnan, A. V., "Techniques of System Identification," Lecture Notes, University of Rome, Rome, Italy, 1968.

6.  Klein, V., "On the Adequate Model for Aircraft Parameter Estimation," Cranfield Report, Aero no. 28, March 1975.

7.  Shinobrot, M., "An Analysis of the Errors in Curve Fitting Problems with an Application to the Calculation of Stability Derivatives from Flight Data," NASA TN 2820.

8.  Parrish, R.V. and G. G. Steinmetz, "Lateral Stability and Control Derivatives of a Jet Fighter Airplane from Flight Data," NASA TN D-6905, Sept. 1972.

FIGURE 1
COST FUNCTIONAL BY ITERATION
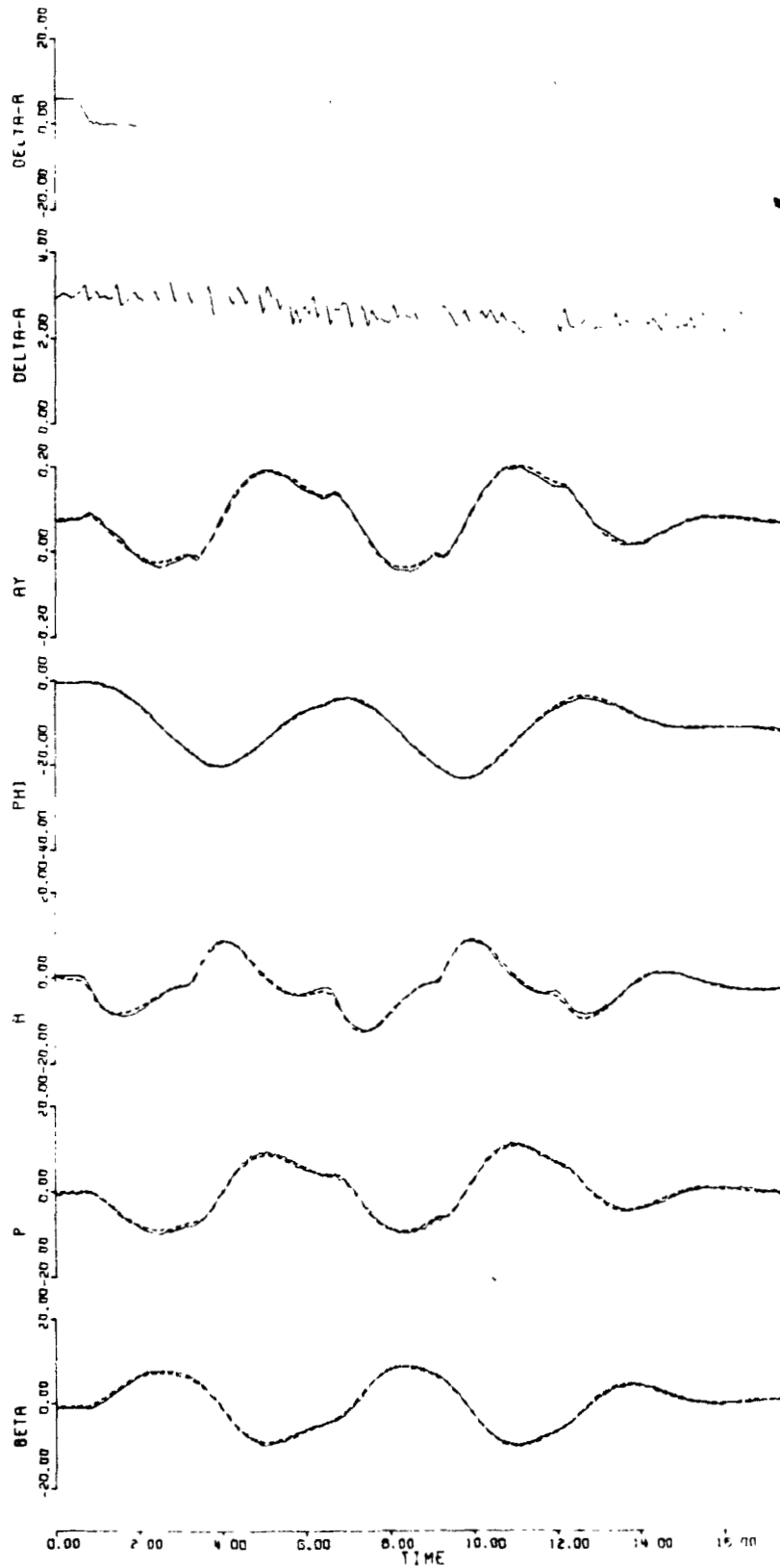C E S S N A   #  11

NOISE VAR. FIXED

LINEAR TERMS ONLY  1ST ITERATION

ITERATION
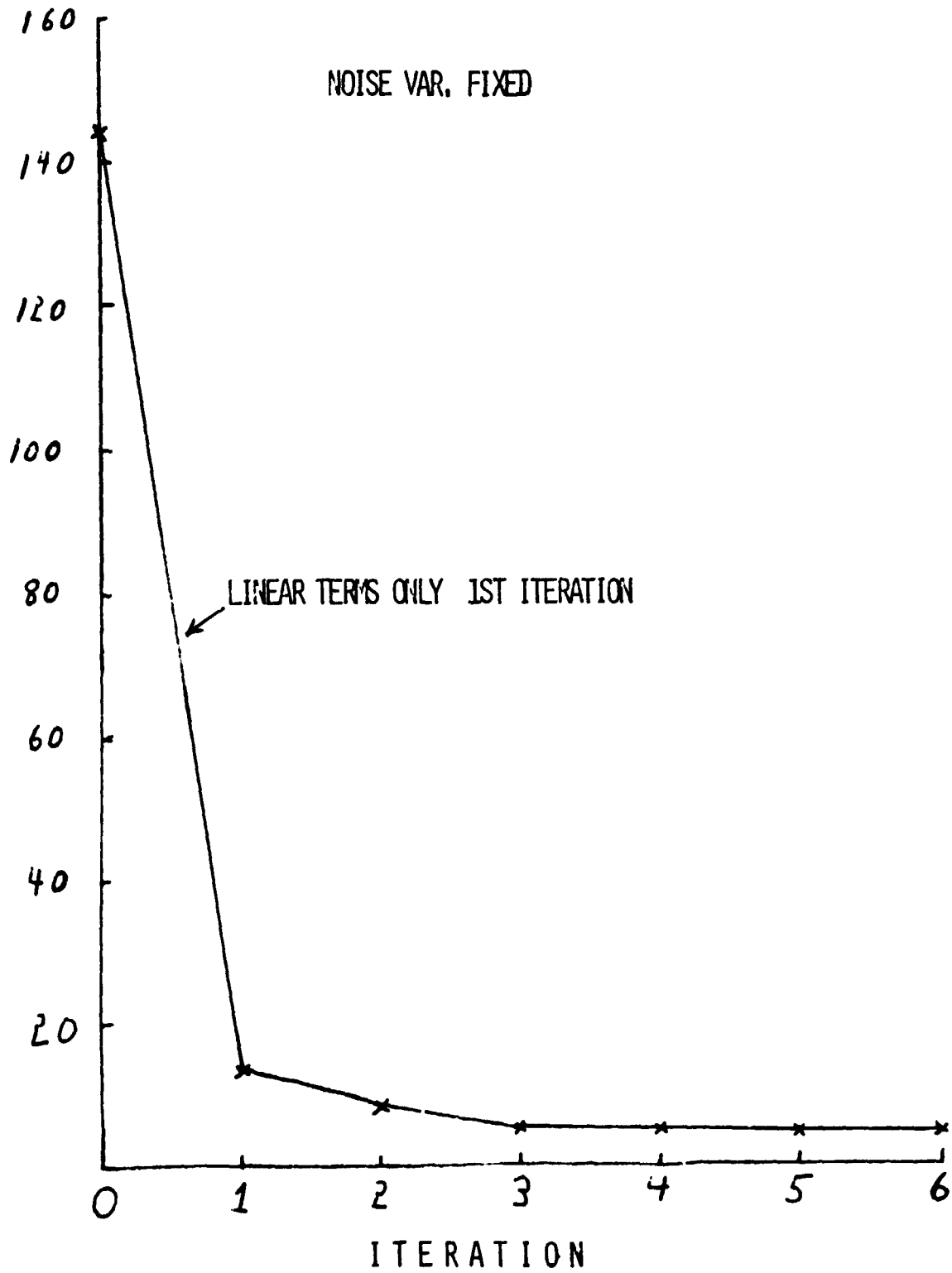
FIG. 2.   TIME HISTORY FIT : CESSNA RUN 10
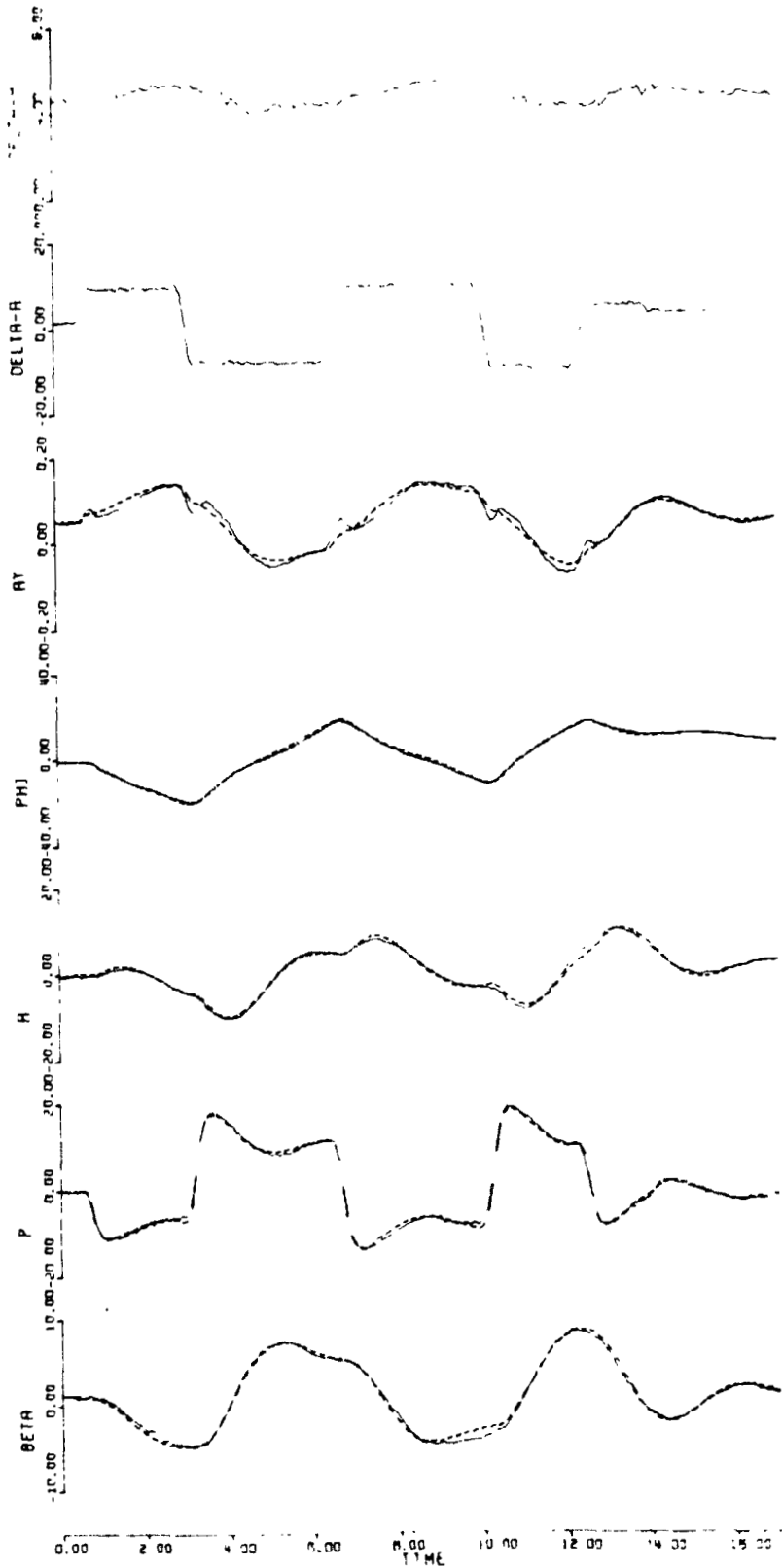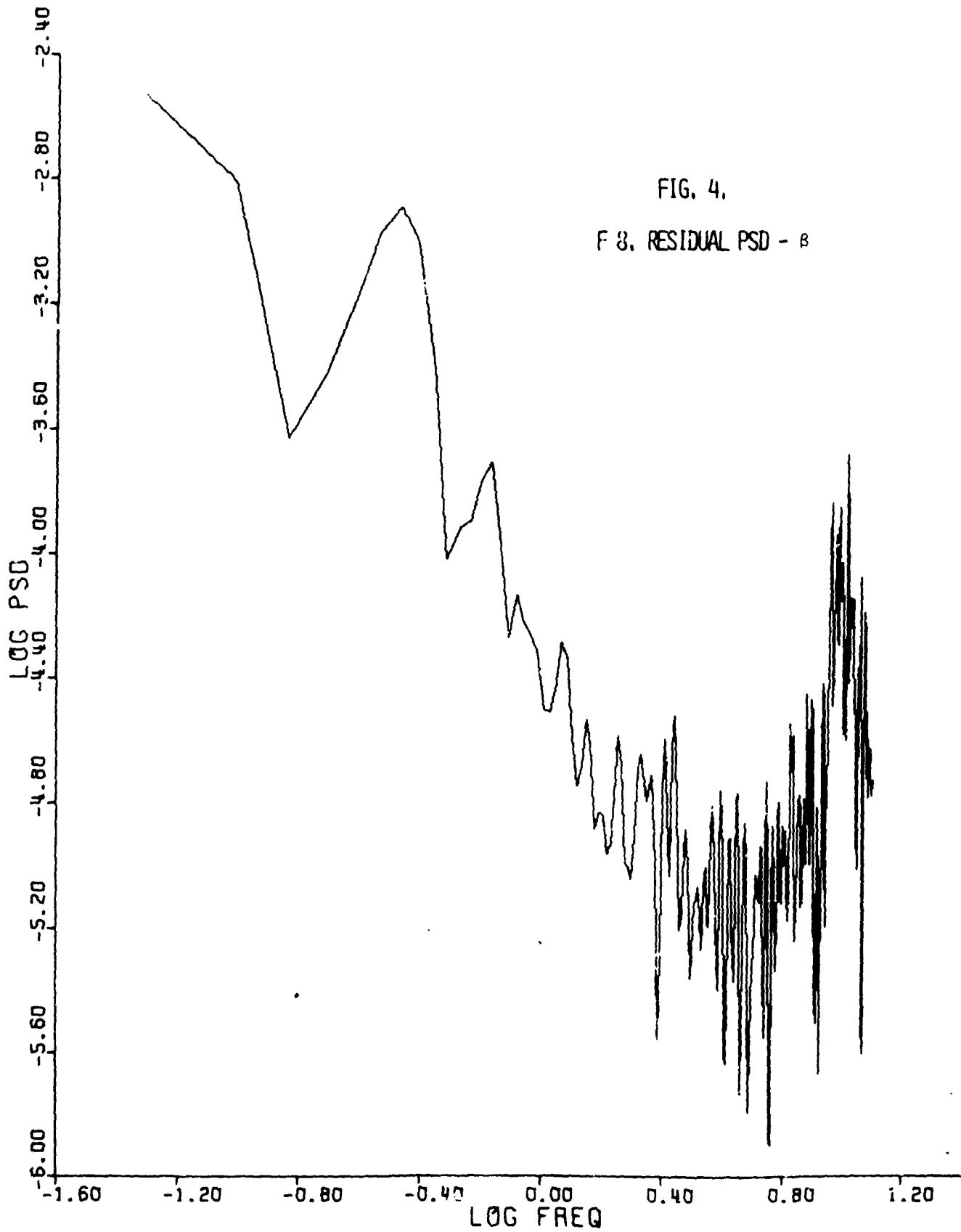
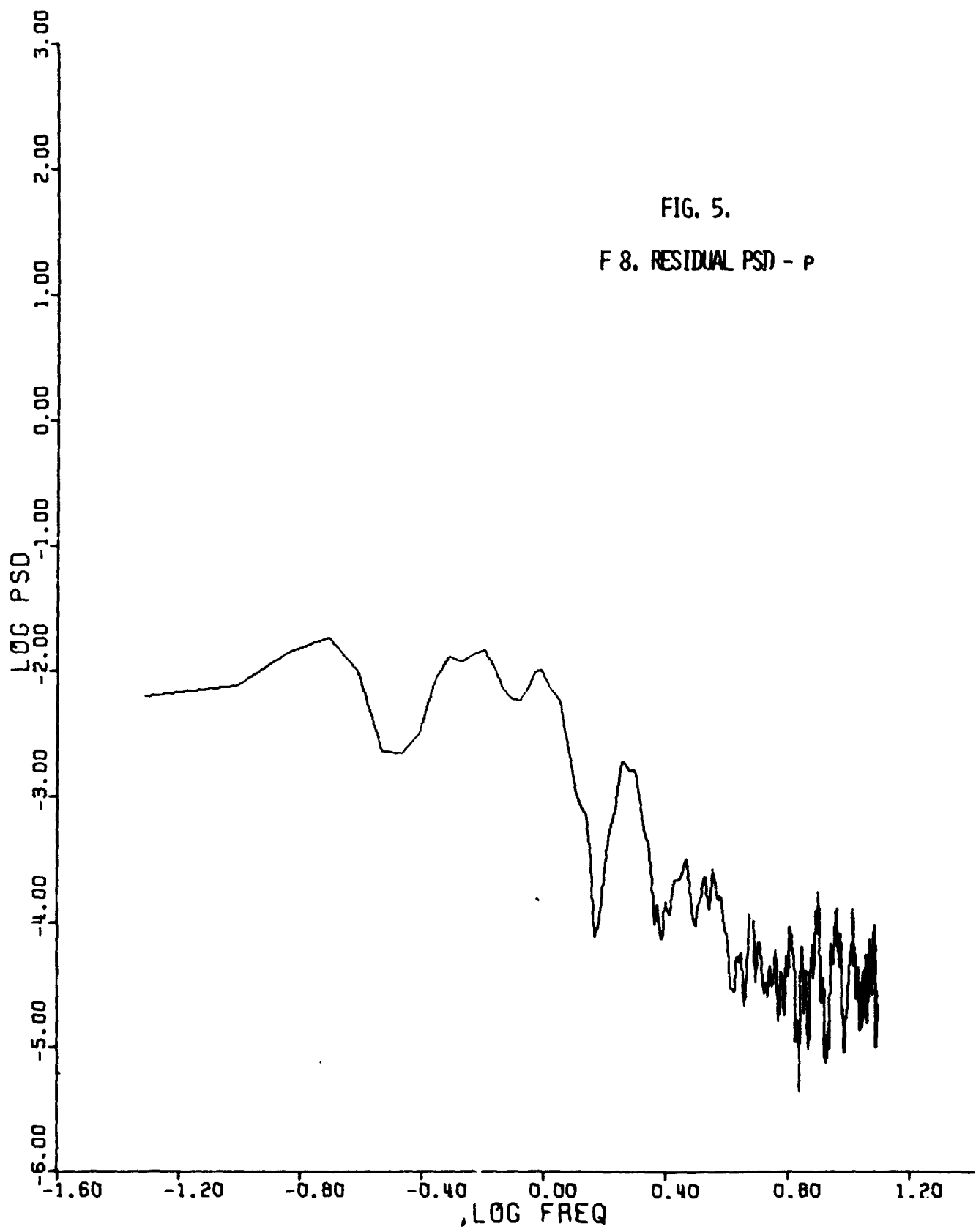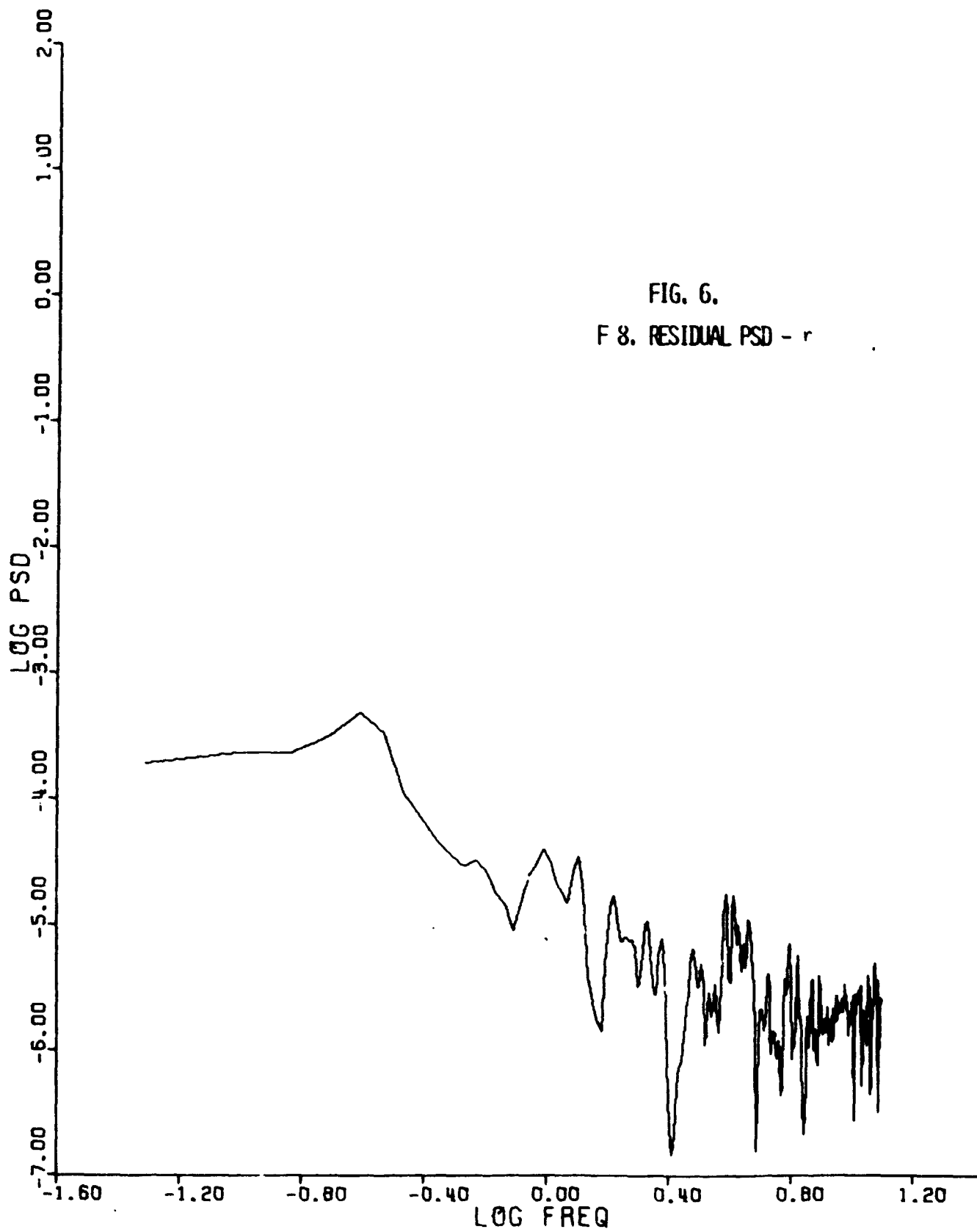FIG. 3.   TIME HISTORY FIT : CESSNA RUN II

57

FIG. 4.

F 8. RESIDUAL PSD - β

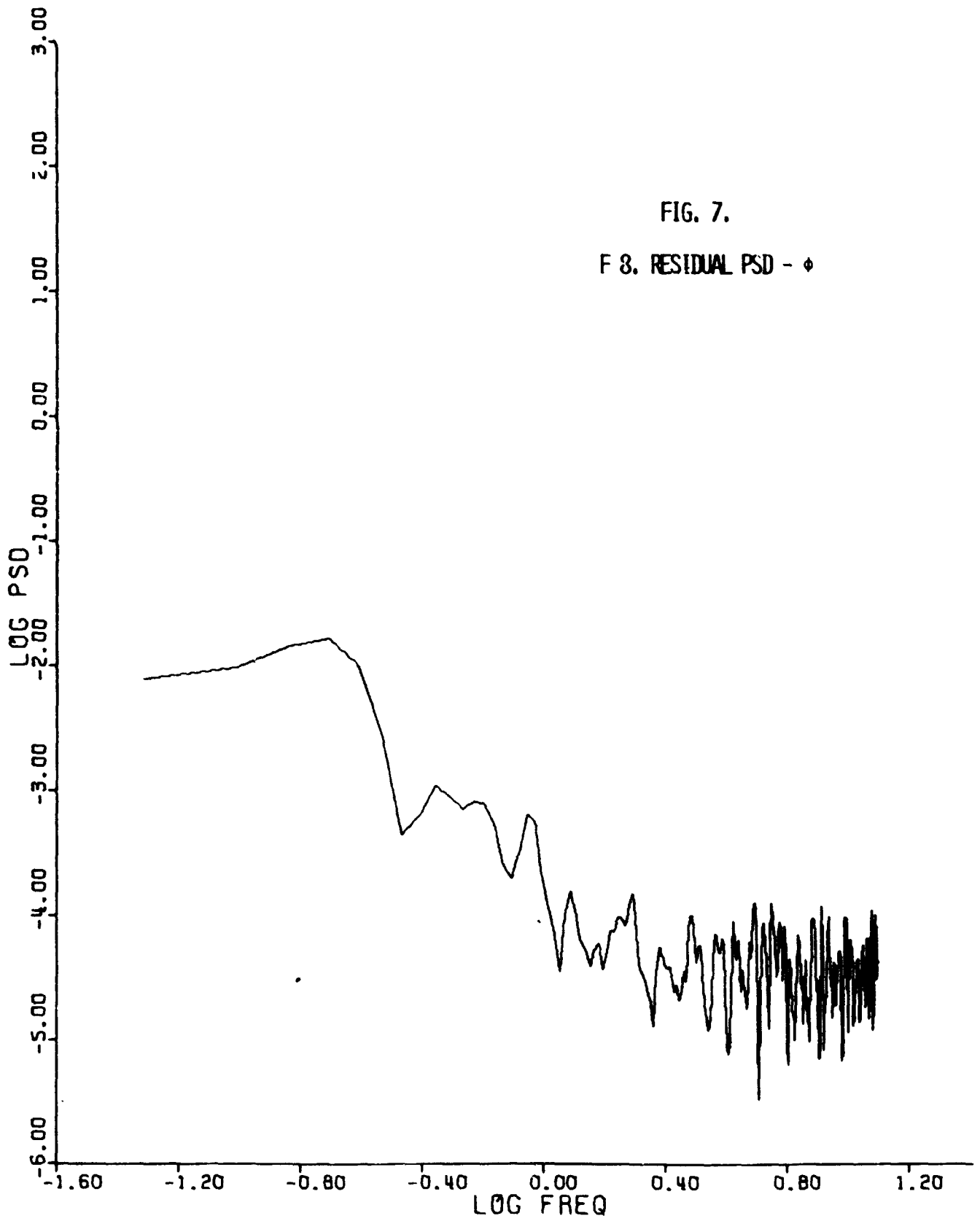FIG. 5.

F 8. RESIDUAL PSD - P

FIG. 6.

F 8. RESIDUAL PSD - r

FIG. 7.

F 8. RESIDUAL PSD - φ

LOG PSD

LOG FREQ

φ
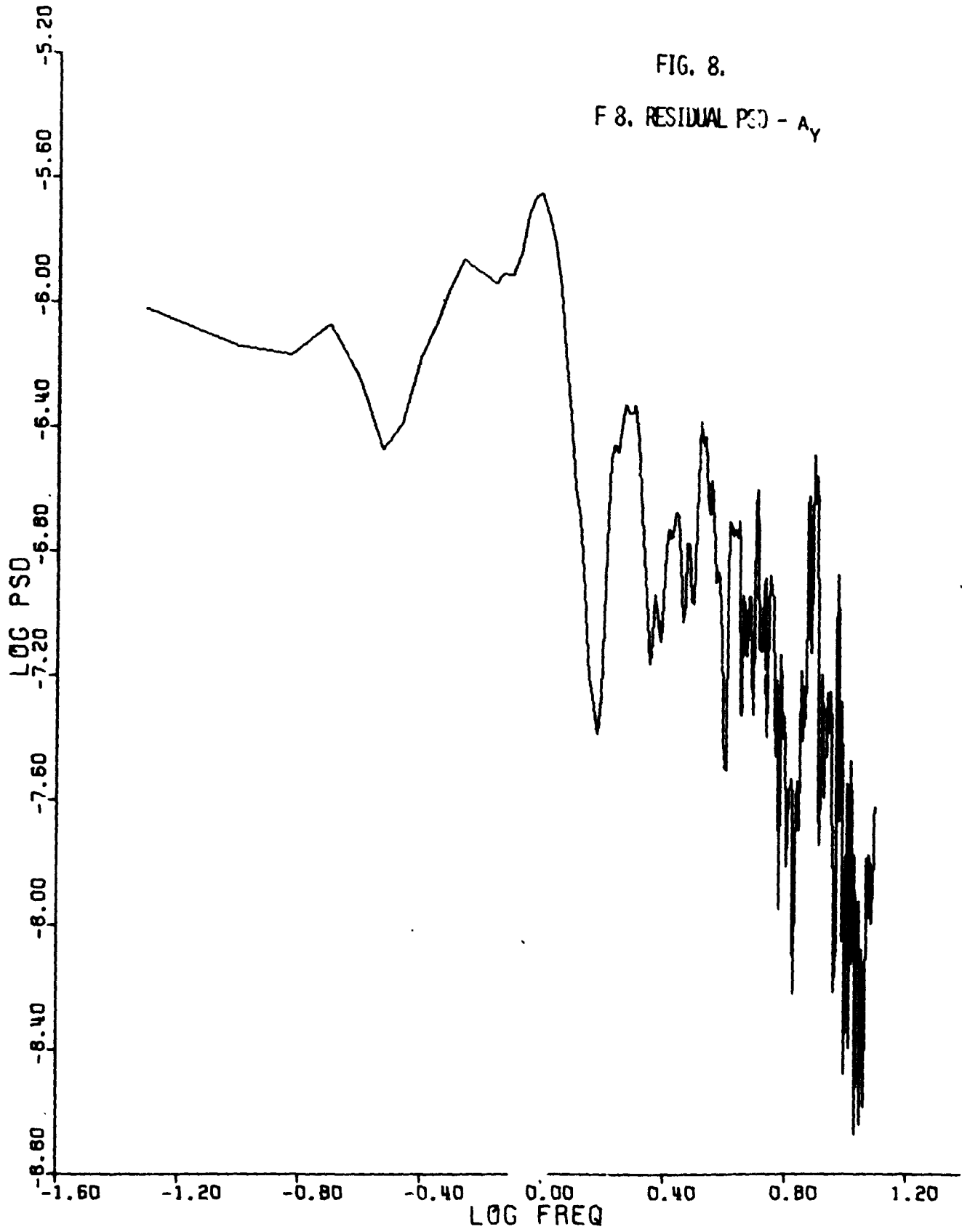
FIG. 8.

F 8. RESIDUAL PSD – $A_Y$

LOG PSD

LOG FREQ
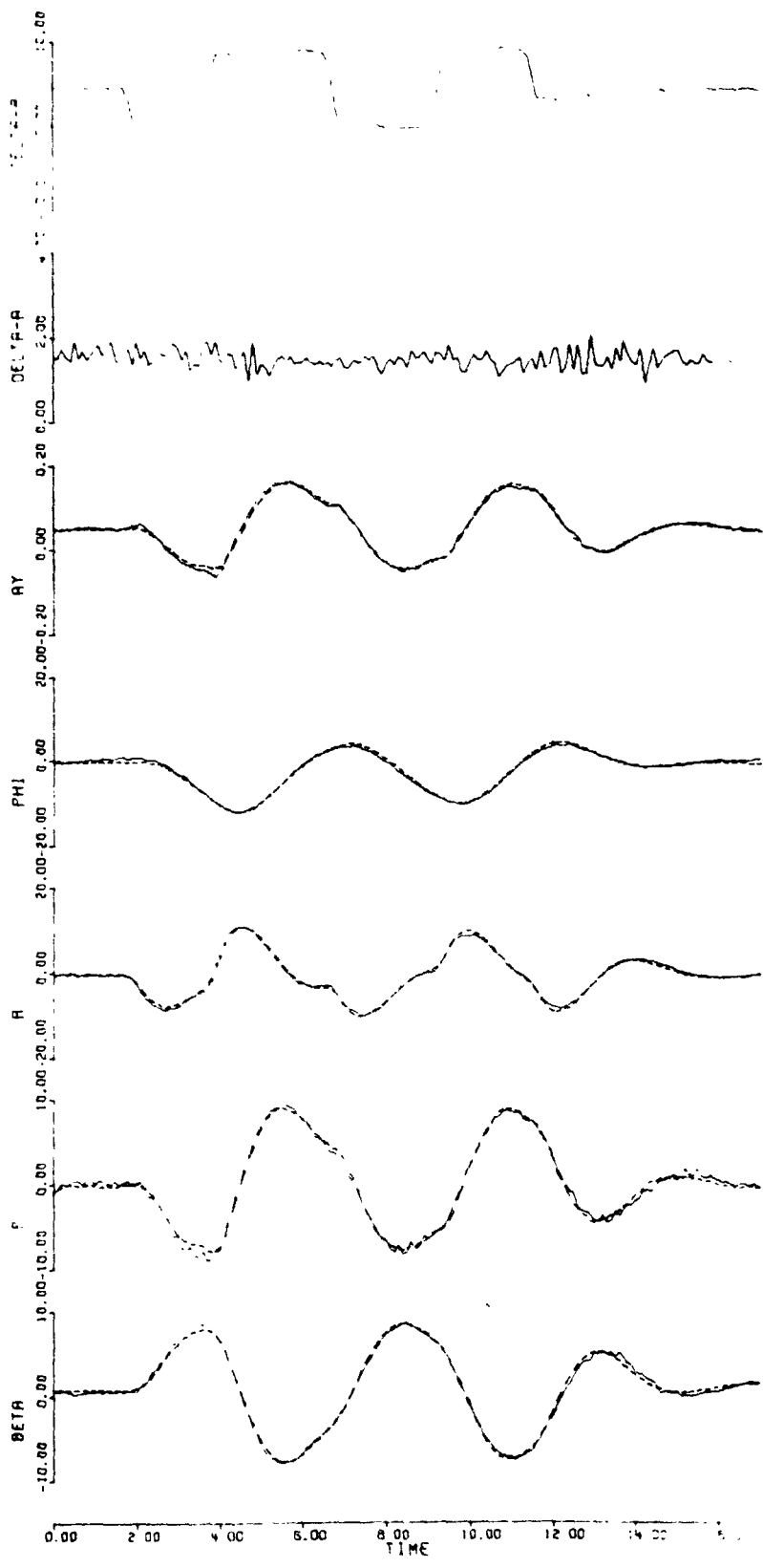
FIG. 9. POOLED MANEUVER: CESSNA RUN 9.
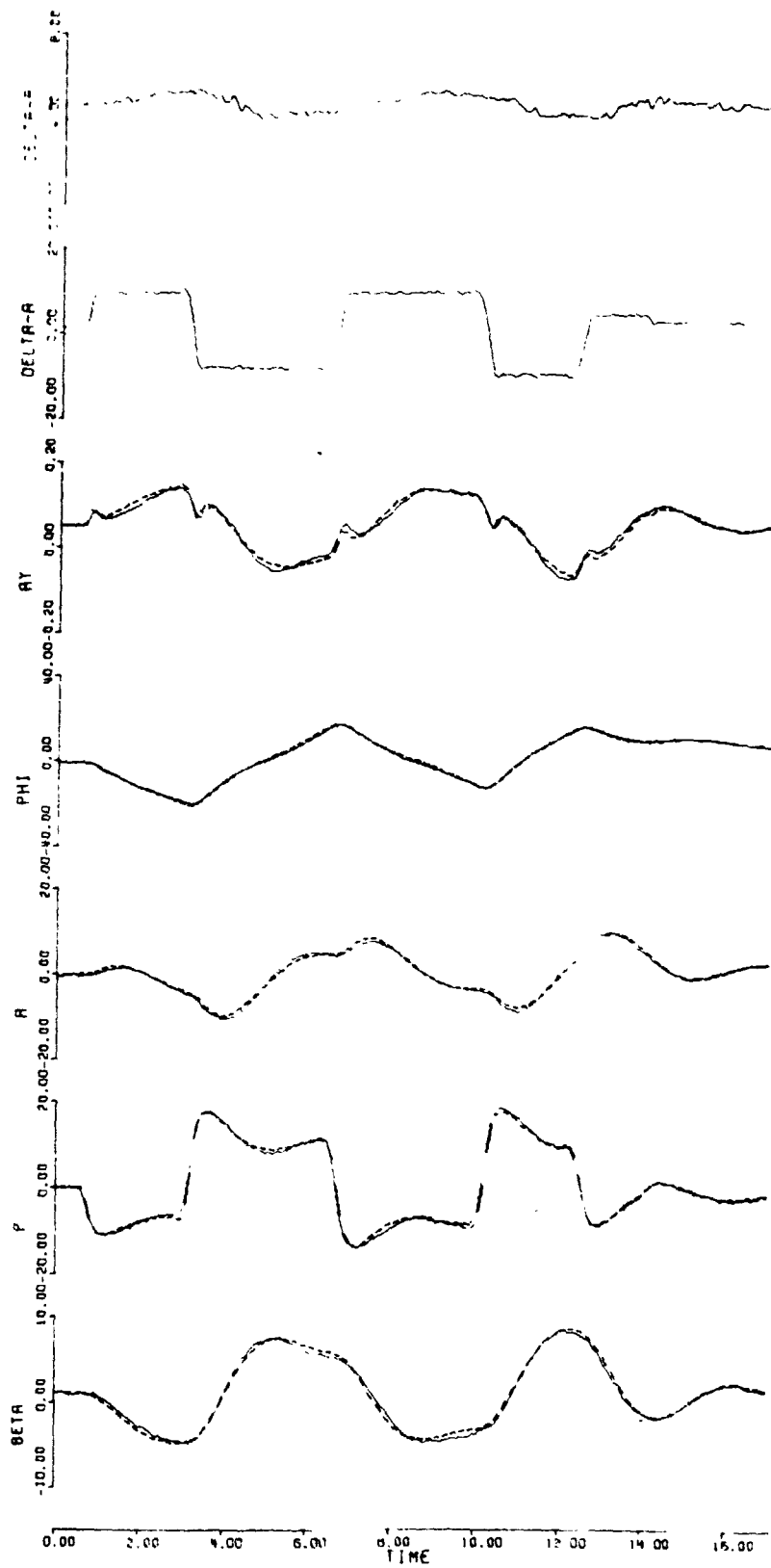
CESSNA 172  FLT 59  CASE 9-11

MANEUVER 1

FIG. 10. INPUT MANEUVER: CESSNA RUN 11.

CESSNA 172   FLT 59   TASK 11

MANEUVER  2

64

FIG. 11. FIXED MANEUVER: CESSNA RUN 10.

CESSNA 172  FLT 54  CASE 10-12

MANEUVER 1

65
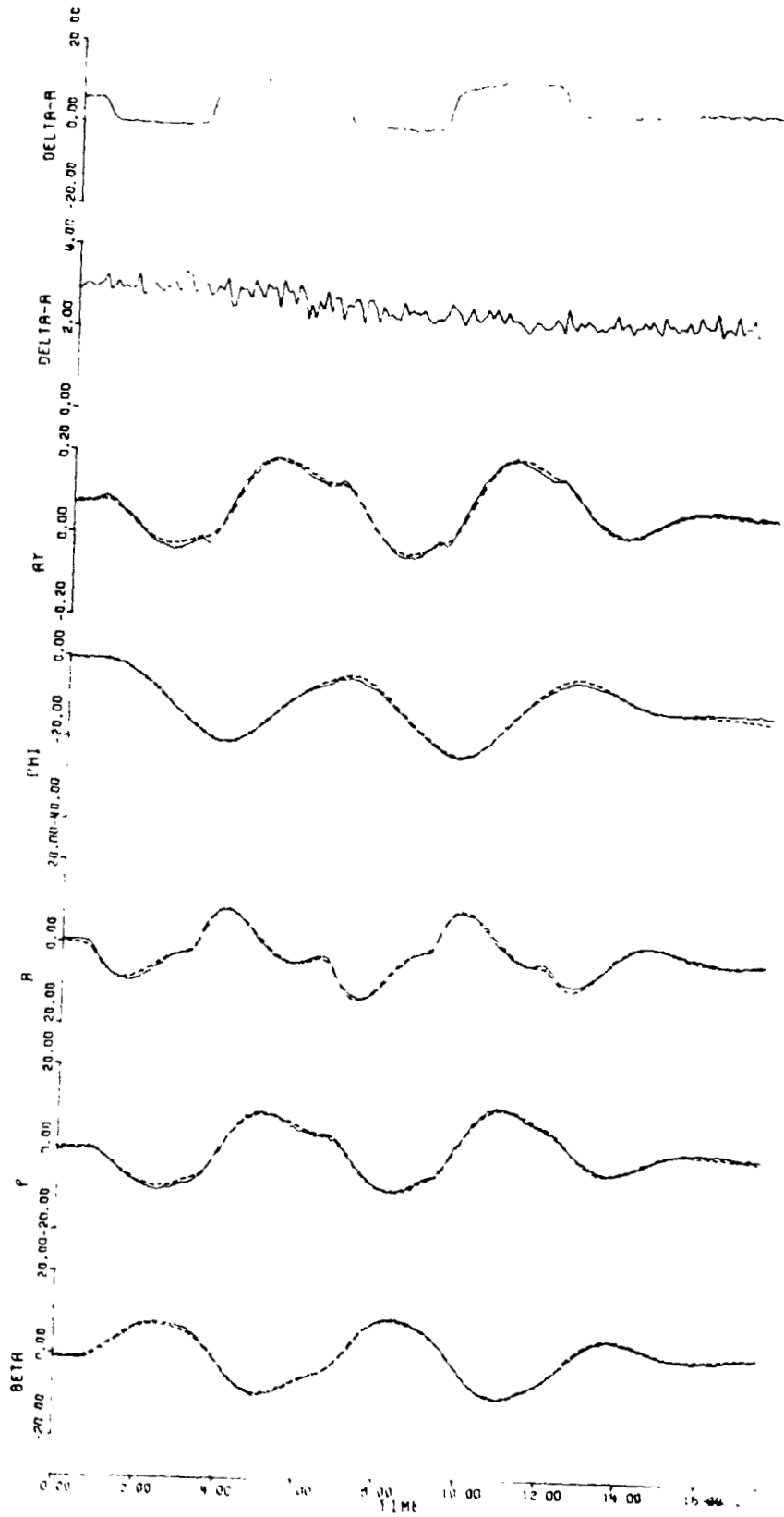
FIG. 1.'. RULED MANEUVER: CESSNA RUN L'.

CESSNA 172  FLT ... ... 10-12
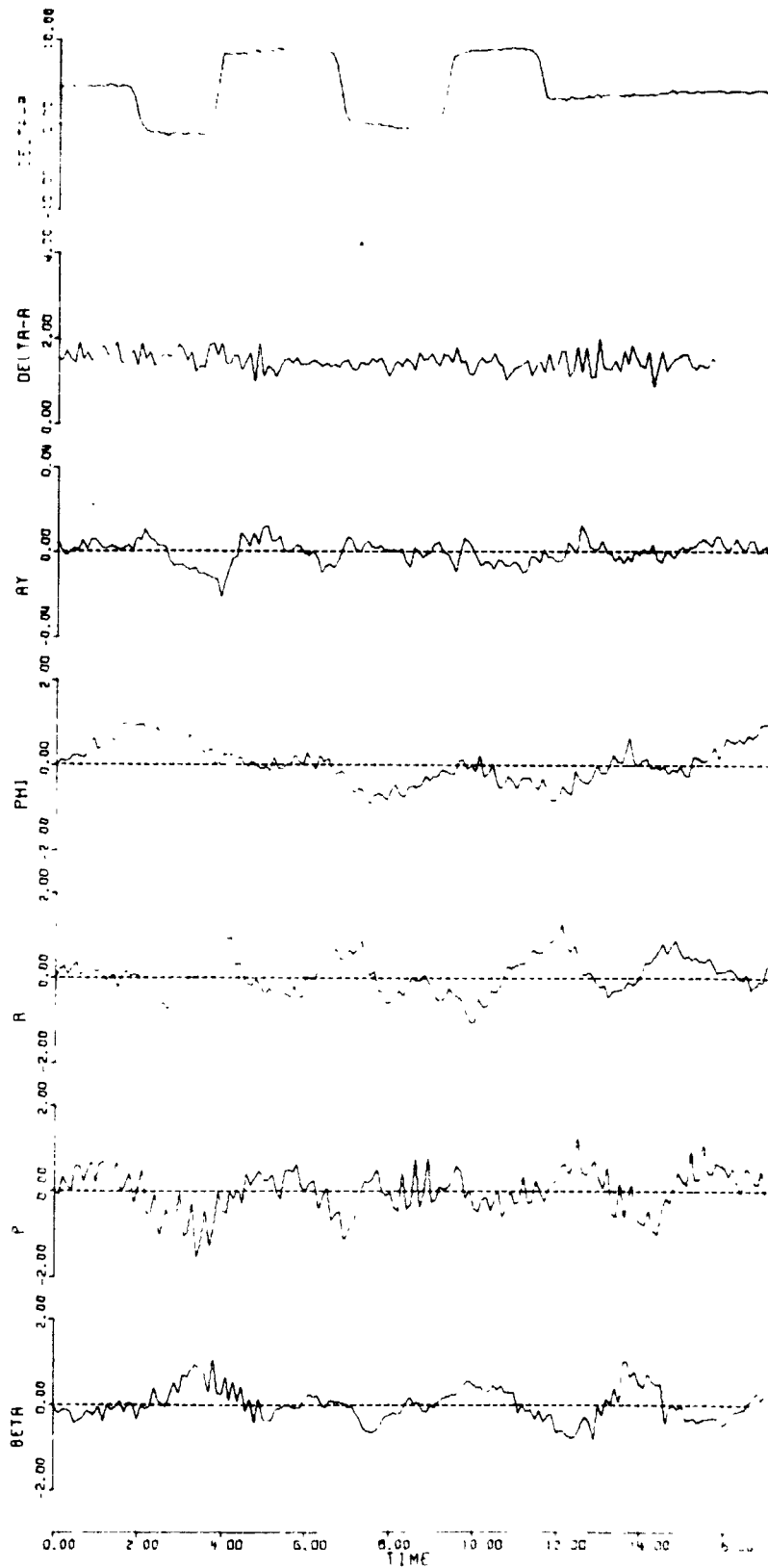
MANEUVER  2

FIG. 13. RESIDUALS: POOLED MAXIMUM R: CESSNA RUN 9.

CESSNA 172   FLT 59   CASE 9-11

MANEUVER :

67

FIG. 14. RESIDUALS: POOLED MANEUVER: CESSNA RUN 11.

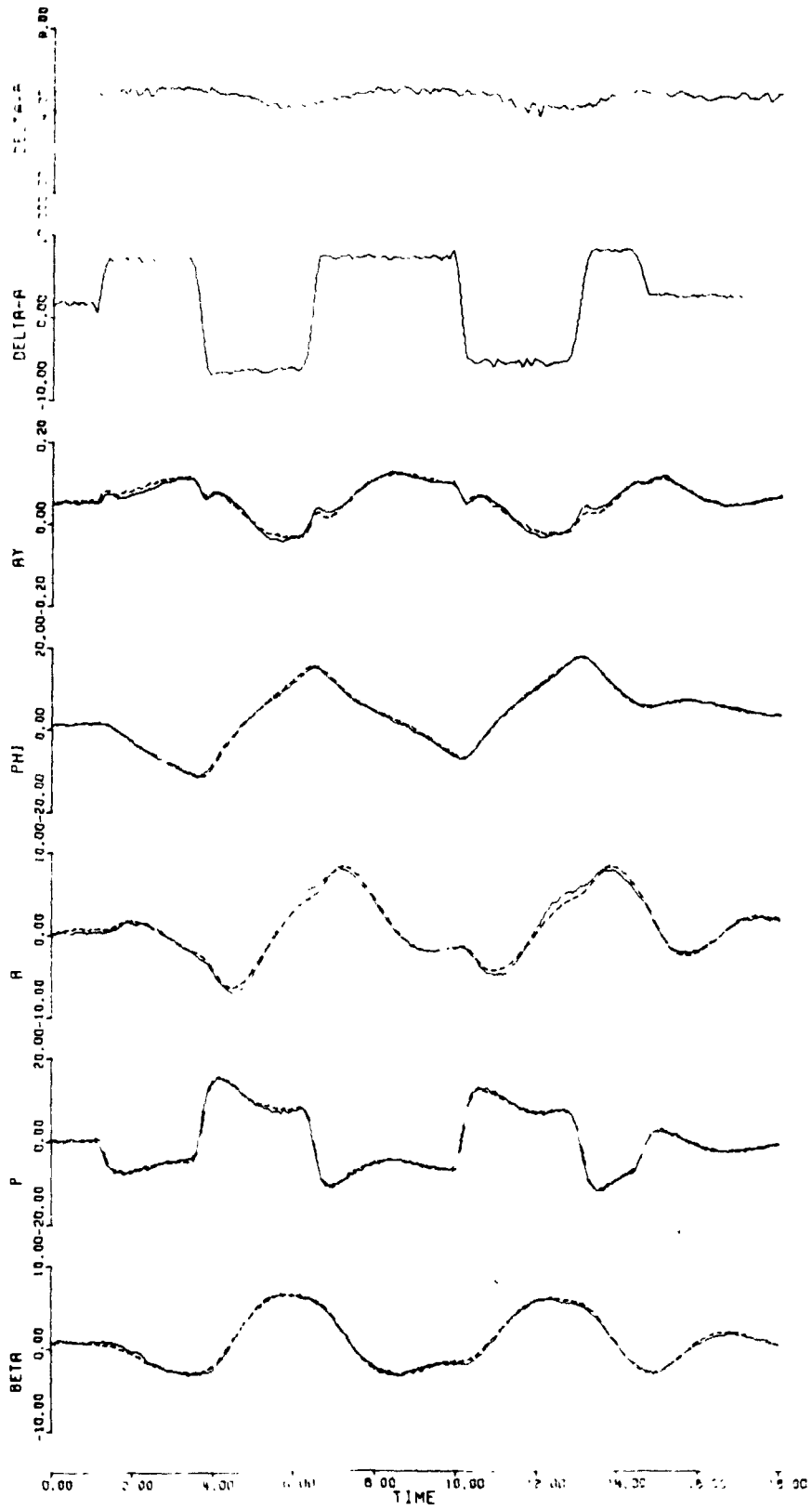CESSNA 172  FLT 59  CASE 9-11

MANEUVER  2

FIG. 15. RESIDUALS: POOLED MANEUVER: CESSNA RUN 10.

CESSNA 172   FLT 59   (RSF 10-12

MANEUVER  1

69

FIG. 1b. RESIDUALS: POOLED MANEUVER: CESSNA RUN 12.

CESSNA 172  FLT 59  CASE 10-12
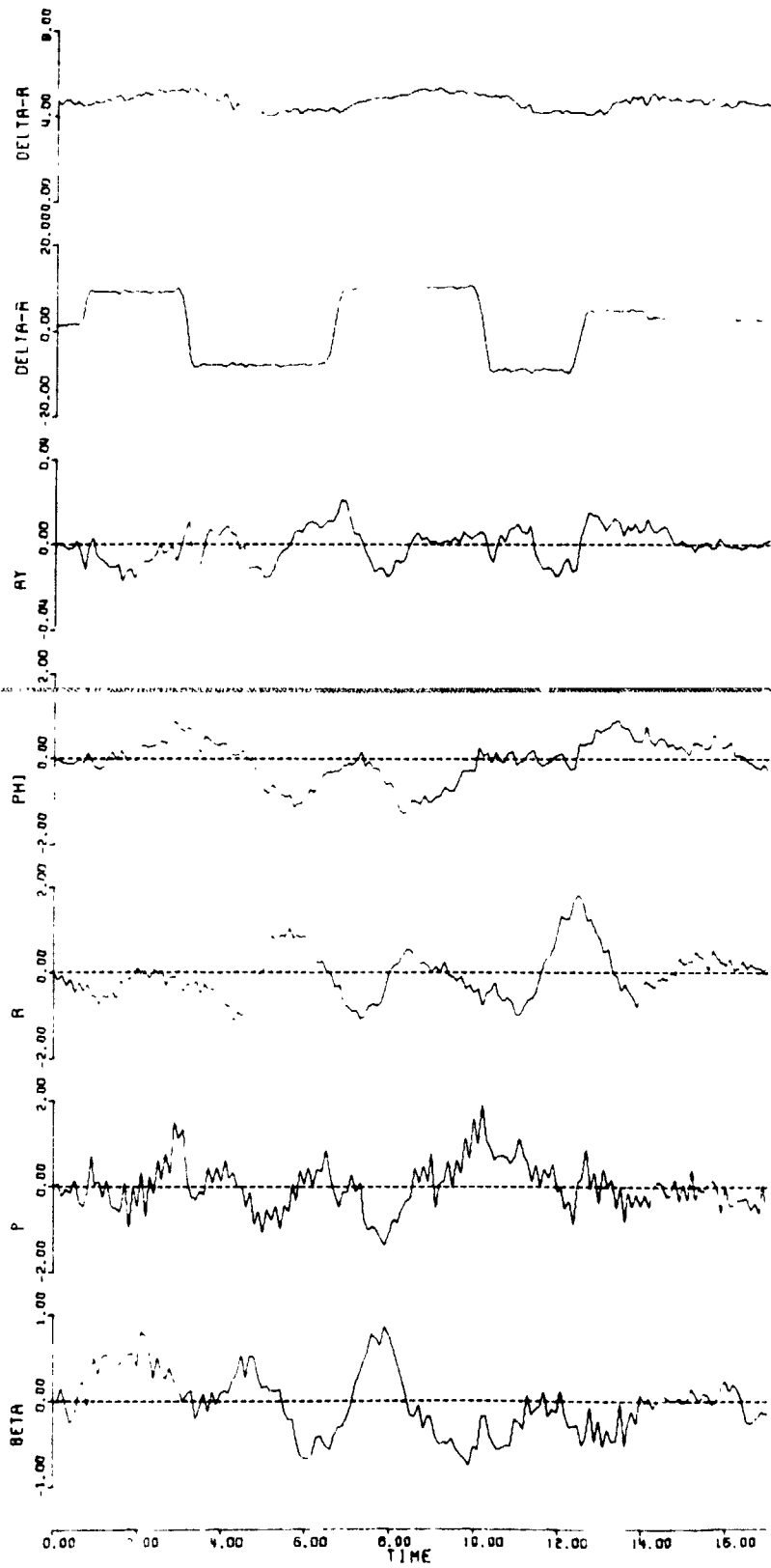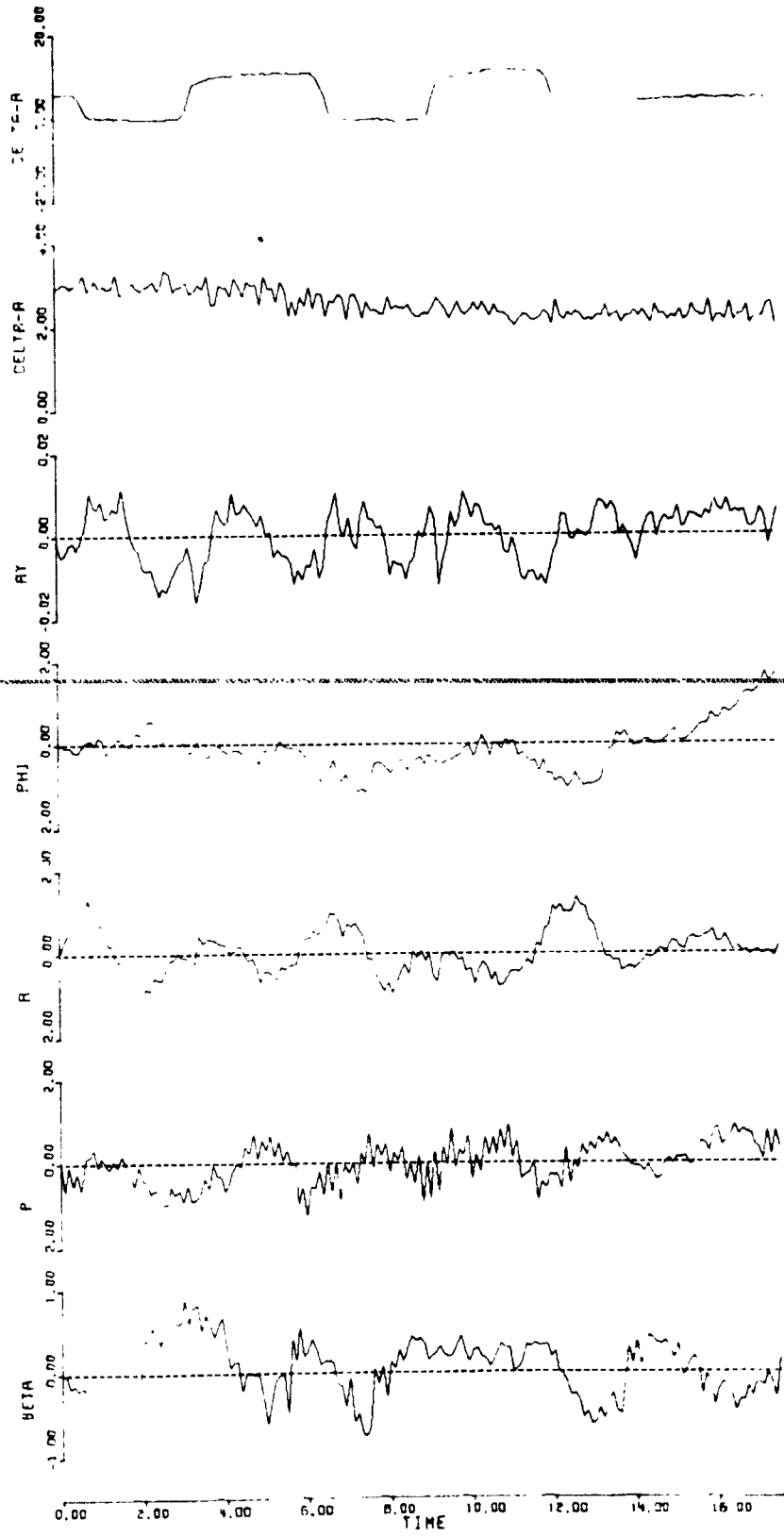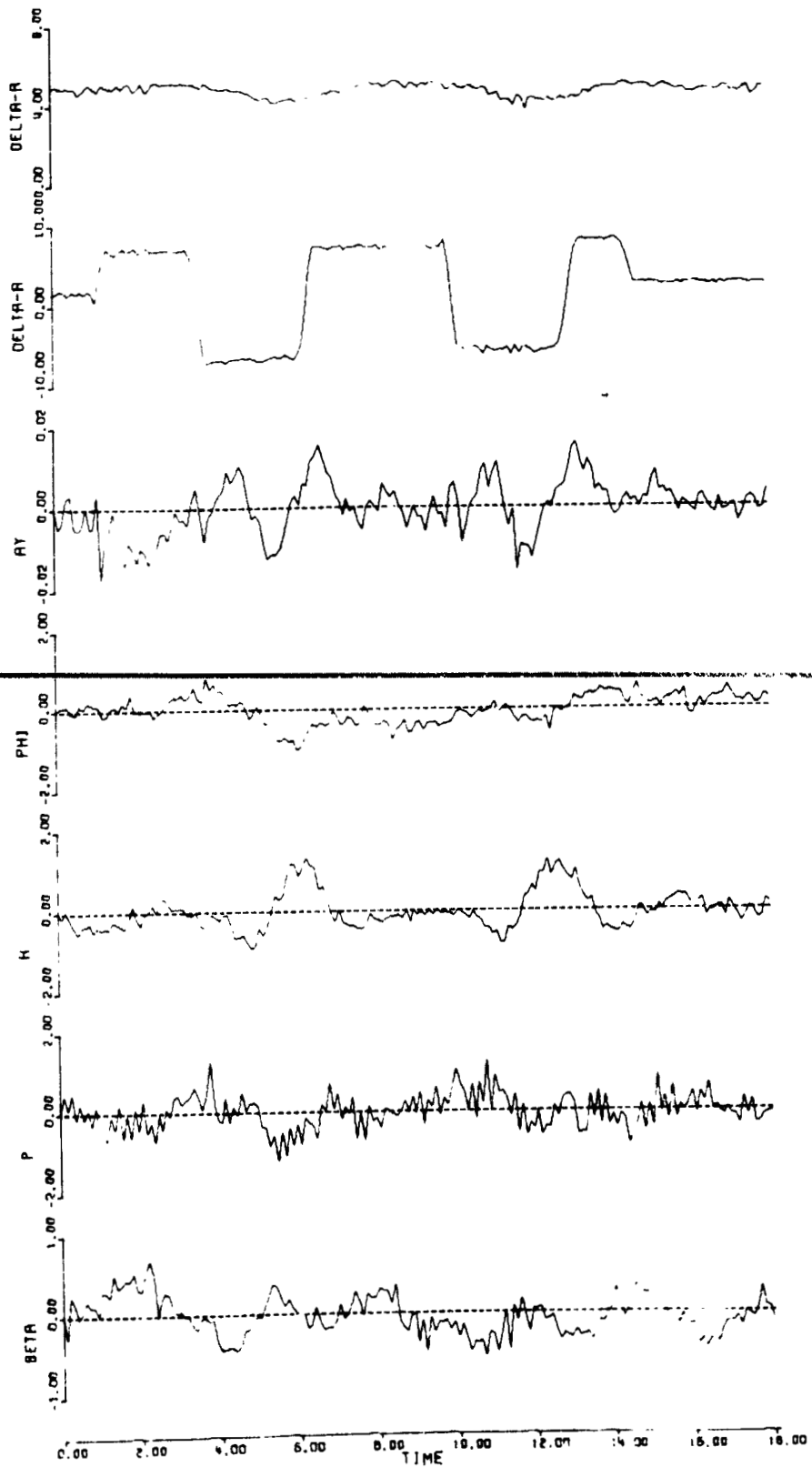
MANEUVER 2

70

F I G. 17
CESSNA: RESIDUAL PSD: β

71

FIG. 18.

CESSNA: RESIDUAL PSD -P

LOG PSD

LOG FREG

72

F I G. 19.
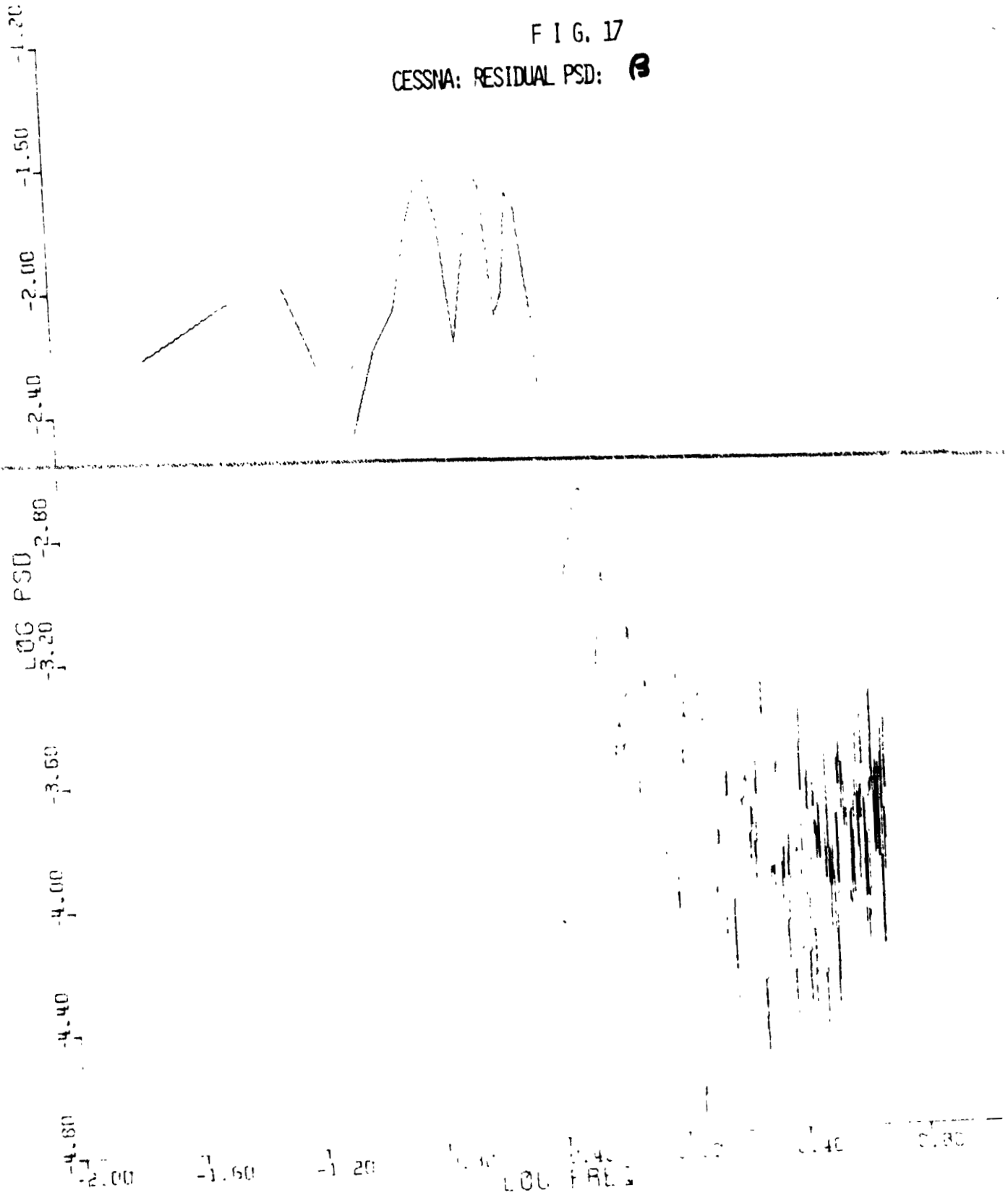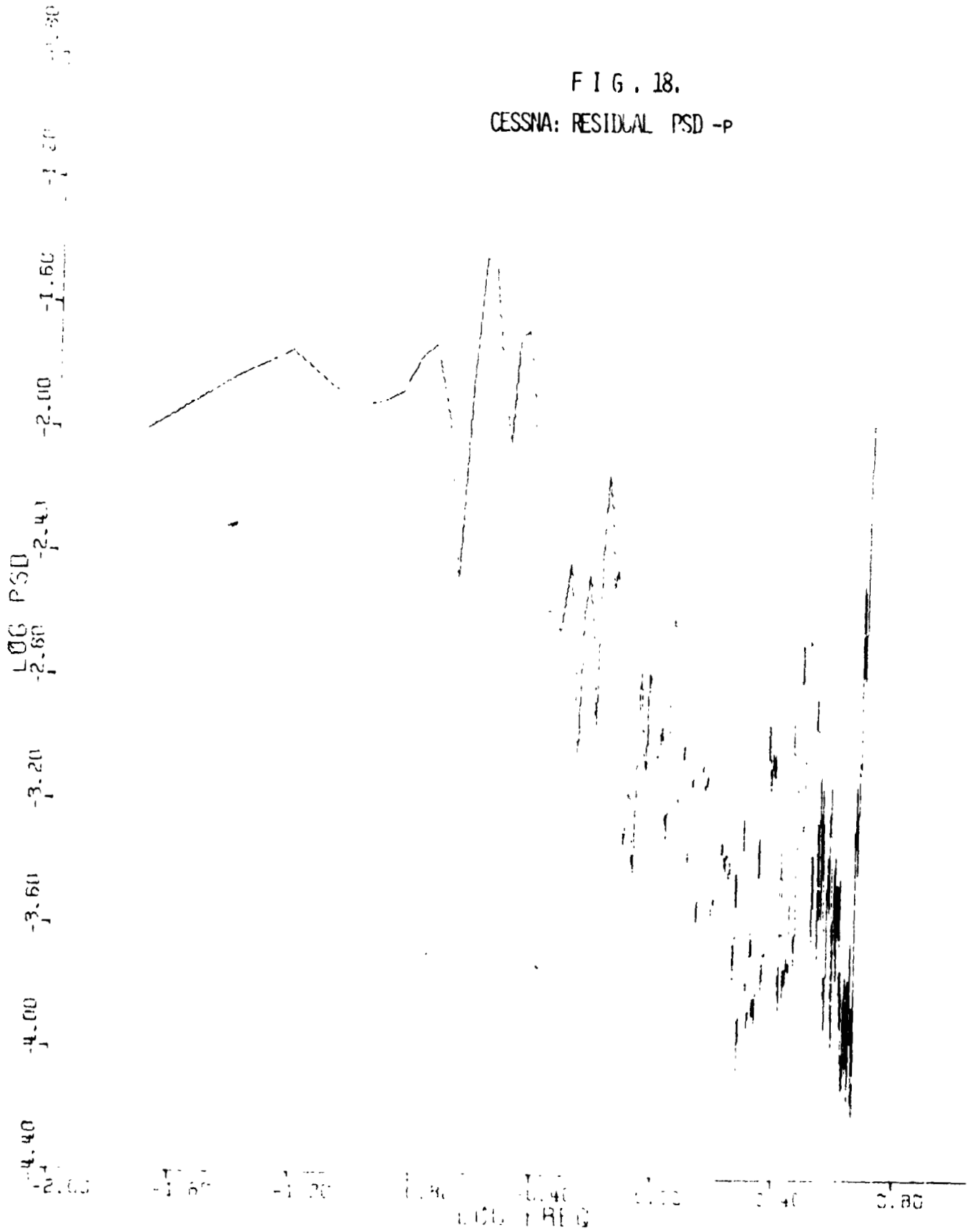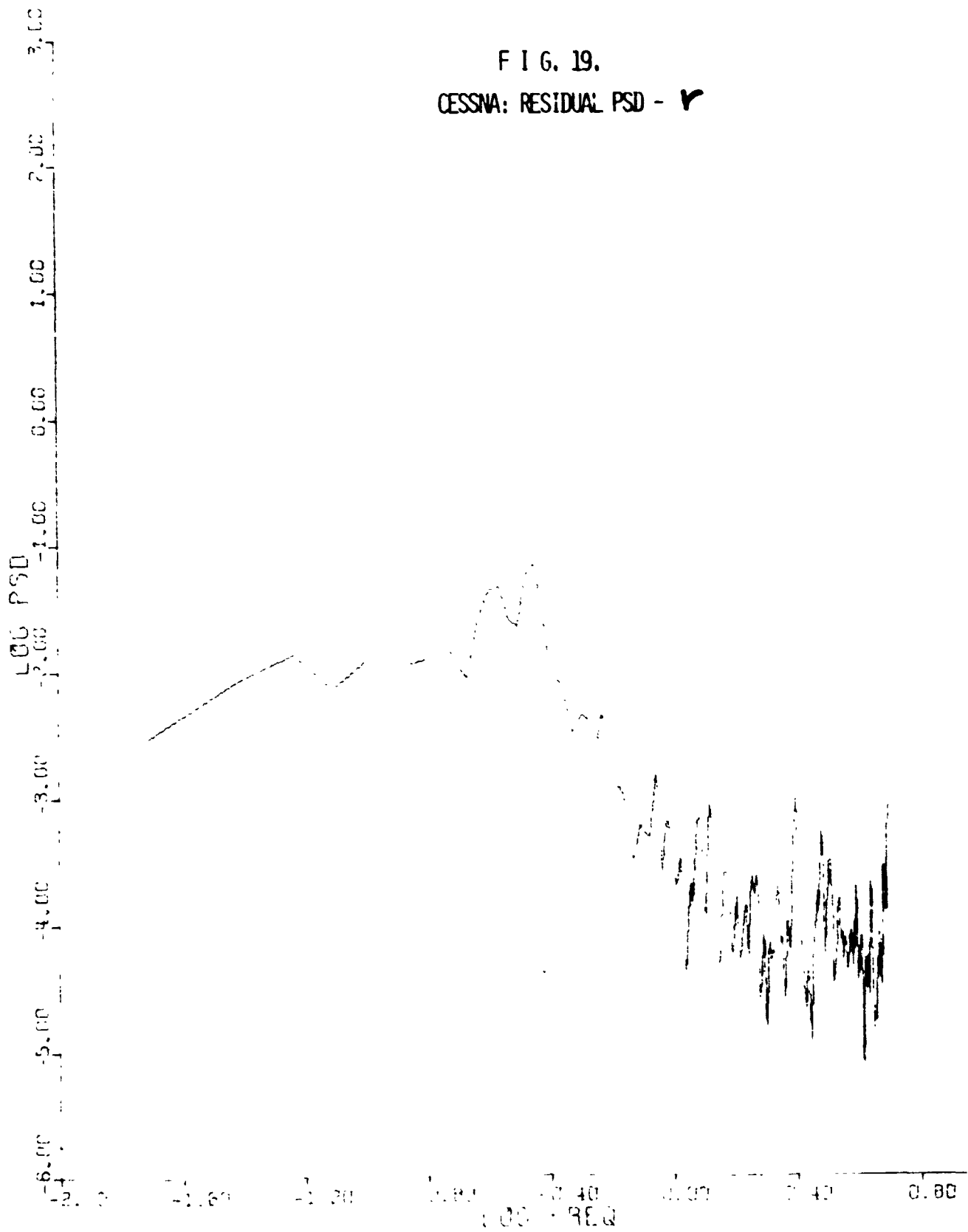
CESSNA: RESIDUAL PSD - $r$

F I G. 20.
CESSNA: RESIDUAL PSD—
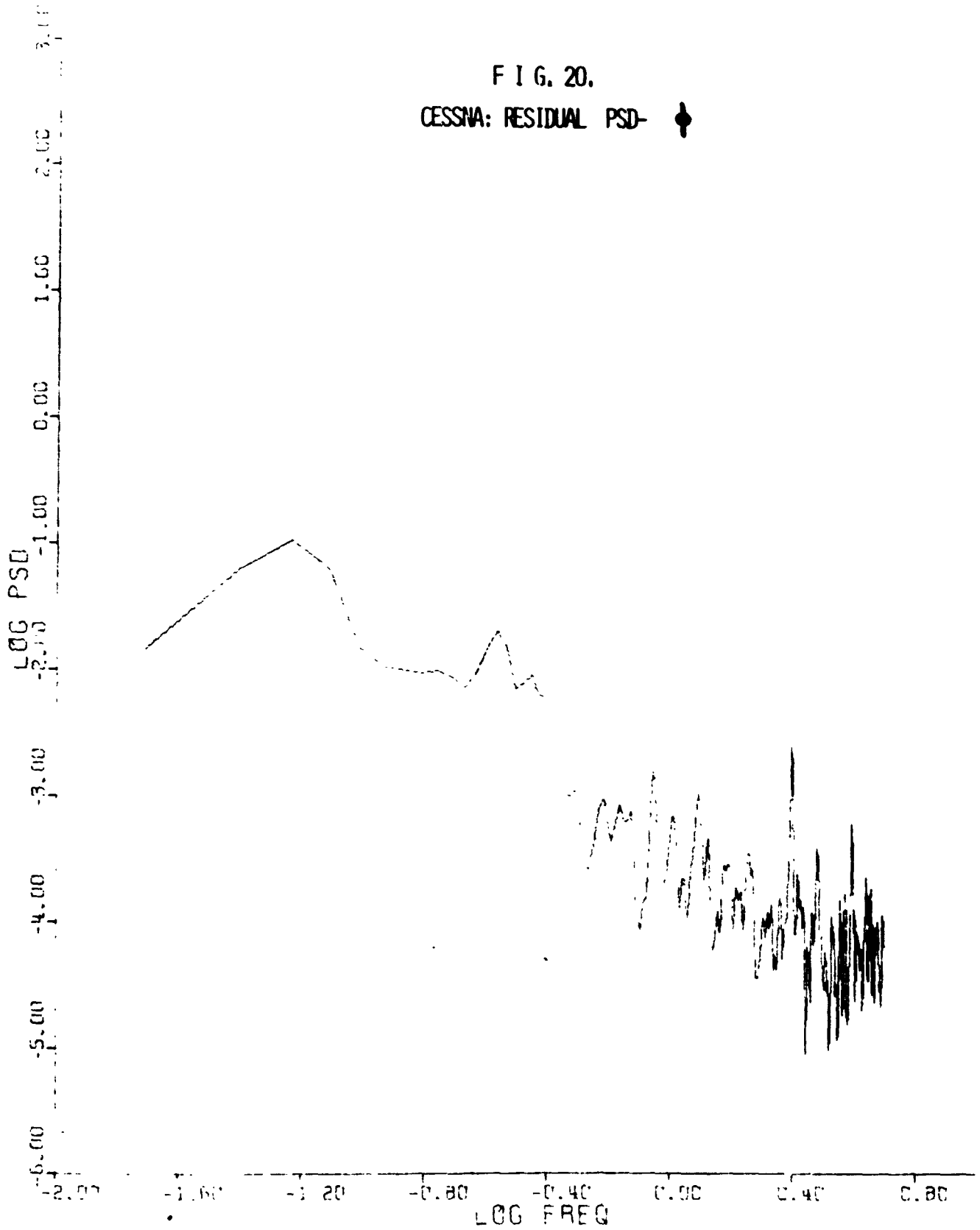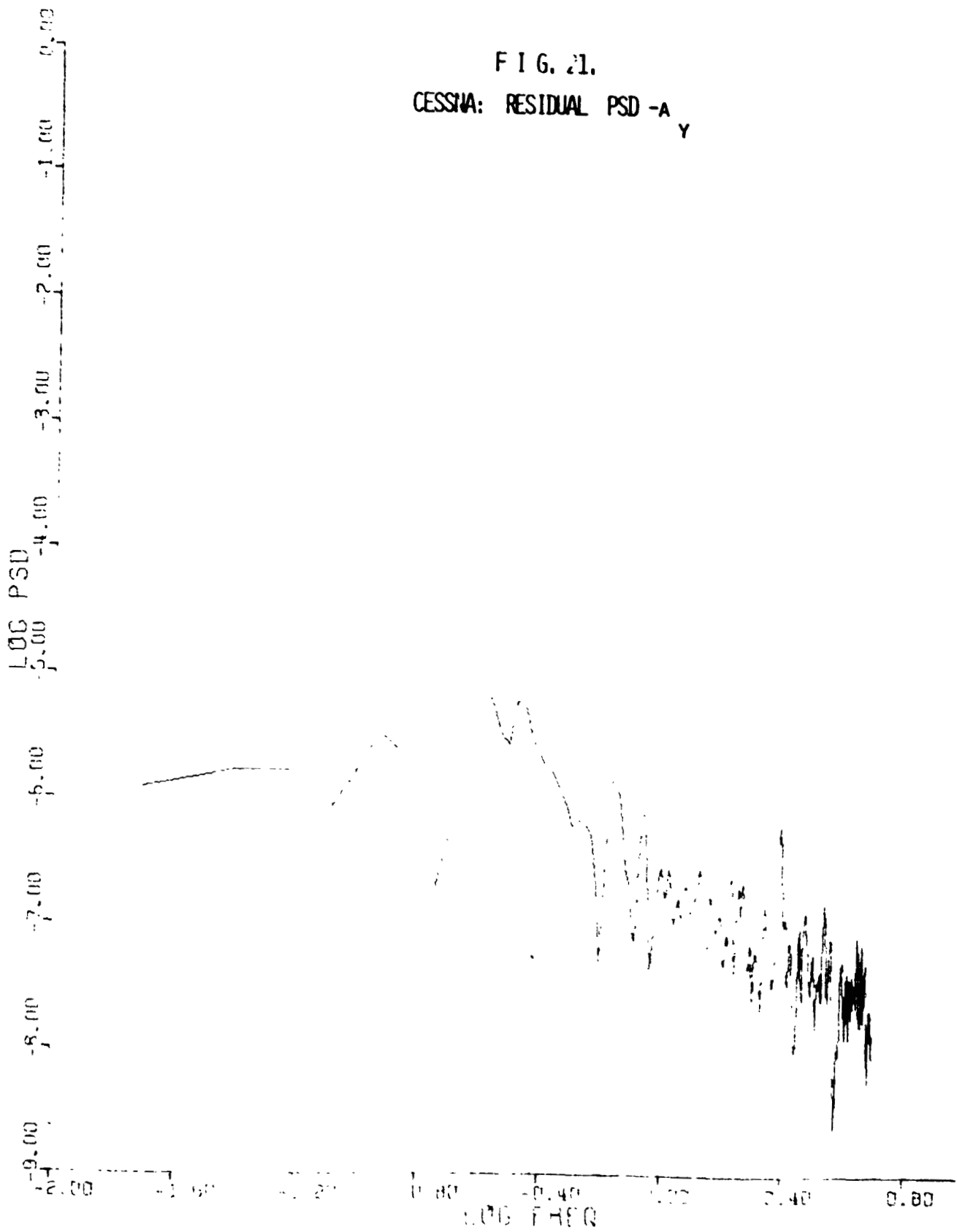
F I G. 21.

CESSNA: RESIDUAL PSD - A<sub>Y</sub>

PARAMETER VALUES BY ITERATION: CESSNA #11

| | START | ITERATION | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| $C_{Y_\beta}$ | -.006 | -.006 | -.005132 | -.005322 | -.005261 | -.005268 | -.005269 |
| $C_{l_\beta}$ | -.001 | -.001 | -.0005543 | -.006450 | -.006462 | -.006446 | -.0006444 |
| $C_{n_\beta}$ | .00032 | .00032 | .0002894 | .0003064 | .0002992 | .0002991 | .0002991 |
| $C_{l_p}$ | -.5 | -.5 | -.2881 | -.2974 | -.2926 | -.2917 | -.2916 |
| $C_{n_p}$ | -.1 | -.1 | -.07248 | -.07877 | -.08575 | -0.8534 | -.08535 |
| $C_{l_r}$ | .1 | .1 | .03943 | .06652 | .06722 | .06684 | .06686 |
| $C_{n_r}$ | -.1 | -.1 | -.05865 | -.07137 | -.07039 | -.07025 | -.07027 |
| $C_{Y_{\delta_a}}$ | 0 | .0008262 | .001008 | .0009679 | .0009883 | .0009863 | .0009861 |
| $C_{l_{\delta_a}}$ | -.0013 | -.001619 | -.0009193 | -.0009871 | -.0009764 | -.0009735 | -.0009731 |
| $C_{n_{\delta_a}}$ | -.00005 | -.0001444 | -.0001118 | -.0001132 | -.0001335 | -.0001327 | -.0001327 |

TABLE 1

# GRADIENTS BY ITERATION: CESSNA #11

| | ITERATION | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| $\nabla C_{Y_\beta}$ | (not computed) | -1000 | 550 | -86 | 5.4 | .99 |
| $\nabla C_{\ell_\beta}$ | - | -14000 | 17000 | 2600 | -350 | 15 |
| $\nabla C_{n_\beta}$ | - | -23000 | -80000 | -4700 | 850 | -67 |
| $\nabla C_{\ell_p}$ | - | -3.2 | -8.5 | -16 | .25 | .012 |
| $\nabla C_{n_p}$ | - | 19 | 17 | 28 | -4.8 | .10 |
| $\nabla C_{\ell_r}$ | - | 11 | -59 | 5.7 | .66 | -.095 |
| $\nabla C_{n_r}$ | - | -280 | 170 | -26 | 1.1 | -.045 |
| $\nabla C_{Y_{\delta_a}}$ | -8200 | .49 | 440 | 12 | -10 | -.062 |
| $\nabla C_{\ell_{\delta_a}}$ | +260000 | -3.7 | -20000 | 3800 | -160 | -18 |
| $\nabla C_{n_{\delta_a}}$ | +1000000 | -14 | -21000 | -4300 | 1400 | 31 |

TABLE 2

77

RMS RESIDUALS BY ITERATION: CESSNA #11

| | | | | | | |
|---|---|---|---|---|---|---|
| β | 2.24 | .524 | .516 | .354 | .350 | .350 | .350 |
| p | 2.89 | 1.11 | .704 | .562 | .563 | .561 | .561 |
| r | 1.85 | .868 | .825 | .582 | .584 | .585 | .585 |
| φ | 3.49 | .920 | .559 | .502 | .496 | .497 | .497 |
| $a_Y$ | .0333 | .0148 | .0108 | .0112 | .0111 | .0110 | .0110 |

TABLE 3

# C E S S N A: Parameter Estimates
## Iteration

| | Start | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|
| $C_{Y_\beta}$ | -.006 | -.007029 | -.006851 | -.005269 | -.00511 |
| $C_{\ell_\beta}$ | -.001 | -.0008411 | -.0007716 | -.0006444 | -.0006358 |
| $C_{n_\beta}$ | .00032 | .0004801 | .0004846 | .0002991 | .0003015 |
| $C_{\ell_\rho}$ | -.5 | -.3616 | -.3178 | -.2916 | -.2856 |
| $C_{n_\rho}$ | -.1 | -.01867 | -.004523 | -.08534 | -.08502 |
| $C_{\ell_r}$ | .1 | .08562 | .08934 | .06686 | .07049 |
| $C_{n_r}$ | -.1 | -.08268 | -.07520 | -.07027 | -.07753 |
| $C_{Y_{\delta_a}}$ | 0 | — | — | .0009861 | .001057 |
| $C_{\ell_{\delta_a}}$ | -.0013 | — | — | -.0009731 | -.0009885 |
| $C_{n_{\delta_a}}$ | -.00005 | — | — | -.0001327 | -.0001347 |
| $C_{Y_{\delta_r}}$ | .001 | -.001379 | -.00131 | — | — |
| $C_{\ell_{\delta_r}}$ | .0002 | -.0001862 | -.0001874 | — | — |
| $C_{a_{\delta_r}}$ | .0002 | .0006266 | .0005536 | — | — |

Table 4

# CESSNA DATA
## SIGNIFICANT SPREADS IN COEFFICIENTS

| Run Number | $C_{\ell_\beta}$ | $C_{n_\beta}$ | $C_{n_p}$ |
|:---:|:---:|:---:|:---:|
| 9 | -.0008411 | .0004801 | -.01867 |
| 10 | -.0007716 | .0004846 | -.004523 |
| 11 | -.0006444 | .0002991 | -.08534 |
| 12 | -.0006358 | .0003015 | -.08502 |

TABLE 5

# F8 FLIGHT CONDITIONS

| Run | Mach | Alpha | $\overline{Q}$ | V | Input |
|-----|------|-------|------|------|-------|
| 4 | .806 | 3.83 | 211.1 | 779.2 | $\delta_a$ |
| 5 | .788 | 4.01 | 200.3 | ⁻⁻?.8 | $\delta_r$ |
| 20 | .793 | 4.48 | 177. | | $\delta_a$ |
| 21 | .796 | 3.89 | 177.9 | | $\delta_r$ |

TABLE 6

## F8 SIMULATED DATA

| COEF | MEAN | C.R. | $\hat{\sigma}$ | $\dfrac{\hat{\sigma}}{C.R.}$ |
|------|------|------|------|------|
| $C_{Y_\beta}$ | -.02022 | .00020 | .00017 | .87 |
| $C_{\ell_\beta}$ | -.003705 | .000043 | .000042 | .98 |
| $C_{n_\beta}$ | .002273 | .000027 | .000021 | .80 |
| $C_{\ell_\rho}$ | -.4367 | .0064 | .0052 | .82 |
| $C_{n_\rho}$ | -.004216 | .0038 | .0033 | .87 |
| $C_{\ell_r}$ | .3230 | .034 | .0033 | .96 |
| $C_{n_r}$ | -.2648 | .021 | .017 | .78 |

TABLE 7

## F8 Flight Data

| COEF | MEAN | $\hat{\sigma}$ | C.R. | $\dfrac{\hat{\sigma}}{C.R.}$ |
|---|---|---|---|---|
| $C_{Y_\beta}$ | -.02038 | .00089 | .00032 | 2.8 |
| $C_{\ell_\beta}$ | -.003890 | .00049 | .000043 | 11.4 |
| $C_{n_\beta}$ | .002150 | .00023 | .000025 | 9.2 |
| $C_{\ell_\rho}$ | -.4680 | .077 | .0059 | 13.1 |
| $C_{n_\rho}$ | -.02421 | .031 | .0032 | 9.7 |
| $C_{\ell_r}$ | .1937 | .41 | .035 | 11.7 |
| $C_{n_r}$ | -.4317 | .25 | .018 | 13.9 |

TABLE 8

CESSNA

CRAMER-RAO BOUNDS

| | 9 | 10 | 11 | 12 |
|---|---|---|---|---|
| $C_{Y_\beta}$ | .000086 | .00011 | .00013 | .00010 |
| $C_{\ell_\beta}$ | .00012 | .000035 | .000011 | .0000091 |
| $C_{n_\beta}$ | .000061 | .000020 | .0000061 | .0000051 |
| $C_{\ell_p}$ | .048 | .014 | .0069 | .0058 |
| $C_{n_p}$ | .025 | .0080 | .0045 | .0041 |
| $C_{\ell_r}$ | .0053 | .0030 | .0031 | .0030 |
| $C_{n_r}$ | .0025 | .0012 | .0013 | .0013 |
| $C_{Y_{\delta_a}}$ | ----- | ----- | .000067 | .000056 |
| $C_{\ell_{\delta_a}}$ | ----- | ----- | .000019 | .000016 |
| $C_{n_{\delta_a}}$ | ----- | ----- | .000013 | .000012 |
| $C_{Y_{\delta_r}}$ | .00011 | .00014 | ------- | ------- |
| $C_{\ell_{\delta_r}}$ | .000028 | .000012 | ------- | ------- |
| $C_{n_{\delta_r}}$ | .000014 | .0000063 | ------- | ------- |

TABLE 9

## CESSNA RUN 9
## PERCENT PARAMETER SPREAD FOR 1% CHANGE IN COST

| COEF | ESTIMATED VALUE | $\frac{1}{S_{ii}}$ PERCENT case 1 | $\lambda_{ii}$ PERCENT case 2 | case 4 |
|---|---|---|---|---|
| $C_{Y_\beta}$ | -.0070 | 3.83 | 5.48 | .089 |
| $C_{\ell_\beta}$ | -.00084 | 1.53 | 62.7 | 54.8 |
| $C_{n_\beta}$ | .00048 | 1.06 | 56.9 | 53.3 |
| $C_{\ell_p}$ | -.36 | 2.58 | 59.9 | 51.6 |
| $C_{n_p}$ | -.019 | 14.2 | 598.6 | 557.7 |
| $C_{\ell_r}$ | .086 | 6.46 | 27.9 | 20.5 |
| $C_{n_r}$ | -.083 | 2.54 | 13.7 | 12.0 |
| $C_{Y_{\delta_r}}$ | -.0014 | 16.4 | 37.3 | 1.93 |
| $C_{\ell_{\delta_r}}$ | -.00019 | 1.81 | 66.8 | 57.39 |
| $C_{n_{\delta_r}}$ | .00063 | 0.31 | 10.4 | 9.32 |
| $C_{Y_0}$ | .017 | 32.8 | 41.8 | 1.89 |
| $C_{L_0}$ | .0036 | 0.40 | 26.9 | 24.4 |
| $C_{l_0}$ | -.0032 | 0.27 | 16.0 | 14.9 |
| $\phi_0$ | .11 | 15.0 | 150.2 | 6.72 |
| $A_Y$ BIAS | .036 | 3.29 | 7.82 | 0.32 |

TABLE 10
85

# CESSNA
# POOLED MANEUVERS (ZAY = .86)

| | ESTIMATES | | C-R BOUNDS (sd) | |
|---|---|---|---|---|
| | 9-11 | 10-12 | 9-11 | 10-12 |
| $C_{Y_\beta}$ | -.006472 | -.006384 | .000066 | .000065 |
| $C_{\ell_\beta}$ | -.0006950 | -.0006769 | .000011 | .000012 |
| $C_{n_\beta}$ | .0003600 | .0003494 | .0000049 | .0000047 |
| $C_{\ell_p}$ | -.3021 | -.2869 | .0049 | .0048 |
| $C_{n_p}$ | -.06398 | -.06317 | .0022 | .0021 |
| $C_{\ell_r}$ | .07112 | .07259 | .0024 | .0023 |
| $C_{n_r}$ | -.07578 | -.07599 | .00097 | .00091 |
| $C_{Y_{\delta_a}}$ | .0006652 | .0007154 | .000044 | .000048 |
| $C_{\ell_{\delta_a}}$ | -.001027 | -.001014 | .000016 | .000015 |
| $C_{n_{\delta_a}}$ | -.00006977 | -.00007758 | .0000072 | .0000072 |
| $C_{Y_{\delta_r}}$ | -.0007583 | -.0008331 | .00011 | .000089 |
| $C_{\delta_r}$ | -.0001335 | -.0001250 | .000013 | .000010 |
| $C_{n_{\delta_r}}$ | .0005766 | .0005460 | .0000054 | .0000044 |

TABLE 11