NASA Reference Publication 1077

The Theory and Practice of Estimating the Accuracy of Dynamic Flight-Determined Coefficients

Richard E. Maine and Kenneth W. Iliff

JULY 1981

NASA

NASA Reference Publication 1077

The Theory and Practice of Estimating the Accuracy of Dynamic Flight-Determined Coefficients

Richard E. Maine and Kenneth W. Iliff
Dryden Flight Research Center
Edwards, California

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1981

CONTENTS

	Page
INTRODUCTION	1
SYMBOLS AND ABBREVIATIONS	2
MAXIMUM LIKELIHOOD ESTIMATION	5
COMMON MEASURES OF ACCURACY	7
Engineering Judgment	7
Bias of the Estimates	8
Scatter	10
Cramér-Rao Inequality	12
Uncertainty Ellipsoid	18
Sensitivity	21
Correlation	23
Cramér-Rao Bound	27
EVALUATION WITH FLIGHT DATA	31
Discrepancy in the Cramér-Rao Bound	31
Previous Attempts at Explanation	35
Explanation for the Discrepancy	36
Suggested Implementations	42
EXAMPLES OF APPLICATION	44
Example 1	45
Example 2	47
Example 3	49
Example 4	51
Example 5	51
Example 6	54
Example 7	55
CONCLUSIONS	56
REFERENCES	57

THE THEORY AND PRACTICE OF ESTIMATING THE ACCURACY OF DYNAMIC FLIGHT-DETERMINED COEFFICIENTS

Richard E. Maine and Kenneth W. Iliff
Dryden Flight Research Center

INTRODUCTION

Parameter estimation is widely used to obtain aircraft stability and control derivatives from dynamic flight maneuvers. Reference 1 describes the details of this application. The estimates are used during flight test programs to permit the safe expansion of the flight envelope, to improve the fidelity of simulators, and to analyze the effects of control system changes. The estimates are also compared with theoretical and wind tunnel predictions to validate and improve prediction and estimation techniques.

Derivatives obtained from any source are by nature only estimates and not exact values, a fact which is often ignored. To make effective use of these estimates, it is necessary to have some gage of their reliability, whether the gage is a statistical measure or a guess based on intuition. If highly accurate estimates cannot be distinguished from worthless estimates, to be safe all estimates must be treated as worthless. Therefore, measures of estimate accuracy are as valuable as the estimates themselves. Nonetheless, the subject of measuring estimate reliability is seldom explored. This paper examines the use of the various measures of accuracy with flight data.

There are three general applications for measures of estimate accuracy. The first is planning the flight tests for the derivative estimation itself. Predictions of estimate accuracy can be used to evaluate proposed maneuvers and instrumentation systems. This application is necessarily limited, because it involves predicting accuracy before the data are obtained. The second application is the comparison of derivative estimates. The comparisons can be between independent flight estimates or between flight estimates and predictions. In either case, the magnitudes of the observed differences can be compared to the accuracy measures to determine whether the differences are significant; if so, the measures of estimate accuracy can also be

used to help decide which of the conflicting values are better or whether some compromise between them should be considered. The third application for measures of estimate accuracy is presentation with the final derivative estimates for the users of the derivatives. If the derivatives are to be used for control system design, for instance, knowledge of the accuracy of the estimates is useful for evaluating the sensitivity of the control system.

Measures of accuracy are particularly important if the estimates are to be used by an explicit adaptive or learning control system (ref. 2). Such immediate use precludes the interjection of engineering judgment; the evaluation of the estimates must be entirely automatic. Such control systems must recognize poor estimates and suitably discount them.

This report discusses both the theoretical and the practical aspects of accuracy estimation. The examples center on the estimation of aircraft stability and control derivatives, but many of the principles have much wider application. The accuracy of maximum likelihood estimates is treated in depth because many concrete results are available on this subject and because maximum likelihood is the estimation method most commonly used for aircraft stability and control derivatives.

The definition of maximum likelihood estimation is first briefly reviewed. The various measures of accuracy in common use are then discussed and compared. Flight data are used to evaluate some of the theoretical results, and a resolution of the discrepancy noted in this evaluation is presented. Several examples of the application of accuracy estimation to real flight data are presented.

Although it discusses ways to evaluate the accuracy of parameter estimates, this report does not directly address the related question of how to improve the accuracy. Reference 1 discusses practical aspects of obtaining accurate estimates of aircraft stability and control derivatives from flight data.

SYMBOLS AND ABBREVIATIONS

a_n	normal acceleration, g
a_y	lateral acceleration, g
B	band limit, Hz
b	bias
C_1, C_2	Cramér-Rao bounds of ξ_1 and ξ_2
C_ℓ	coefficient of rolling moment
C_m	coefficient of pitching moment

C_n	coefficient of yawing moment
C_Y	coefficient of lateral force
$E\{\ \}$	expected value
e_i	unit vector along i-axis
FF^*	state noise power spectral density matrix
$f(\ )$	general function
GG^*	measurement noise covariance matrix
$g(\ )$	general function
H	approximation to information matrix
I	identity matrix
J	cost functional
L	likelihood function
M	Fisher information matrix
m	length of observation vector
N	length of arbitrary vectors
N	number of time points
n	state noise vector
PSD	power spectral density
p	roll rate, deg/sec
$p(\ )$	probability density
q	pitch rate, deg/sec
R	prediction error covariance matrix
r	yaw rate, deg/sec
S_1, S_2	insensitivities of ξ_1 and ξ_2
t	time, sec

u	control vector
X	arbitrary vector
x	state vector or arbitrary vector
Y	arbitrary vector
y	arbitrary vector
Z	complete time history of response
z	observation vector
$\hat{z}_\xi$	prediction of z based on ξ
α	angle of attack, deg
β	angle of sideslip, deg
Δt	sampling interval, sec
δ_a	aileron deflection, deg
δ_r	rudder deflection, deg
η	measurement noise vector
Λ	arbitrary matrix
λ	Lagrange multiplier
ξ	vector of unknowns
φ	bank angle, deg
∇_ξ	gradient with respect to ξ

Superscripts:

$\tilde{}$	predicted estimate
$\hat{}$	estimate
*	adjoint (transpose)

Subscripts:

i, j, o	general indices

max	maximum
min	minimum
p, q, r	derivative with respect to indicated quantity, per rad
t	true value
α, β, δ_a, δ_r	derivative with respect to indicated quantity, per deg

MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood estimation is by far the most commonly used technique for estimating parameters from dynamic flight data. The theory of maximum likelihood estimation is extensive, so a lot can be inferred about the accuracy of maximum likelihood estimates. Some of the accuracy measures discussed in this report can be readily applied only to maximum likelihood or related estimators. The emphasis of this paper is on maximum likelihood estimates, so the most basic results on maximum likelihood estimation are briefly reviewed below. No attempt is made here to derive these results or to discuss them in detail (refs. 3 to 5). References 6 and 7 present background information on probability theory in an easily understood manner.

The derivations in this paper all use continuous-time models with discrete observations.

The basic form of the model is

$$\left.\begin{aligned} \dot{x}(t) &= f[x(t),\ u(t),\ t,\ \xi] + Fn(t) \\ z(t_i) &= g\left[x(t_i),\ u(t_i),\ t_i,\ \xi\right] + G\eta(t) \end{aligned}\right\} \quad (1)$$

The unknown vector ξ is to be estimated. The F- and G-matrices can also be functions of ξ; if so, they usually receive special treatment, and are not of concern to this report. The specific forms of equation (1) used for aircraft stability and control analysis are given in references 1 and 8.

For each value of ξ, one can define the probability distribution $p(Z|\xi)$ of the response Z. The likelihood function is defined as follows:

$$L(\xi) = p(Z|\xi)\Big|_{Z=Z_{\text{measured}}} \quad (2)$$

The likelihood function is thus a measure of the relative plausibility of the measured response for each value of ξ. The maximum likelihood estimate of ξ is defined as the value that maximizes the likelihood function; in other words, it is the value that makes the measured data most plausible.

To derive the form of the likelihood function for the system described by equation (1), assume that the state noise vector n is a Gaussian white noise vector with identity covariance and zero mean. Also assume that the measurement noise vector η is a zero mean, identity covariance, Gaussian vector which is independent at each time point. The likelihood functional can then be evaluated as follows:

$$L(\xi) = [(2\pi)^m |R|]^{N/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\tilde{z}_\xi(t_i) - z(t_i)\right]^* R^{-1}\left[\tilde{z}_\xi(t_i) - z(t_i)\right]\right\} \tag{3}$$

where $\tilde{z}_\xi$ is the Kalman-filter-predicted z based on the postulated ξ, and R is the prediction error covariance matrix.

It is convenient to work with the negative of the logarithm of the likelihood functional. Naturally, minimizing the negative of the logarithm is equivalent to maximizing the original function. If R is known, the term multiplying the exponential in equation (3) is a constant and does not affect the maximization. The maximum likelihood estimates can then be obtained by minimizing the following cost functional:

$$J(\xi) = \frac{1}{2}\sum_{i=1}^{N}\left[\tilde{z}_\xi(t_i) - z(t_i)\right]^* R^{-1}\left[\tilde{z}_\xi(t_i) - z(t_i)\right] \tag{4}$$

Many algorithms can be used to minimize J. Reference 8 describes the Newton-Balakrishnan algorithm, which has been found to work well.

Two special cases of equation (1) are often used. If F is 0, $\tilde{z}$ can be obtained simply by integration, and R is GG^*. This is referred to as an output error method (ref. 9). If G is 0, and g can be inverted to obtain x from z, the estimator is essentially an equation error method (ref. 9). The general estimator derived from equation (2) is referred to as a prediction error method (ref. 3). Although most of the examples in this paper use an output error method, the principles are also applicable to the equation error and prediction error methods.

Three of the statistical properties of maximum likelihood estimates are of significance to this paper. All are asymptotic properties; that is, properties that are exactly true only in the limit of infinite data time.

First, the estimates are asymptotically unbiased; that is, the expected value is asymptotically equal to the true value. Second, the estimates asymptotically approach a Gaussian distribution. Third, the estimates are asymptotically efficient; that is, their variance approaches the limit given by the Cramér-Rao inequality. For finite data time, these properties are not exactly true. If the data time is long enough, the approximation will be good; experience supported by simple theoretical arguments suggests that "long enough" can usually be equated to a few periods of the lowest system natural frequency. These properties are further discussed later.

COMMON MEASURES OF ACCURACY

Various means of judging the accuracy of derivative estimates are in current use. This section discusses and compares the application of these accuracy measures to flight data.

First, it is necessary to define the term accuracy as it applies to real data. An aircraft is never described exactly by the simplified models used for analysis. Some of the sources of modeling error can be examined through simulations that use more complex models than are practical for use in flight data analysis. Such simulations can include, for example, models of specific postulated instrumentation errors (refs. 10 to 12). Regardless of the sophistication of the model, however, unexplained sources of error will always remain. Therefore, there is not a unique correct model for an aircraft.

The concept of accuracy is difficult to define if no correct model exists. It is most easily approached by considering the problem in two parts: estimation and modeling. For the estimation problem, the model is assumed to describe the airplane exactly. The accuracy of the estimates can then be precisely and quantitatively defined. Many results are available on the subject of estimation accuracy. Several of them are discussed below.

The modeling problem addresses the question of whether the form of the model can describe the aircraft behavior adequately for the use intended. There is very little guidance from the theory in this area. Studies such as references 13 to 15 discuss the selection of the best model from a set of candidates, but do not address the more basic problem of defining the candidate models. For the most part, the determination of model adequacy is based on engineering judgment. This paper does not address the modeling problem in detail, because few general conclusions can be drawn.

After the analyst considers estimation and modeling as separate problems, it is necessary to look at their interactions in order to complete the discussion of accuracy. One must examine the estimates that result from a model that is judged to be adequate, although not exact. As in the modeling problem, this process involves a large amount of engineering judgment. Some quantitative results can be obtained, however. The effects of specific postulated errors can be evaluated by sensitivity studies, as in reference 11. The section of this report entitled Explanation for the Discrepancy presents more general, but less rigorous, results.

Engineering Judgment

Engineering judgment is the oldest measure of estimate reliability. Even with the theoretical measures of accuracy now available, the need for judgment remains; the theoretical measures are merely tools that supply more information on which judgment can be based. By definition, the process of applying engineering judgment cannot be precisely and quantitatively described. Algorithms can be devised to search

for specific problems, but a final unautomated judgment is usually necessary. Therefore, this section is of necessity somewhat vague. It simply lists some of the factors often considered in making a judgment.

One of the most fundamental factors in judging the accuracy of the estimates is the anticipated accuracy. The engineer usually has *a priori* knowledge of how accurately he can reasonably expect to be able to estimate the derivatives. The advance knowledge can be based on previous experience, knowledge of the relative importance and linear dependence of the derivatives, and the quality of the data obtained. Another basic criterion is the reasonability of the estimated derivatives. Before analysis is begun, the approximate values of the derivatives are usually known to within some range. Drastic deviations from this range are reason to suspect the estimates unless the explanation for the poor prediction is discovered or the suspect value is independently verified.

The role of engineering judgment in evaluating model adequacy has already been mentioned. The engineer will be looking for violations of the assumptions made in deriving the model and for unexplained problems that may indicate modeling errors. Both the estimator and the theoretical measures of accuracy can be invalidated by modeling errors. In any case, the magnitude of the effects of modeling error must be judged.

The engineer judges the quality of the fit of the measured and estimated time histories. The characteristics of this fit can be used as indicators of many problems. Many modeling error problems first become apparent as poor time history fits. Failed sensors and data processing errors or omissions are among the other problems that can be diagnosed using the time history fits.

Finally, engineering judgment is used to assemble and weigh all of the available information about the estimates. The judgmental factors mentioned above must be combined with information from the theoretical tools described below to arrive at final best estimates of both the derivatives and their accuracy.

Bias of the Estimates

A bias is defined statistically as a consistent or repeatable error. The statistical definition of bias is more general than the commonly understood meaning of the term. Let an estimator be described as a function g (not necessarily expressed in a closed form) that operates on the measured system response to obtain the estimate of the parameter vector

$$\hat{\xi} = g(Z) \tag{5}$$

For a specific input, Z is a function of the parameter vector ξ and the random noise signals n and η, as follows:

$$Z = Z\left(\xi_t, n, \eta\right) \tag{6}$$

The bias is then defined as the function

$$b(\xi_t) = E\ \hat{\xi} - \xi_t\ = E\left\{g\left[Z(\xi_t, n, \eta)\right]\right\} - \xi_t \tag{7}$$

or, in other words, the expected value of the estimation error. Note that the bias is, in general, a function of the true value. It is also a function of the input signal, although this dependence is not explicit in the equation above.

An unbiased estimator is defined as an estimator for which $b(\xi_t)$ is 0 for all values of ξ_t. This requirement is quite stringent, and unbiased estimators are rare for complicated problems. It is usually difficult to compute the bias because g is a complicated function of Z and commonly involves the solution of a nonlinear optimization problem.

With reasonable assumptions, maximum likelihood estimators can be shown to be asymptotically unbiased. This means that the bias approaches 0 as the time interval analyzed becomes longer. For finite time, the bias is not 0, and it is impractical to compute. Balakrishnan proves asymptotic unbiasedness and the stronger result of asymptotic consistency in reference 4. The proofs are too involved to repeat in this report.

Although the bias is only 0 in the limit of infinite time, it will be small for long enough finite time intervals. A few cycles of the longest period of oscillation are usually sufficient to insure that the bias is negligible.

The proof that maximum likelihood estimators are asymptotically unbiased implicitly assumes that the model used to define the likelihood function is correct. The model incorporates assumptions about the noise statistics as well as about the system dynamics. Different noise assumptions result in different estimators for the same dynamic model. Basically, equation error (regression) estimates are asymptotically biased in the presence of measurement noise and output error estimates are asymptotically biased in the presence of process noise; prediction error estimators allow for both measurement and process noise without asymptotic bias. Noise that is not white and Gaussian will result in biases for all of these algorithms unless special provisions are made. Errors in modeling aircraft dynamics will result in a bias for any estimation algorithm.

Bias is not widely used as a measure of derivative accuracy for two reasons. First, it reflects only the consistent errors and ignores random scatter. In fact, if the bias were known, it could, except in rare cases, be subtracted from the estimates to obtain revised estimates that were unbiased. Second, the bias is, in general, difficult to compute. The preferred approach is to use an estimator that is matched to the observed noise characteristics in order to keep the bias arising from the noise small enough to neglect. Bias arising from errors in the dynamic model is unavoidable; the magnitude of such bias is usually judged by the engineer.

The analyst should always consider the possibility of significant bias due to modeling error. Bias errors are added to all other types of error in the estimates. Unfortunately, some bias errors are impossible to detect solely by analyzing the data.

The estimates can be repeatable with very little scatter and appear to be accurate by all other measures, and still have large biases. An example of this type of problem is a calibration error in a nonredundant instrument. The only way to avoid such problems is to use meticulous care and documentation in every step of the application, including modeling, instrumentation, and data handling. No automatic tests exist that substitute for such care.

One source of bias error is easy to analyze and should be considered—errors in the measured flight condition or vehicle mass properties. Nondimensional derivatives are directly affected by errors in velocity, dynamic pressure, mass, moments of inertia, and center of gravity. The possible biases in the measurement of these quantities can easily be translated into corresponding biases in the nondimensional derivatives, giving bounds on the bias errors from this source.

Scatter

When several maneuvers are made at a given flight condition, the scatter of the resulting estimates is an indication of accuracy. Data acquired from a series of maneuvers with gradually changing flight conditions can also be evaluated for scatter if it is assumed that derivatives change smoothly with flight condition. In that case the scatter about a smooth fairing is examined.

The scatter has a significant advantage over many theoretical measures of accuracy: it measures the actual performance that some of the theoretical measures try to predict. For this reason the scatter manifests several effects, such as random errors in the measurement of flight condition, that are ignored in the theoretical predictions. The most informative approach, of course, is to consider both the observed scatter and the theoretical predictions.

An inherent weakness in the use of the scatter as a gage of accuracy is that several data points are necessary to define it. Depending on the application, this weakness can range from inconsequential to insurmountable. A related problem is that the scatter does not give any indication of the accuracy of individual points, some of which may be more accurate than others. For instance, if only two conflicting data points are available, the scatter gives no indication as to which is more reliable. Figure 1, which is taken from reference 1, shows estimates of C_{n_p} obtained from flight data for a PA-30 aircraft. The scatter is large and shows estimates of both signs.

Figure 2 shows the same data segregated into rudder and aileron maneuvers. In this case, the scatter makes it evident that the aileron maneuvers result in far more consistent estimates of C_{n_p} than the rudder maneuvers. Had only one or two aileron and one or two rudder maneuvers been available, however, there would have been no way to deduce from the scatter that the aileron maneuvers were superior.

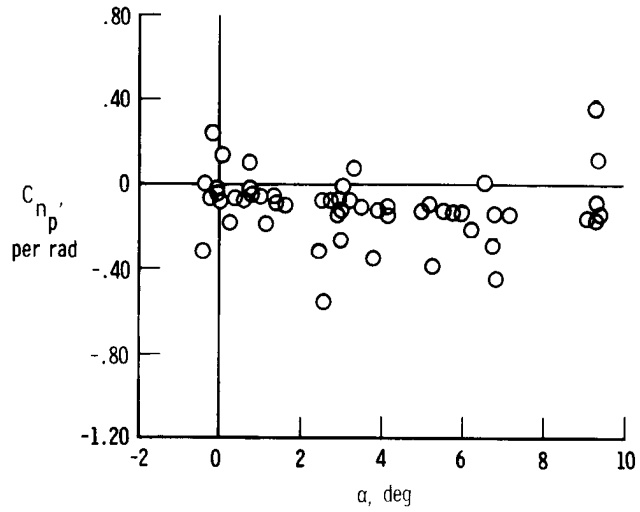


Figure 1. Estimates of C_{n_p} for PA-30 airplane.

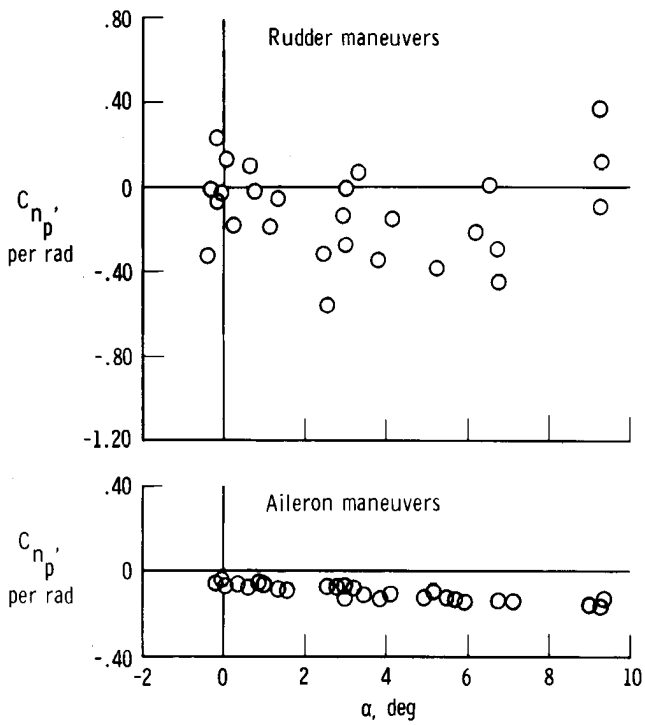


Figure 2. Estimates of C_{n_p} segregated by input used.

The scatter shares a weakness with the theoretical accuracy measures discussed below in that it does not account for consistent errors (biases). Many situations can result in small scatter about an incorrect value. A simple example of such a situation would be the presence of an error in the measurement of the moment of inertia. The scatter should therefore be regarded as a lower bound on the error in the estimates. The estimates may be worse than is indicated by the scatter, but they are seldom better.

In spite of its deficiencies, the data scatter is an easily used tool for evaluating accuracy, and it should always be examined when sufficient data points are available to define it.

Cramér-Rao Inequality

The Cramér-Rao inequality is the most important result in the theory of accuracy estimation. It is the basis for the parameter sensitivities, correlations, and Cramér-Rao bounds. The Cramér-Rao inequality gives a theoretical limit for the accuracy that is possible, regardless of the estimator used. In a sense, the Cramér-Rao inequality gives a measure of the information content of the data.

This section gives the derivation of the Cramér-Rao inequality and describes some of the factors that link the Cramér-Rao inequality to maximum likelihood estimators. The derivation of the inequality is essentially that of Balakrishnan (ref. 4), with some changes in notation.

A brief lemma is first proven.

LEMMA: *Let X and Y be two random N-vectors. Then*

$$E\{XX^*\} \geq E\{XY^*\}(E\{YY^*\})^{-1} E\{YX^*\}$$

assuming that the inverse exists.

Proof (by completing the square): Let Λ be any nonrandom N-by-N matrix. Then

$$E\{(X - \Lambda Y)(X - \Lambda Y)^*\} \geq 0 \tag{8}$$

because it is a covariance matrix. Expand the expression as follows:

$$E\{XX^*\} \geq \Lambda E\{YX^*\} + E\{XY^*\}\Lambda^* - \Lambda E\{YY^*\}\Lambda^* \tag{9}$$

Choose Λ such that

$$\Lambda = E\{XY^*\}(E\{YY^*\})^{-1} \tag{10}$$

Then

$$E\{XX^*\} \geq E\{XY^*\}(E\{YY^*\})^{-1} E\{YX^*\} + E\{XY^*\}(E\{YY^*\})^{-1} E\{YX^*\} - E\{XY^*\}(E\{YY^*\})^{-1} E\{YY^*\}(E\{YY^*\})^{-1} E\{YX^*\} \tag{11}$$

or

$$E\{XX^*\} \geq E\{XY^*\}(E\{YY^*\})^{-1} E\{YX^*\} \qquad \square \tag{12}$$

The Cramér-Rao inequality is a bound on $E\left\{\left(\hat{\xi} - \xi_t\right)\left(\hat{\xi} - \xi_t\right)^* | \xi_t\right\}$, the covariance of the estimate. (It should be noted that the covariance is a function of true value.)

THEOREM (Cramér-Rao): *Assume that* $p\left(Z|\xi_t\right)$ *exists and is sufficiently smooth to allow the operations required. Then, for any unbiased estimator,*

$$\text{var}\ (\hat{\xi}) \geq M\left(\xi_t\right)^{-1}$$

where

$$M\left(\xi_t\right) \equiv E\left\{\left[\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)\right]\left[\nabla_{\xi_t}^* \ln p\left(Z|\xi_t\right) |\xi_t\right]\right\}$$

Proof: Recall that $\hat{\xi}$ is a function of Z, the system response. Let X and Y from the lemma be $\hat{\xi} - \xi_t$ and $\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)$. Let all of the expectations in the lemma be conditioned on ξ_t. Concentrate first on the term,

$$E\left\{XY^*|\xi_t\right\} \equiv E\left\{\left(\hat{\xi} - \xi_t\right)\left[\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)\right]^*|\xi_t\right\} \equiv \int\left(\hat{\xi} - \xi_t\right)\left[\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)\right]^* p\left(Z|\xi_t\right) d|Z| \tag{13}$$

where $d|Z|$ is the volume element in the space of Z. Substituting the relation

$$\nabla_{\xi_t} \ln p\left(Z|\xi_t\right) = \frac{\nabla_{\xi_t} p\left(Z|\xi_t\right)}{p\left(Z|\xi_t\right)} \tag{14}$$

into equation (13) gives

$$E\left\{XY^*|\xi_t\right\} = \int\left(\hat{\xi} - \xi_t\right)\left[\nabla_{\xi_t} p\left(Z|\xi_t\right)\right]^* d|Z|$$

$$= \int\hat{\xi}\left[\nabla_{\xi_t} p\left(Z|\xi_t\right)\right]^* d|Z| - \int\xi_t\left[\nabla_{\xi_t} p\left(Z|\xi_t\right)\right]^* d|Z| \tag{15}$$

Observe that $\hat{\xi}$ is a function of Z only, not of ξ_t. Then, if $p\left(Z|\xi_t\right)$ is sufficiently smooth, the first term becomes

$$\int\hat{\xi}\nabla^*_{\xi_t} p\left(Z|\xi_t\right) d|Z| = \nabla^*_{\xi_t}\int\hat{\xi}\, p\left(Z|\xi_t\right) d|Z|$$

$$\equiv \nabla^*_{\xi_t} E\left\{\hat{\xi}|\xi_t\right\} \tag{16}$$

Use the definition of the bias (eq. (7)) to obtain

$$\int\hat{\xi}\nabla^*_{\xi_t} p\left(Z|\xi_t\right) d|Z| = \nabla^*_{\xi_t}\left[\xi_t + b\left(\xi_t\right)\right] = I + \nabla^*_{\xi_t} b\left(\xi_t\right) \tag{17}$$

Since ξ_t is not a function of Z, the second term of equation (15) becomes

$$\int\xi_t\nabla^*_{\xi_t} p\left(Z|\xi_t\right) d|Z| = \xi_t\nabla^*_{\xi_t}\int p\left(Z|\xi_t\right) d|Z| = \xi_t\nabla^*_{\xi_t} 1 = 0 \tag{18}$$

Thus, using equations (17) and (18) in equation (15) results in

$$E\left\{XY^*|\xi_t\right\} = I + \nabla^*_{\xi_t} b\left(\xi_t\right) \tag{19}$$

Define the Fisher information matrix as follows:

$$M\left(\xi_t\right) \equiv E\left\{YY^*|\xi_t\right\} \equiv E\left\{\left[\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)\right]\left[\nabla^*_{\xi_t} \ln p\left(Z|\xi_t\right)\right]|\xi_t\right\} \tag{20}$$

Then, from the lemma,

$$E\left\{\left(\hat{\xi} - \xi_t\right)\left(\hat{\xi} - \xi_t\right)^*|\xi_t\right\} \geq \left[I + \nabla_{\xi_t} b\left(\xi_t\right)\right]M\left(\xi_t\right)^{-1}\left[I + \nabla_{\xi_t} b\left(\xi_t\right)\right]^* \tag{21}$$

Equation (21) is the Cramér-Rao inequality. For unbiased estimators, $b\left(\xi_t\right)$ is 0, so

$$\mathrm{var}(\hat{\xi}) = E\left\{\left(\hat{\xi} - \xi_t\right)\left(\hat{\xi} - \xi_t\right)^*|\xi_t\right\} \geq M\left(\xi_t\right)^{-1} \qquad \square \tag{22}$$

Equation (21) gives a lower bound for the variance of any estimator with a given bias; equation (22) gives the specific form for unbiased estimators. Any unbiased estimator that attains the equality in equation (22) for every value of ξ_t is called an efficient estimator. The development above gives no guarantee that equality can be attained by any estimator; thus, efficient estimators do not always exist. When they do exist, efficient estimators are of particular interest, because no estimator can attain a lower variance without introducing a bias in the estimates. In this sense, efficient estimators extract all of the information available in the data.

The Cramér-Rao inequality applies to any estimator, but the theorems below make it particularly useful for maximum likelihood estimators.

THEOREM: *If an efficient estimator exists for a problem, that estimator is a maximum likelihood estimator.*

Proof: An estimator will be efficient if and only if equality holds in the lemma. Equality holds if X equals ΛY in equation (8). Substituting for Λ from equation (10) gives

$$X = E\{XY^*\}(E\{YY^*\})^{-1}\, Y \tag{23}$$

Substituting for X and Y and using equations (19) and (20) gives

$$\hat{\xi} - \xi_t = \left[I + \nabla^*_{\xi_t}\, b\left(\xi_t\right)\right] M\left(\xi_t\right)^{-1}\, \nabla_{\xi_t} \ln p\left(Z|\xi_t\right) \tag{24}$$

For unbiased estimators $b(\xi)$ is 0, so

$$\hat{\xi} - \xi_t = M\left(\xi_t\right)^{-1}\, \nabla_{\xi_t} \ln p\left(Z|\xi_t\right) \tag{25}$$

Equation (25) must hold for all values of ξ_t and Z ($\hat{\xi}$ is a function of Z). For each Z, therefore, it must hold when ξ_t equals $\hat{\xi}(Z)$. The left side of the equation is then 0, so

$$\nabla_{\xi_t} \ln p\left(Z|\xi_t\right)\Big|_{\xi_t=\hat{\xi}} = 0 \tag{26}$$

The estimate is thus at a stationary point of the likelihood function. Taking the gradient of equation (25) with respect to ξ_t, evaluating at $\xi_t = \hat{\xi}$, and using equation (26) gives

$$-I = M(\hat{\xi})^{-1}\, \nabla_{\xi_t}^{\,2} \ln p\left(Z|\xi_t\right)\Big|_{\xi_t = \hat{\xi}} \tag{27}$$

Since M is positive definite, the stationary point is a local maximum. It must in fact be the only local maximum, because equation (25) is valid for all ξ_t and Z. The requirement for $\int p(Z|\xi_t)\, d|Z|$ to be finite implies that it is a global maximum. The estimator is therefore a maximum likelihood estimator. □

This result is not as useful as it might first seem, because there is no guarantee that an efficient unbiased estimator exists for most problems. It motivates, however, the following important theorem.

THEOREM: *Under mild regularity conditions, maximum likelihood estimators for dynamic systems are asymptotically efficient; that is, the Cramér-Rao inequality approaches an equality for large data time.*

Proof: See reference 16.

Reference 16 also proves that the maximum likelihood estimates are asymptotically Gaussian. Experience, supported by simple theoretical arguments, suggests that a few cycles of the longest system natural period are usually adequate for these asymptotic results to become very accurate.

The consequence of the results above is that the Cramér-Rao inequality gives a close approximation to the variance of maximum likelihood estimates, rather than just a lower bound.

The form of the Fisher information matrix given in equation (20) is not ideal for computational purposes. If $p(Z|\xi_t)$ is sufficiently smooth, a useful alternative expression for the Fisher information matrix can be obtained as shown in the following theorem.

THEOREM: *If $p(Z|\xi_t)$ is sufficiently smooth, the Fisher information matrix can be expressed as follows:*

$$M(\xi_t) = -E\left\{\nabla_{\xi_t}^{\ 2} \ln p(Z|\xi_t)\,\middle|\,\xi_t\right\}$$

Proof: Applying equation (14) to the definition of M gives

$$M(\xi_t) = E\left\{\frac{\left[\nabla_{\xi_t} p(Z|\xi_t)\right]\left[\nabla_{\xi_t}^{*} p(Z|\xi_t)\right]}{p(Z|\xi_t)^2}\,\middle|\,\xi_t\right\} \tag{28}$$

Examine the expression

$$E\left\{\nabla_{\xi_t}{}^2 \ln p\left(Z|\xi_t\right)\Big|\xi_t\right\} = E\left\{\nabla_{\xi_t} \frac{\nabla^*_{\xi_t} p\left(Z|\xi_t\right)}{p\left(Z|\xi_t\right)}\Bigg|\xi_t\right\}$$

$$= E\left\{\frac{\nabla_{\xi_t}{}^2 p\left(Z|\xi_t\right)}{p\left(Z|\xi_t\right)}\Bigg|\xi_t\right\} - E\left\{\frac{\left[\nabla_{\xi_t} p\left(Z|\xi_t\right)\right]\left[\nabla^*_{\xi_t} p\left(Z|\xi_t\right)\right]}{p\left(Z|\xi_t\right)^2}\Bigg|\xi_t\right\} \tag{29}$$

The second term is equal to M, as shown in equation (28). The first term is evaluated as follows:

$$\int \frac{\nabla_{\xi_t}{}^2 p\left(Z|\xi_t\right)}{p\left(Z|\xi_t\right)} p\left(Z|\xi_t\right) d|Z| = \int \nabla_{\xi_t}{}^2 p\left(Z|\xi_t\right) d|Z|$$

$$= \nabla_{\xi_t}{}^2 \int p\left(Z|\xi_t\right) d|Z| = \nabla_{\xi_t}{}^2 1 = 0 \tag{30}$$

Thus, the information matrix can be expressed

$$M\left(\xi_t\right) = -E\left\{\nabla_{\xi_t}{}^2 \ln p\left(Z|\xi_t\right)|\xi_t\right\} \qquad \square \tag{31}$$

By using this form on the system of equation (1), the information matrix can be closely approximated by

$$H \equiv \sum_{i=1}^{N} \nabla_{\xi} \tilde{z}^*_{\xi}\left(t_i\right) R^{-1} \nabla^*_{\xi} \tilde{z}_{\xi}\left(t_i\right) \tag{32}$$

The H-matrix is identical to the dominant term of the Hessian matrix $\nabla_{\xi}{}^2 J(\xi)$. The H-matrix or a suitable approximation is thus required by most algorithms for minimizing the cost functional of equation (4). When F is 0 and G is known, H is an exact expression for the information matrix. Otherwise H is a good approximation to the information matrix.

Uncertainty Ellipsoid

The uncertainty ellipsoid is an approximation of the confidence region based on the Cramér-Rao inequality. It can also be defined by using concepts from optimization. This section reviews both approaches to the definition.

Throughout this section it is necessary to use the long form of the notation for the probability density function. The abbreviated notation used in other sections is not precise enough to express the concepts discussed.

To define the confidence region, first examine the probability density function of the estimate. As shown above, for maximum likelihood estimates this density function is approximately Gaussian, with mean equal to the true value and covariance equal to $M(\xi_t)^{-1}$.

$$p_{\hat{\xi}}(\xi) \cong \left[(2\pi)^m \left| M(\xi_t)^{-1} \right| \right]^{-N/2} \exp\left\{-\frac{1}{2}(\xi - \xi_t)^* M(\xi_t)(\xi - \xi_t)\right\} \tag{33}$$

Estimates of equal likelihood lie on isometric surfaces of the density function

$$p_{\hat{\xi}}(\xi) = \text{Constant} \tag{34}$$

For the density of equation (33), these isometric surfaces are the ellipsoids

$$(\xi - \xi_t)^* M(\xi_t)(\xi - \xi_t) = \text{Constant} \tag{35}$$

The constant in equation (35) is most commonly chosen to be 1. This gives a 1σ (63.2 percent) probability that $(\hat{\xi}, X)X$ lies within the ellipsoid for any fixed unit vector X. The expression $(\hat{\xi}, X)X$ is the component of $\hat{\xi}$ in the direction given by X. If given that the distribution is Gaussian, this ellipsoid uniquely defines the density function. If repeated samples were taken and plotted, the resulting scatter would approximate the shape and size of the ellipsoid. It is easy to prove that this ellipsoid has the least volume of all regions that have an equal probability of containing $\hat{\xi}$.

When actual flight data are analyzed, the value of ξ_t is not known, so the ellipsoid described above cannot be drawn. This problem is handled by reversing the roles of ξ_t and $\hat{\xi}$. First, substitute $(\xi - \hat{\xi})$ for $(\xi - \xi_t)$. This centers the ellipsoid at $\hat{\xi}$. One can then define the probability that the ellipsoid covers ξ_t. Attention must be paid to the semantics of this statement, because ξ_t is not a random variable. Second, substitute $M(\hat{\xi})$ for $M(\xi_t)$; it is assumed that $M(\hat{\xi})$ is a good approximation of $M(\xi_t)$. This assumption is usually found to be good in practice, even when $\hat{\xi}$ is fairly far from ξ_t, because $M(\xi)$ tends to be relatively insensitive to changes in ξ for such cases.

The ellipsoid defined by

$$(\xi - \hat{\xi})^* M(\hat{\xi})(\xi - \hat{\xi}) = 1 \tag{36}$$

is called the uncertainty ellipsoid. As discussed in the previous section, $M(\hat{\xi})$ can be well approximated by H. This paper uses H to define the ellipsoid because M is difficult to compute except when it equals H exactly. The uncertainty ellipsoid as used in this report is thus defined by

$$(\xi - \hat{\xi})^* H(\xi - \hat{\xi}) = 1 \tag{37}$$

This ellipsoid gives a theoretical prediction, on the basis of the estimates from one maneuver, of the shape and size of the scatter plot.

The form of the maximum likelihood estimator suggests an alternate, non-statistical derivation of the uncertainty ellipsoid. The derivation is heuristic, but enlightening. The maximum likelihood estimates are obtained by minimizing the negative log likelihood functional, which can be expanded about the estimate as follows:

$$J(\xi) = J(\hat{\xi}) + \frac{1}{2}(\xi - \hat{\xi})^* \nabla_{\xi}^2 J(\hat{\xi})(\xi - \hat{\xi}) + \text{Higher order terms} \tag{38}$$

Since ξ_t will almost never equal $\hat{\xi}$, $J(\xi_t)$ will not be the minimum value of the cost function. It is reasonable, however, to assume that $J(\xi_t)$ will be close to the minimum. More precisely, suppose that $J(\xi_t) - J(\hat{\xi})$ will usually be less than some constant. This situation is closely related to the suboptimal control problem, where the exact optimum is not required, but performance is required to be within a specified range of the optimum. If terms of higher than second order are ignored, this condition is written

$$\frac{1}{2}(\xi_t - \hat{\xi})^* \nabla_{\xi}^2 J(\hat{\xi})(\xi_t - \hat{\xi}) \leq \text{Constant} \tag{39}$$

This inequality describes the interior of an ellipsoid. It was previously stated that $\nabla_{\xi}^2 J(\xi)$ closely approximates H, so this ellipsoid is essentially the same as the uncertainty ellipsoid if the constant is chosen to be $\frac{1}{2}$. Thus, the uncertainty ellipsoid can be thought of as the result of approximately maximizing the likelihood functional. This approximate maximization is a form of a sensitivity analysis based on the principle that two values of ξ cannot be reliably distinguished unless they result in sufficient differences in the likelihood functional.

Figure 3 is a one-dimensional picture of the construction of the uncertainty ellipsoid. In this case, the "ellipsoid" is the line segment from ξ_{min} to ξ_{max}, which is actually a degenerate one-dimensional case of an ellipsoid. Figure 4 shows the same construction with two unknowns. The interior of the uncertainty ellipse is shaded. Most of the concepts discussed below can be illustrated by using such a two-dimensional ellipse.

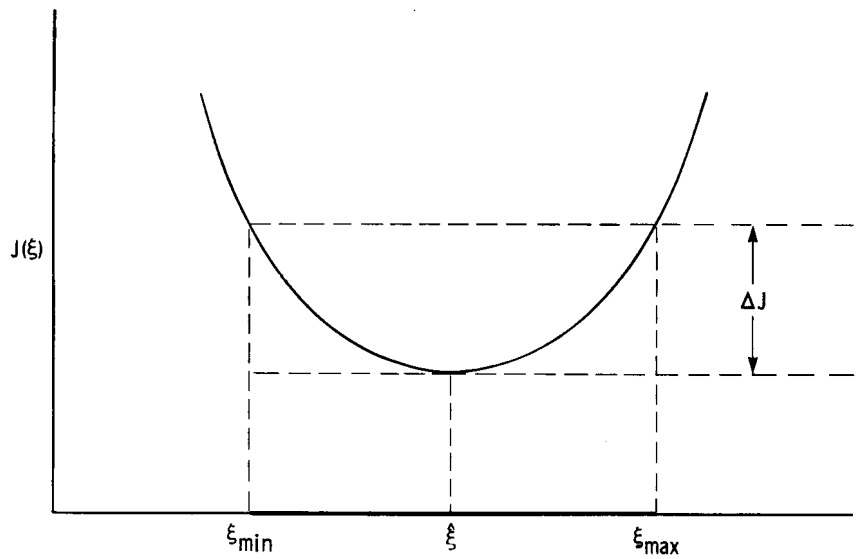


Figure 3. Construction of one-dimensional uncertainty ellipsoid.

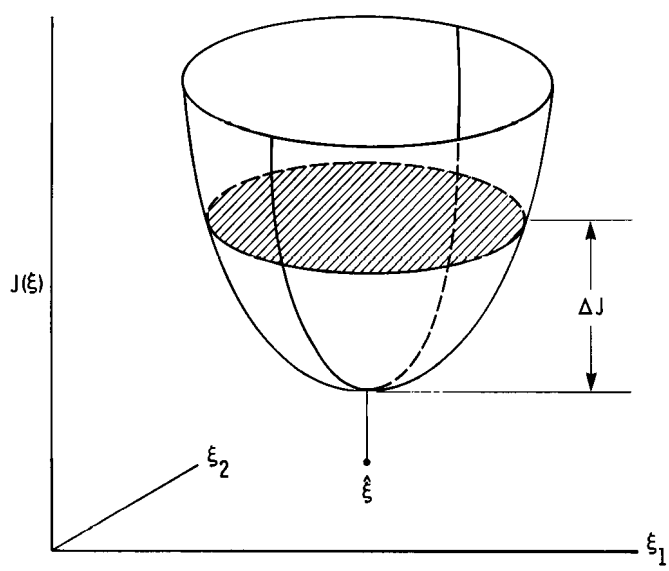


Figure 4. Construction of two-dimensional uncertainty ellipsoid.

The uncertainty ellipsoid gives the most complete picture of the theoretical accuracy of the estimates. It is difficult, however, to display the information given by the ellipsoid on a two-dimensional sheet of paper. Cross sections of the ellipsoid can be informative, but choosing the appropriate cross sections to examine is difficult, because there are typically from 10 to 30 parameters; that is, the ellipsoid is 10 to 30 dimensional. The problem of presenting the results in an understandable form is further complicated when many maneuvers are used, as would be required to investigate the full flight envelope of a high performance vehicle.

To make the most effective use of the uncertainty ellipsoid, it is necessary to find a simplified format for presenting the information. The next three sections of the report discuss the most commonly used of these formats: sensitivity, correlation, and Cramér-Rao bound.

Sensitivity

Sensitivity is a measure of how much the fit error, J, increases for a given change in a parameter. The sensitivity can be defined for any estimator, such as maximum likelihood, that involves minimizing a cost function. The inverse of the sensitivity, called the insensitivity, is a measure of how much a parameter can be changed from the estimated value without causing the fit error to increase by more than a given amount. All of the other parameters are held fixed at the estimates during this change. The insensitivity is more convenient than the sensitivity for purposes of graphic presentation and comparison with other measures.

Instead of actually recomputing the fit error for parameter changes, the insensitivity is generally defined by a second-order approximation (eq. (38)). As discussed above, the restrictions on the fit error can be (approximately) interpreted geometrically by constraining the parameters to lie within the uncertainty ellipsoid. An expression for the insensitivity is easy to obtain from the equation for the uncertainty ellipsoid (eq. (37)). Investigate changes in $\hat{\xi}$ along the direction of the ith parameter—that is, $\xi - \hat{\xi} = \alpha e_i$, where e_i is a unit vector in the direction of the ith parameter and α is an arbitrary scalar. To stay inside the ellipsoid, the following must be true:

$$\alpha^2 e_i^* H e_i \leq 1 \tag{40}$$

Since

$$e_i^* H e_i = H_{ii} \geq 0 \tag{41}$$

then

$$|\alpha| \leq \left(H_{ii}\right)^{-1/2} \tag{42}$$

The maximum magnitude of α is given by the equality, which occurs at the intercept of the e_i axis with the ellipsoid. This value is defined as the insensitivity with respect to ξ_i, and the inverse of this value is the sensitivity. It should be noted that

the insensitivity is often defined as $\left[\nabla_{\xi}^{2} J(\xi)_{ii}\right]^{-1/2}$ instead of $\left(H_{ii}\right)^{-1/2}$, but the difference is generally of small import, and the definition using H is more convenient for the purposes of this paper. Figure 5 illustrates the geometric interpretation of the insensitivity for a problem with two unknowns. For convenience, the origin has been redefined to be the center of the ellipse, which is the maximum likelihood estimate. The insensitivities of ξ_1 and ξ_2 are labeled as S_1 and S_2.

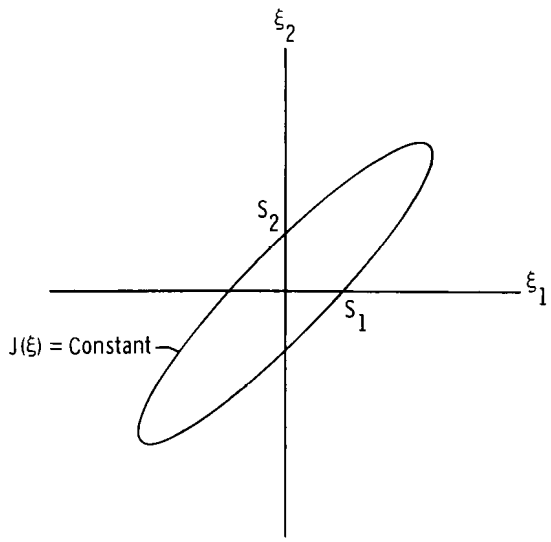


Figure 5. Geometric interpretation of insensitivity.

Statistically, the insensitivity is approximately the conditional standard deviation of the parameter estimate, given that all of the other parameters are known. The expression for this conditional standard deviation is based on the Cramér-Rao inequality. It was shown earlier that the covariance matrix of the estimates can be approximated by H^{-1} and that the estimates are approximately Gaussian. Letting ξ_o represent all of the parameters except ξ_i and using Bayes' rule (refs. 6 and 7), the conditional density of $\hat{\xi}_i$ given $\hat{\xi}_o$ is

$$p\left(\hat{\xi}_i \,|\, \hat{\xi}_o\right) = \frac{p\left(\hat{\xi}_i, \hat{\xi}_o\right)}{p\left(\hat{\xi}_o\right)} \tag{43}$$

This is to be evaluated at a known value of $\hat{\xi}_o$. For a known value of $\hat{\xi}_o$, $p\left(\hat{\xi}_o\right)$ is a constant, and $p\left(\hat{\xi}_i, \hat{\xi}_o\right)$ is given by equation (33). Thus,

$$p\left(\hat{\xi}_i \,|\, \hat{\xi}_o\right) = \text{Constant} \exp\left[-\frac{1}{2}\left(\hat{\xi}_i - \xi_{i_t}\right)^{*} H_{ii}\left(\hat{\xi}_i - \xi_{i_t}\right)\right] \tag{44}$$

where H and the constant depend on $\hat{\xi}_o$. The Cramér-Rao inequality thus gives the approximation $\left(H_{ii}\right)^{-1/2}$ for the conditional standard deviation of $\hat{\xi}_i$, which is equal to the insensitivity.

The insensitivity is a reasonable measure of accuracy only when a single parameter is being estimated, because the estimates of the other parameters are never exact, as the analysis assumes. Any effect of correlation between parameters is ignored; in fact, the insensitivity to a given parameter can be evaluated without even knowing what other parameters are being estimated. When several parameters are estimated, the insensitivity gives only a lower bound for the error band. The error band is always at least as large as the insensitivity regardless of what other parameters are simultaneously estimated; correlation effects between the parameters can increase, but never decrease, the error band.

In practice, correlation effects tend to increase the error band so much that the insensitivity is virtually useless as an indicator of accuracy. An insensitivity analysis will detect the obvious problems that exist when a parameter has little effect on aircraft response; an example of such a problem would arise if an attempt were made to estimate a control derivative for a control that was zero for an entire maneuver. The insensitivity will not indicate, however, when the effect of a parameter cannot be distinguished from the effects of other parameters, a much more common problem. The simplest example of the inadequacy of the insensitivity is the estimation of a control derivative of a control that is fixed at a constant but nonzero value for an entire maneuver. The insensitivity would indicate there to be no problem in estimating the derivatives because the control derivatives do affect aircraft response. The control derivatives cannot be estimated, however, because their effect cannot be distinguished from the effects of biases.

Because of these well known problems, the insensitivity is seldom used to gage the accuracy of parameter estimates. The insensitivity is useful only as a one-dimensional tool, whereas typical aircraft problems are 10 to 30 dimensional.

Correlation

The inadequacy of the insensitivity as a measure of estimate accuracy has led to the widespread use of parameter correlations. It was noted in the previous section that the correlations among parameters result in much larger error bands than the insensitivities. The use of correlations as indicators of accuracy has therefore been advanced (refs. 17 to 20).

The correlation between two parameters ξ_i and ξ_j is defined as $E\left\{\left(\hat{\xi}_i - \xi_{i_t}\right)\left(\hat{\xi}_j - \xi_{j_t}\right)\right\} \Big/ \sqrt{E\left\{\left(\hat{\xi}_i - \xi_{i_t}\right)^2\right\} E\left\{\left(\hat{\xi}_j - \xi_{j_t}\right)^2\right\}}$. In other words, it is an off-diagonal element of the covariance matrix of the estimate, normalized by the corresponding diagonal elements. For maximum likelihood estimates, the covariance matrix is approximated by H^{-1} using the Cramér-Rao inequality. The resulting approximation for the correlation between ξ_i and ξ_j is $\left(H^{-1}\right)_{ij} \Big/ \sqrt{\left(H^{-1}\right)_{ii}\left(H^{-1}\right)_{jj}}$. The

quantities computed from H are often referred to as the correlations (the qualifier "approximate" is omitted). The limitations of these quantities are better understood if it is realized that they only approximate the true correlations of the estimates. Geometrically, the correlations are related to the eccentricity of the uncertainty ellipsoid, provided that the sensitivities of the unknown parameters are all equal. (This can always be made true by a scale change.) The closer the magnitudes of the correlations are to 1, the more eccentric the scaled uncertainty ellipsoid becomes. The magnitude of the correlations can never exceed 1, except by round-off errors in the computation.

Occasionally, the conditional correlations are used. The conditional correlations are defined the same as correlations except that all of the expectations are taken on the assumption that all except the two parameters under consideration are known. (Statistically speaking, the expectations are conditioned on the other parameters.) For maximum likelihood estimates, the conditional correlation would be approximated by $-H_{ij} \Big/ \sqrt{H_{ii} H_{jj}}$, the same expression as before except that H^{-1} is replaced by H and the sign has changed. The conditional correlations are somewhat easier to compute than the full correlations, since it is not necessary to invert H. The computational cost of inverting H is usually negligible, however. If there are only two unknowns, the full and conditional correlations are identical. If there are more than two unknowns, the conditional correlation can give a quite different picture than the full correlation. A well known example is the case when H is an N-by-N matrix with 1's on the diagonal and all of the off-diagonal elements are equal to X. As X (the conditional correlation) approaches $\frac{-1}{N-1}$, the full correlation approaches 1. In the limit, when X equals $\frac{-1}{N-1}$, the matrix H is singular. Thus, for large N, the full correlations can be quite high even when the conditional correlations are low. This example can easily be turned around to show that the converse is also true. All of the comments made herein about the correlations also apply to the conditional correlations if H^{-1} is replaced by H.

There are three objections to using the correlations as a measure of accuracy. First, although the correlations give information about the shape of the uncertainty ellipsoid, they completely ignore its size. Figure 6 shows two uncertainty ellipses. Ellipse A is completely contained within ellipse B, and is therefore clearly preferable; yet ellipse B has zero correlation, and ellipse A has high correlation. From this example, it is obvious that excellent estimates can have high correlations and poor estimates can have low correlations. In order to evaluate the quality of the estimates, information is needed about the sensitivities as well as the correlations. Neither is adequate alone.

Examples of the problem shown in figure 6 are easily found in realistic situations. Figure 7(a) shows a typical elevator pulse input. The steady-state elevator deflection at this flight condition is 10°. A large portion of the energy in this input is from the steady-state value of 10°; the energy in the pulse is much smaller. Thus, the elevator derivative estimates will be highly correlated with the constant bias estimates. Because of the high energy of the total input signal, the sensitivity to the elevator derivative estimates will be high. The high sensitivity compensates for the high correlation.

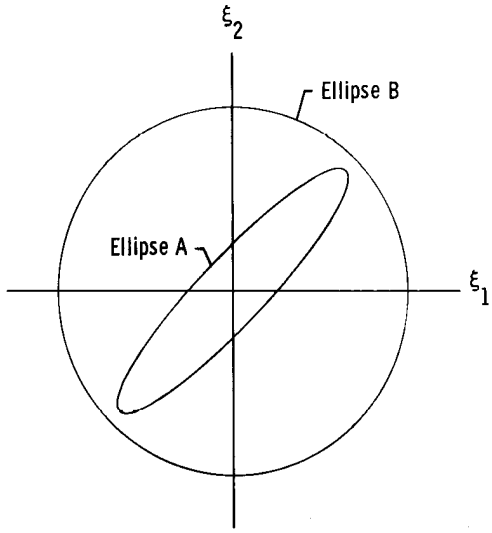
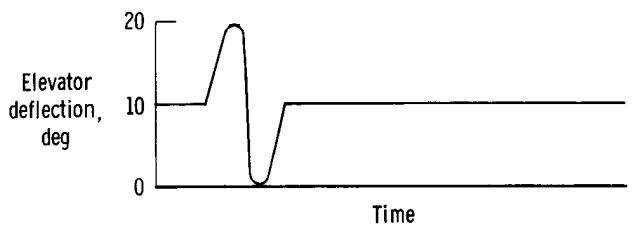
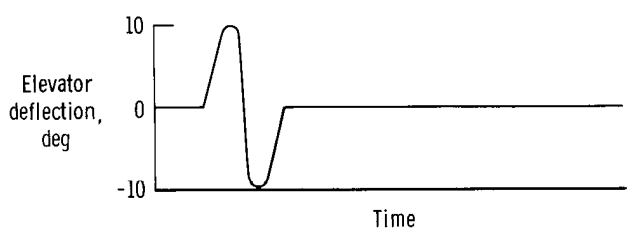


Figure 6. Need for sensitivity information.

Figure 7(b) shows the same input as in figure 7(a), except that the zero reference line has been redefined so that the steady-state value of the elevator input is 0°; this redefinition would correspond to analyzing the maneuver with perturbation equations. The correlation between the elevator derivative and bias estimates will be low for the input in figure 7(b). The sensitivities will also be lower for the input in figure 7(b) than for figure 7(a), balancing the change in the correlations.



(a) High correlation, high sensitivity.



(b) Low correlation, low sensitivity.

Figure 7. Interaction of correlation and sensitivity.

Figures 7(a) and 7(b) illustrate that the correlation alone is not a reasonable measure of accuracy. The correlation in this example can be changed at will to any value from -1 to 1 simply by redefining the reference axis used for analysis. Of course, the estimates, and thus their accuracy, are not really affected by this change in axis. The change in the correlation is balanced by a corresponding change in the sensitivity.

The second objection to the use of the correlations is more serious because it cannot be dealt with simply by looking at sensitivities and correlations together. In the same way that the sensitivities are one-dimensional tools, the correlations are two-dimensional tools. Linear dependence of higher order is not manifested in the correlations. Since aircraft stability and control problems typically involve 10 to 30 parameters, the utility of a tool restricted to two-dimensional subspaces is limited. Three simple examples of realistic situations serve to illustrate the dimensional limitations of the correlations. All of these examples involve free lateral-directional oscillation without pilot inputs.

For the first example, suppose that there is a yaw rate feedback to the rudder and an aileron-to-rudder interconnect. Both the aileron and the rudder signals are, therefore, proportional to yaw rate. In this case, the conditional correlations of the aileron, rudder, and yaw rate derivatives are 1 (or nearly so with imperfect data). Conditioned on the aileron derivatives being known exactly, changes in the rudder derivative estimates can be exactly compensated for by changes in the yaw rate derivative estimates; thus, the conditional correlation is 1. The unconditional correlations, however, are easily seen to be only $\frac{1}{2}$. In general, changes in the rudder derivative estimates must be compensated for by some combination of changes in the aileron and yaw rate derivative estimates. Since there are no constraints on the proportions of the compensation to come from the aileron and the yaw rate derivative estimates, the unconditional correlations would be $\frac{1}{2}$ (because, on the average, $\frac{1}{2}$ of the compensation would come from each source).

For the second example, suppose that there is no feedback and that there is a neutrally damped Dutch roll oscillation (or a wing rock). The signals β, p, and r are thus all sinusoids of the same frequency, with different phases and amplitudes. Taken two at a time, these signals have low correlations. The conditional correlations consider only two parameters at a time, so the conditional correlations of the β, p, and r derivatives are all low. Nonetheless, the β, p, and r signals are linearly dependent if they are all considered together, because they can all be written as linear combinations of a sine wave and a cosine wave at the Dutch roll frequency. Thus, the unconditional correlations of the β, p, and r derivatives are 1 (or nearly so with imperfect data).

Both of the examples above have three-dimensional correlation problems that prevent the parameters from being identifiable. The conditional correlations are low in one case, and the unconditional correlations are low in the other. Although either alone is insufficient, examination of both the conditional and unconditional correlations will always reveal three-dimensional correlation problems.

For the third example, suppose that a wing leveler feeds back bank angle to the ailerons and that a neutrally damped Dutch roll is present with the feedback on. There are then four pertinent signals (β, p, r, and δ_a) that are sinusoids with the same frequency and different phases. In this case, both the conditional and the unconditional correlations are low. Nonetheless, there is a correlation problem, and it causes some parameters to be unidentifiable. This correlation problem is four dimensional, and it cannot be seen by using the two-dimensional correlations.

The estimated correlations or conditional correlations are closely related to the eigenvalues (ref. 21) of 2-by-2 submatrices of the matrices H^{-1} or H, respectively, normalized to have unity diagonal elements. Specifically, the eigenvalues are 1 plus the correlation and 1 minus the correlation; thus, high correlations correspond to large eigenvalue spreads. Correlations of higher order would be investigated by using the eigenvalues of larger submatrices. If it is looked at in this light, the investigation of 2-by-2 submatrices is revealed as an arbitrary choice dictated more by familiarity than by any objective criterion. The eigenvalues of the full normalized H- or H^{-1}-matrix would seem more appropriate tools. These eigenvalues and the corresponding eigenvectors can provide some information, but they are seldom used. In principle, small eigenvalues of the normalized H-matrix or large eigenvalues of the normalized H^{-1}-matrix indicate correlations among the parameters with significant components in the corresponding eigenvectors. Note that the eigenvalues of the unnormalized H- and H^{-1}-matrices are of very little use in studying correlations, because scaling effects tend to dominate.

The last objection to the use of the correlations is the difficulty of presentation. For obvious reasons, no attempt has ever been made to display the estimated correlations between all of the parameters graphically. The usual presentation is simply a printing of the estimated correlation matrix. This is feasible when the number of maneuvers used is very small. If a flight test program with many maneuvers is used to investigate the behavior of a vehicle, however, printing all of the estimated correlation matrices would be pointless: the data could not be incorporated into a coherent picture of the vehicle's flight characteristics.

Cramér-Rao Bound

The Cramér-Rao bound is also based on the uncertainty ellipsoid. This measure of accuracy is often referred to by other names, including estimated variance and standard deviation. These quantities are the same as the Cramér-Rao bound, except that the variance is the Cramér-Rao bound squared. Cramér-Rao bounds are also equivalent to F-statistics.

By the definition of the covariance matrix, the standard deviation of the ith parameter is the square root of the ith diagonal element. For maximum likelihood estimates, the Cramér-Rao inequality gives the approximation H^{-1} for the covariance matrix. The quantity $\sqrt{\left(H^{-1}\right)_{ii}}$ is called the Cramér-Rao bound. (There is some justification for defining the bound as $\sqrt{\left(M^{-1}\right)_{ii}}$, but M is difficult to compute except

when it is equal to H; thus the definition used here.) The Cramér-Rao bounds are often referred to as standard deviations without the qualifier "approximate". The limitations of Cramér-Rao bounds are better understood if it is realized that they are approximations of the true standard deviations of the estimates.

The approximation of the standard deviation discussed here should not be confused with the sample standard deviation, although they are closely related in that both are estimates of the same quantity. The discussion in this section concerns an estimate of the true standard deviation that is based on the theoretical assumptions made about the system, and is computed for each data point using the Cramér-Rao inequality. The sample standard deviation is an estimate of the same quantity, but it is based on the observed scatter of the data; no assumptions about the system are made, and only a single value is computed on the basis of all of the data points. Naturally, since the Cramér-Rao bound and the sample standard deviation are different estimates of the same quantity, it is worthwhile to compare them.

The Cramér-Rao bounds are closely related to the insensitivities. Recall that the insensitivity is approximately the conditional standard deviation of the parameter estimate, given that all of the other parameters are known. The Cramér-Rao bound is an approximation of the unconditional standard deviation. Whenever the term "standard deviation" is used, it is assumed that it is unconditional unless otherwise specified. The computational relationship is also interesting. The Cramér-Rao bound is $\sqrt{\left(H^{-1}\right)_{ii}}$, whereas the insensitivity is $\sqrt{\left(H_{ii}\right)^{-1}}$. The geometric relationship between the quantities is often overlooked. The insensitivity can be defined as the solution to an optimization problem: maximize $\left|\xi_i - \hat{\xi}_i\right|$ while staying within the uncertainty ellipsoid and holding all of the other parameters fixed at the estimates. The Cramér-Rao bound is the solution to the same problem without holding the other parameters fixed. The removal of the restrictions on the optimization is directly analogous to removing the statistical conditioning. A two-dimensional sketch of the geometric relationship between the two measures is shown in figure 8. The insensitivities of ξ_1 and ξ_2 are labeled S_1 and S_2, while the Cramér-Rao bounds are labeled C_1 and C_2.

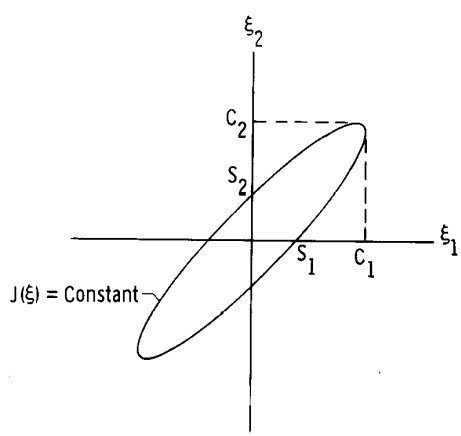


Figure 8. Geometric relationship between Cramér-Rao bounds and insensitivities.

To prove that the Cramér-Rao bound is the solution to the optimization problem above, we will state and prove a more general result. The general result is actually easier to prove.

THEOREM: *Given a fixed vector y and a positive definite symmetric matrix H, then the maximum of $x^* y$, subject to the constraint that $x^* Hx \leq 1$, is given by $\sqrt{y^* H^{-1} y}$.*

Proof: Since $x^* y$ has no unconstrained local extrema, the solution must lie on the constraint boundary, so the inequality in the constraint can be replaced by an equality. This constrained optimization problem can be restated by the use of Lagrange multipliers (ref. 22) as the unconstrained minimization of

$$f(x, \lambda) = x^* y - \frac{1}{2}\lambda(x^* Hx - 1) \tag{45}$$

where λ is the scalar Lagrange multiplier. The maximum is found by setting the gradients to zero, as follows:

$$0 = \nabla_x f(x, \lambda) = y - \lambda Hx \tag{46}$$

$$0 = \frac{\partial}{\partial \lambda} f(x, \lambda) = -\frac{1}{2}(x^* Hx - 1) \tag{47}$$

From equation (46),

$$x = \lambda^{-1} H^{-1} y \tag{48}$$

Substituting this into equation (47) gives

$$y^* H^{-1} \lambda^{-1} H\lambda^{-1} H^{-1} y - 1 = 0 \tag{49}$$

or

$$\lambda^{-2} y^* H^{-1} y = 1 \tag{50}$$

or

$$\lambda = \sqrt{y^* H^{-1} y} \tag{51}$$

Substituting into equation (48) gives

$$x = \frac{H^{-1} y}{\sqrt{y^* H^{-1} y}} \tag{52}$$

and thus

$$x^* y = \frac{y^* H^{-1} y}{\sqrt{y^* H^{-1} y}} = \sqrt{y^* H^{-1} y} \tag{53}$$

at the solution. □

For the specific problem, y is e_i, a unit vector along the axis of the ith parameter, and x is replaced by $\xi - \hat{\xi}$. We are then seeking to maximize

$$(\xi - \hat{\xi})^* e_i = \xi_i - \hat{\xi}_i \tag{54}$$

The solution is given by

$$\xi_i - \hat{\xi}_i = \sqrt{e_i^* H^{-1} e_i} = \sqrt{\left(H^{-1}\right)_{ii}} \tag{55}$$

which is the assertion we set out to prove. Note that the solution point is not, in general, along the e_i axis, in contrast to the sensitivity problem (see fig. 8). In fact, the solution point is

$$\xi - \hat{\xi} = \frac{H^{-1} e_i}{\sqrt{\left(H^{-1}\right)_{ii}}} \tag{56}$$

Recognize $H^{-1} e_i$ as the ith column of H^{-1}. This completes the solution of the specific problem.

The solution of the general problem has other applications. The value of any linear combination of the parameters can be expressed as $\xi^* y$ for some fixed y-vector. Thus, the general solution shows how to evaluate the accuracy of arbitrary linear combinations of the parameters. Such analysis might be used to show, for instance, that the estimate of $C_{\ell_p} + C_{\ell_{\delta_a}}$ was quite accurate, even though the individual estimates of C_{ℓ_p} and $C_{\ell_{\delta_a}}$ were poor. This situation would be expected if roll rate feedback provided the only aileron input. Reference 23 suggests that a sensitivity analysis be made of situations where the correlation is high. The Cramér-Rao bound provides exactly such a sensitivity analysis.

On the basis of this geometric picture, the Cramér-Rao bounds can be thought of as insensitivities that are computed accounting for all parameter correlations. The computation and interpretation are valid in any number of dimensions, unlike the sensitivities, which are one-dimensional tools, or the correlations, which are two-

dimensional tools. The Cramér-Rao bounds are thus the best of the theoretical measures of accuracy that can be evaluated for a single maneuver.

Some authors have advocated the examination of the sensitivities and estimated correlations in addition to or instead of the Cramér-Rao bounds in order to evaluate the accuracy of the estimates (refs. 17 to 20). However, the Cramér-Rao bounds already include all of the effects of sensitivities and correlations. The sensitivities and estimated correlations may be useful in finding the source of a problem indicated by large Cramér-Rao bounds, but they do not provide additional information as to whether a problem exists. Any argument about the validity of the theory used to compute the Cramér-Rao bounds must take into account that the same theory is used to compute the estimated correlations.

EVALUATION WITH FLIGHT DATA

Because of the numerous qualifications about modeling error and noise statistics, theoretical measures of accuracy must be validated before much confidence can be placed in them. The Cramér-Rao bounds have been advanced as the best of the theoretical measures of accuracy. The comparison of the Cramér-Rao bounds with the sample standard deviations obtained from the data scatter gives a good indication of the adequacy of the assumptions made in the theoretical development.

It has long been known that this comparison shows significant discrepancies if actual flight data are used. This section examines these discrepancies and advances an explanation for them. Approximate corrections for the problem are then suggested and evaluated.

Discrepancy in the Cramér-Rao Bound

Data from a PA-30 aircraft were chosen to evaluate the Cramér-Rao bounds. (This example was previously reported in ref. 24.) Data from 18 maneuvers made with this vehicle were obtained. Each maneuver consisted of an aileron input initiated from steady flight. The derivatives were estimated by the MMLE3 program (ref. 8) using an output error maximum likelihood method (measurement noise only). A typical match of the measured and estimated time histories is shown in figure 9.

Figure 10 presents the estimates of C_{n_β} and C_{ℓ_β} and the Cramér-Rao bounds obtained from these maneuvers. The vertical scales in these plots are exaggerated to show the scatter, and the angle of attack scale is exaggerated to separate the maneuvers. No significant differences in the derivatives are expected over this small angle of attack range; the 18 maneuvers can be regarded as being at essentially the same flight condition.

The Cramér-Rao bounds in this plot are so small that they are difficult to see; they are roughly the same size as the symbols. The data scatter is much greater than indicated by the Cramér-Rao bounds. Quantitatively, the sample standard deviation is about nine times the average Cramér-Rao bound. The existence of this discrepancy

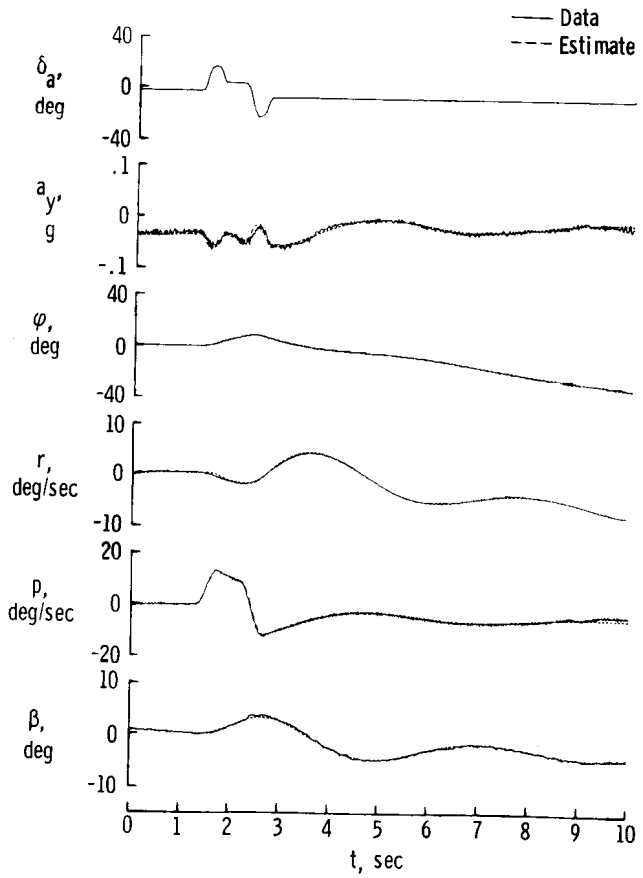


Figure 9. Comparison of measured and estimated time histories for flight data.

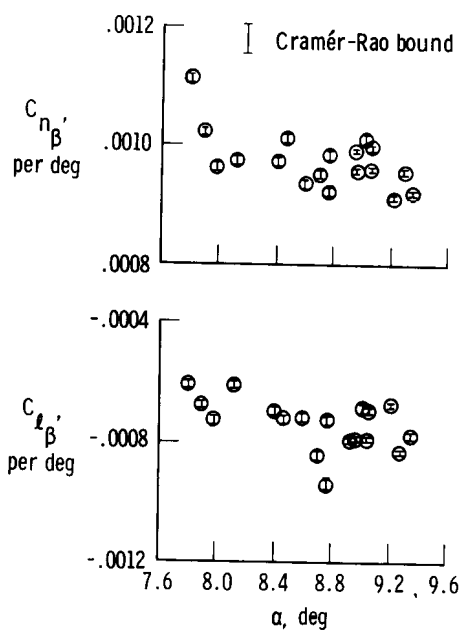


Figure 10. Estimates from flight data.

between the data scatter and the Cramér-Rao bound has long been known. It has become common practice to multiply the Cramér-Rao bounds by a "fudge factor" of 5 to 10 (ref. 24). The resulting values proved useful for evaluating the accuracy of estimates, but the necessity of the unexplained fudge factor detracted from the confidence. Because of this problem with the Cramér-Rao bounds, several reports used the estimated correlations as primary indicators of accuracy (refs. 20 and 25) in spite of the previously mentioned problems with the correlations. The estimated correlations are based on the same theoretical foundation as the Cramér-Rao bounds and thus should be equally suspect if errors are known to exist. Other reports ignored the fudge factor and quoted overly optimistic values of accuracy (ref. 18). Discrepancies larger than the quoted accuracy have then been attributed to various effects without sufficient data points to establish whether the observed differences were significant or lay within the scatter band.

The evaluation of accuracy measures with actual flight data is complicated by the impossibility of establishing true values for comparison and by the inevitable presence of unmodeled effects. Although tests with actual flight data are necessary for final validation, preliminary work is aided by the more controlled environment provided by simulated data. Therefore, the experiment described above was repeated with simulated data.

To mimic the flight experiment as closely as feasible, the control inputs measured from the flight data were used to create simulated data. Eighteen maneuvers were simulated, using the same flight conditions as for the 18 actual maneuvers. The same model was used for the simulation as for the estimation. The same true values of the nondimensional derivatives were used for all 18 maneuvers. A pseudorandom noise generator was used to add simulated white Gaussian measurement noise to the computed responses. The measurement noise power for each signal was proportional to the average residual power observed in the flight data for that signal. The proportionality constant was adjusted to obtain the same magnitude of scatter in the estimates from the simulated data as was observed in the flight data. The derivatives and Cramér-Rao bounds were estimated from the simulated data using the same program as for the flight data. A typical comparison of the simulated measured time history and the estimated one is shown in figure 11.

The estimates of C_{n_β} and C_{ℓ_β} from the simulated data are shown in figure 12 plotted with the same scales as those used in figure 10. The true value of C_{n_β} is 0.001, and that of C_{ℓ_β} is -0.0007.

The scatter in this plot is about the same as for the flight data because the simulated noise power was adjusted to achieve this scatter. The Cramér-Rao bounds for the flight and simulated data differ drastically. The Cramér-Rao bounds for the simulated data are about 10 times as great as for the flight data and agree well with the observed scatter. For the 11 derivatives estimated (not including bias terms), the ratios of the sample standard deviations to the Cramér-Rao bounds range from 0.68 to 1.32. This is excellent agreement for a sample of only 18 maneuvers.

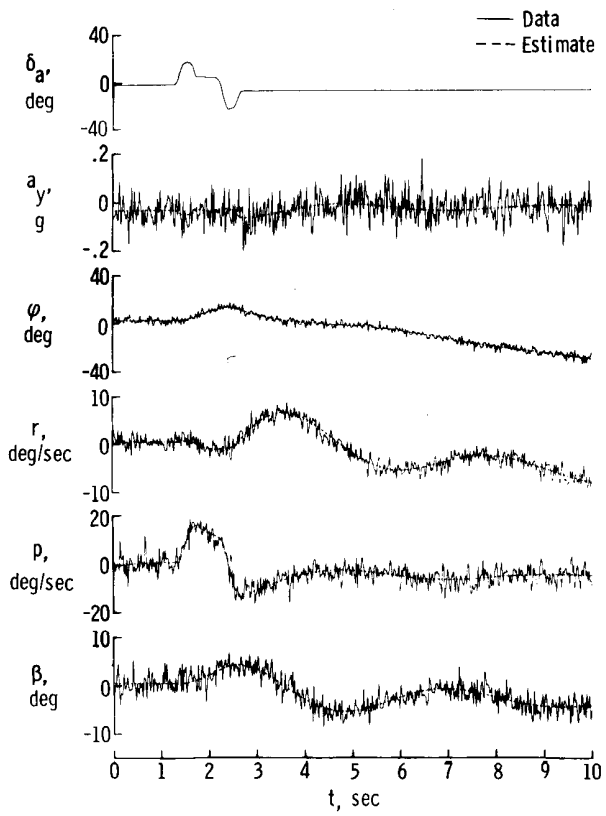


Figure 11. Comparisons of measured and estimated time histories for simulated data with white noise.

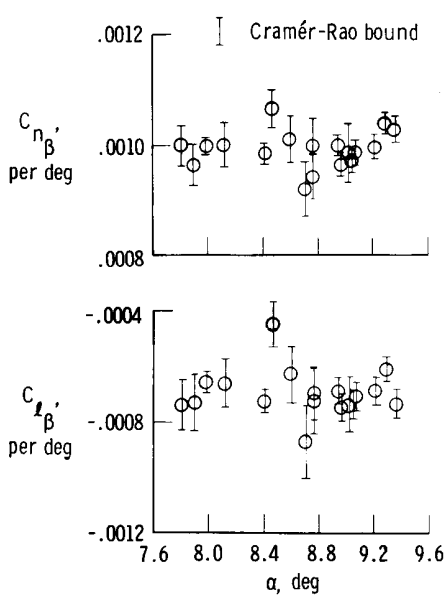


Figure 12. Estimates from simulated data with white noise.

The Cramér-Rao bound thus agrees very well with the scatter in the simulated data, but disagrees drastically with the flight data. The following sections examine the reasons for this discrepancy.

Previous Attempts at Explanation

Several explanations have been advanced to attempt to resolve the discrepancy in the Cramér-Rao bound. Few of these ideas have been discussed extensively in formal publications because of their speculative nature. As shown below, none of the proposals provides a satisfactory explanation for discrepancies of the magnitude observed in practice.

The excellent performance of the Cramér-Rao bound with simulated data invalidates many of the suggestions that could be advanced to explain the discrepancy. One such explanation would be an error in the formulation or computer programming. An error of this type would affect the simulated as well as the flight results, since the same computer program was used for both.

Another explanation is that the Cramér-Rao bound is only a lower bound. The maximum likelihood estimator has been proved to be asymptotically efficient; that is, the Cramér-Rao inequality approaches an equality as time approaches infinity. For finite data time the equality does not hold. Intuition, which can be backed up by analysis in this case, suggests that a few periods of the natural frequency should be enough for the asymptotic result to be closely approached. The fact that the simulated data agree so well with the scatter verifies that the time is long enough to make the equality a very good approximation.

The above attempts to explain the discrepancy between the Cramér-Rao bounds and the scatter implicitly assume that the scatter is a reasonable measure of accuracy. For the simulated data, where the true values are known to be constant, this statement is almost a definition of accuracy for an unbiased estimator. For the flight data, other possibilities must be considered. As stated above, the simulated data noise level was chosen to make the scatter match that of the flight data, and the resulting Cramér-Rao bounds of the simulated data were larger than those of the flight data. As is evident from a comparison of figures 9 and 11, the noise power in the simulated data is much larger than the power of the flight data residuals. The difference in the Cramér-Rao bounds arises directly from the difference in the noise power. If the simulated data noise power were reduced to the same level as that of the flight data residuals, the simulated and flight Cramér-Rao bounds would be the same. Of course, the scatter of the simulated data would be much less than that of the flight data.

This suggests the possibility that the Cramér-Rao bounds are not too small but rather that the scatter in the flight data is too large to represent the accuracy properly. It might be true, for instance, that the individual estimates are as accurate as indicated by the Cramér-Rao bounds and that the scatter reflects actual changes in the coefficients. In principle, this would explain the discrepancy. However, there is no physical reason to suspect such large variations in the aerodynamic derivatives at essentially the same flight condition. Furthermore, no ascertainable pattern can be detected in the scatter that relates to any flight condition parameter. Although only

a single example is shown here, the same discrepancy is noted in every class of vehicle tested, including small general aviation aircraft (ref. 24), airliners (ref. 26), military fighters (ref. 27), large supersonic aircraft (ref. 28), and unconventional vehicles (refs. 29 to 30). The universality of the discrepancy is strong evidence against the possibility that there are apparently random changes in the actual derivatives.

A related possibility is that the scatter in the estimates results from unmodeled errors that would not be reflected in the Cramér-Rao bounds. An obvious example of such a problem is an error in the measurement of the flight condition. For instance, if the dynamic pressure measurement were inaccurate, the Cramér-Rao bound and the estimates of the dimensional derivatives would not be affected. The nondimensional derivatives, however, would have larger errors than otherwise predicted. The occurrence of the same discrepancy with many different aircraft and data systems is strong evidence that this is not the case. Most of the data systems used are believed to be accurate enough to eliminate problems of error in the measurement of flight condition.

Thus, none of the suggestions advanced above provides a satisfactory explanation of the discrepancy observed in the flight data. For several years the necessity for the fudge factor on flight data was left essentially without explanation. It was argued that modeling errors existed which invalidated the Cramér-Rao bound. Such modeling errors, of course, were not present in the simulated data. Although this argument is virtually irrefutable, it does little to explain the problem. The subject of what types of modeling error might exist that would have such effects was not addressed. This argument, amounting to a dismissal of the problem, does not give any basis for confidence in the use of the fudge factor. Although some authors found the values to be of empirical use when the fudge factor was applied, others rejected the Cramér-Rao bound as invalid.

This state of affairs has more far-reaching implications than the invalidation of the Cramér-Rao bound for use with flight data. The theoretical derivation of the Cramér-Rao bound rests on the likelihood functional. If this theory is inadequate, the theoretical justification for the use of maximum likelihood estimators must also be questioned, since the estimators are based on the same likelihood functional. Thus an invalidation of the Cramér-Rao bound might imply an invalidation of the estimates, leaving the analyst with nothing theoretically worthwhile.

Explanation for the Discrepancy

In reference 31, the authors advanced the first satisfactory explanation for the discrepancy in the Cramér-Rao bound. The source of the discrepancy was traced to the theoretical assumptions about the independence of the noise samples.

The length of the period before this explanation was advanced reveals the need for researchers to be well grounded in both theory and practice, because the problem lay essentially in a lack of communication between the theoreticians and the practitioners. The theory is most naturally developed assuming independence of the noise samples, and the theory was passed to the practitioners in this form. Experi-

enced practitioners were well aware that residual spectra are seldom even close to white, but accepted the assumptions as necessary to the theory. In fact, the theory can easily give much information about the effects of colored noise. The theoreticians were unaware that such information was needed, and the practitioners were unaware that the information was available.

The exact discrete-time theory of estimation in the presence of colored noise is trivial when the spectral shape of the noise is known. The application is also easy in the frequency domain, but such considerations as nonlinearities and time variation severely limit the application of frequency domain estimation. The time domain application of the exact theory of estimation with colored noise is overly cumbersome and the effort involved is not justified by the small benefits expected. Furthermore, the spectral characteristics of the noise are seldom precisely known, and incorrect specification can actually make the estimates worse. The estimation of the spectral characteristics is possible in principle, but it adds further unacceptable computational complications. The theory of estimation with colored noise has therefore been largely relegated to textbook examples.

Abandon the exact approach, and recognize that approximating the effects of colored noise would still provide more useful results than ignoring the question. For a first approximation, assume that the noise is band limited white with band limit B. Then, instead of deriving a maximum likelihood estimator for this system, analyze the performance of the estimator based on white noise when the actual noise is colored in this manner. The results of this analysis (ref. 32) agree well with intuitive expectations. As long as the noise bandwidth is much larger than the system bandwidth, the effect on the estimates is negligible. Stated loosely, the estimation errors are caused by the noise near the natural system frequencies; the estimator will always mistake some percentage of this noise as actual system response and vice versa. Noise far above the system bandwidth is readily identified as noise rather than as system response. A good estimator should be little influenced by such high frequency noise.

This result is easy to generalize. It is obvious that the exact shape of the noise rolloff is of little consequence in the analysis above. If the analysis is repeated with different rolloff characteristics, the same results will be obtained. In short, the high frequency characteristics of the noise do not materially affect the estimates.

This conclusion is a welcome validation of the practice of using the maximum likelihood estimator derived on the basis of independent noise samples, even though the actual residuals are known to be correlated significantly. The quantitative interpretation of high frequency is somewhat difficult, and skepticism is prudent when the noise bandwidth nears the system bandwidth, as often occurs. Nonetheless, this theory provides a much stronger base than failing to address the question.

Since the estimates are essentially unaffected by high frequency noise, it immediately follows that all functions of the estimates are equally unaffected. Thus, in particular, the same expression for the Cramér-Rao bound should still be valid. This would seem to refute the hypothesis that the discrepancy in the Cramér-Rao bound is related to the noise spectrum; in fact, the relationship is so elementary that it has been overlooked.

In stating that the estimates are essentially unaffected by high frequency noise, we are comparing noise spectra that are equal to each other at low frequencies. The high frequency spectra, and thus the total noise power, may differ. The low frequency spectral density, rather than the total power, is the important statistic. However, all of the programs in use are written in terms of the total noise power (or, equivalently, the noise variance). Programs based on the continuous time theory use the spectral density, but the spectral density is estimated in practice by dividing the discrete total power estimate by the Nyquist frequency.

Let us consider the effects of using the total power instead of the spectral density. Imagine a system with total noise power R. If the noise samples are independent (the discrete time equivalent of white noise), the classic analysis is valid and the Cramér-Rao bounds should be correct. The noise power spectral density of this system is $2R\Delta t$, since the noise spectrum is flat out to the Nyquist frequency $\frac{1}{2\Delta t}$. Now imagine a second system with the same total power, but with a one-sided noise bandwidth of B. Figure 13 shows the noise spectra of these two systems.

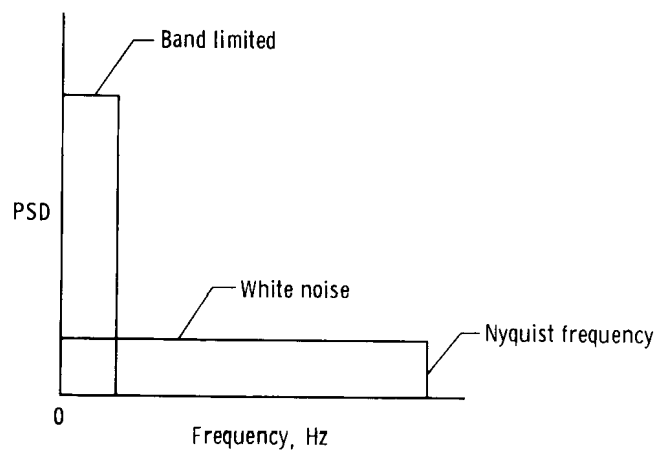


Figure 13. Assumed noise spectra.

The Cramér-Rao inequality for both systems, expressed in the usual way, is

$$\operatorname{var}(\hat{\xi}) \geq \left[\sum_{i=1}^{N} \left(\nabla_{\xi}\, \tilde{z}^{*}\right) R^{-1} \left(\nabla_{\xi}\, \tilde{z}^{*}\right)^{*}\right]^{-1} \tag{57}$$

where $\nabla_{\xi}\, \tilde{z}^{*}$ does not depend on the noise statistics. Since the total power, R, is the same for both systems, the same values are computed for Cramér-Rao bounds. This is the computation used by current programs. Expressed in terms of the power spectral density GG^{*}, the Cramér-Rao inequality should be

$$\operatorname{var}(\hat{\xi}) \geq \left[\sum_{i=1}^{N} \left(\nabla_{\xi}\, \tilde{z}^{*}\right)\left(\frac{1}{2\Delta t}\, GG^{*}\right)^{-1} \left(\nabla_{\xi}\, \tilde{z}^{*}\right)^{*}\right]^{-1} \tag{58}$$

For the first system, $\frac{1}{2\Delta t}\, GG^*$ equals R, so this computation is the same as the previous one. For the second system, however, $\frac{1}{2\Delta t}\, GG^*$ equals $\frac{1}{2B\Delta t}\, R$; thus

$$\operatorname{var}(\hat{\xi}) \geq \frac{1}{2B\Delta t}\left[\sum_{i=1}^{N} \left(\nabla_{\xi}\, \tilde{z}^{*}\right) R^{-1} \left(\nabla_{\xi}\, \tilde{z}^{*}\right)^{*}\right]^{-1} \tag{59}$$

The Cramér-Rao bound computation based on the total power was, therefore, too small by a factor of $\frac{1}{2B\Delta t}$ for the variance (or $\sqrt{\frac{1}{2B\Delta t}}$ for the standard deviation). The Cramér-Rao bound computation based on the total noise power assumes that the power is spread evenly over the Nyquist range, as in the first case. Thus, in equation (59), B is assumed to be $\frac{1}{2\Delta t}$. If this assumption is incorrect, the resulting computation of variance will be proportionally incorrect.

These results provide an explanation for the discrepancies previously observed. The noise samples in the simulated data were independent; thus, the corresponding Cramér-Rao bound computations were valid. Modern flight test instrumentation is accurate enough that the largest component of the residual error in flight data is from the modeling error rather than true measurement error. The statistics for so-called "measurement noise" must include all such unmodeled effects. The philosophical question of what to include in "measurement noise" is not addressed in detail in this paper. The classical viewpoint on this subject is propounded by Cramér (ref. 5), from whom the following is quoted.

> It is, of course, clear that there is no sharp distinction between these various modes of randomness. Whether we ascribe e.g. the fluctuations observed in the results of a series of shots at a target mainly to small variations in the initial state of the projectile, to the complicated nature of the ballistic laws, or to the action of small disturbing factors, is largely a matter of taste. The essential thing is that, in all cases where one or more of these circumstances are present, an exact prediction of the results of individual experiments becomes impossible, and the irregular fluctuations characteristic of random experiments will appear.
>
> We shall now see that, in cases of this character, there appears amidst all irregularity of fluctuations a certain typical form of regularity, that will serve as the basis of the mathematical theory of statistics.

Suffice it to say that since the theory under discussion contains only the linear system model and measurement noise, anything not included in the system model must be considered part of "measurement noise". Otherwise the theory denies its existence; such solipsism seems unwise. The "measurement noise" in real flight test data therefore tends to be quite colored. Figure 14 shows power spectral density plots of the residuals from the flight test data used in figure 10. A precise break point is not obvious, but a value in the neighborhood of 0.3 hertz seems reason-

able. The Nyquist frequency for this 50-sample-per-second data is 25 hertz, so the computed Cramér-Rao bounds should be increased by a factor of $\sqrt{\frac{25}{0.3}}$, or about 9. Increased by this factor, the agreement between the Cramér-Rao bounds and the scatter is quite reasonable, considering the vaguely defined value of the break frequency. The noise break frequency is close enough to the system frequencies so that the approximations in the theory are subject to question, but the experimental results hold up well.

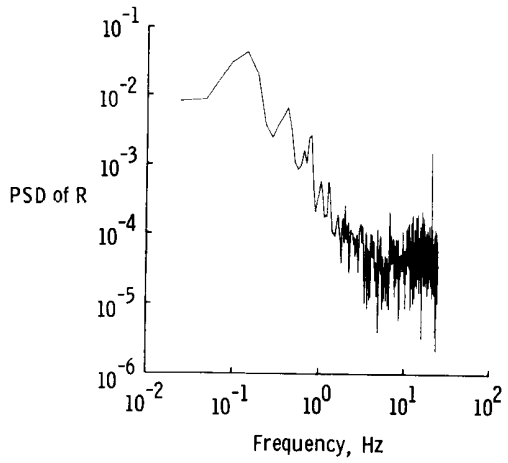


Figure 14. Power spectral density of yaw rate residual from flight data.

To verify the theoretical results described in this section, a new set of simulated data was created. The noise for these data was created by passing the pseudorandom independent noise through a fifth order Chebychev filter with a break frequency of 1 hertz. This filter has a sharp break at 1 hertz as shown by figure 15, which is a

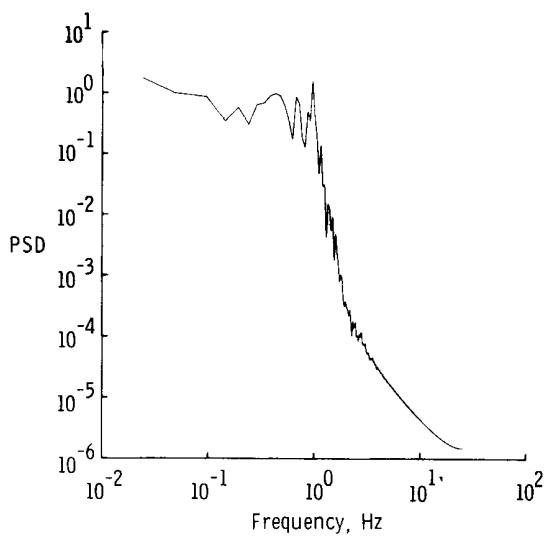


Figure 15. Power spectral density of simulated colored noise.

power spectral density of one of the resulting measurement noise signals. A typical time history match to these data is shown in figure 16. Note that this match exhibits deterministic-appearing characteristics such as phase shift and flattened peaks.

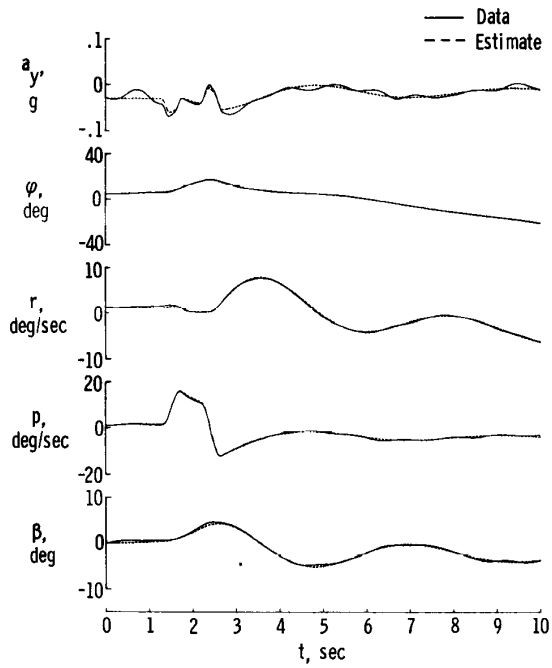


Figure 16. Comparison of measured and estimated time histories for simulated data with filtered noise.

The matches of β, p, r, and φ are more typical of flight data than the simulations shown in figure 11. Accelerometers, by their nature, have more high frequency noise than such instruments as rate gyros. Therefore, the flight accelerometer matches tend to be of a character intermediate between that of figures 11 and 16. The Cramér-Rao bounds and estimates of C_{n_β} and C_{ℓ_β} from the simulated data with filtered noise are shown in figure 17. This figure shows that the discrepancy has been duplicated successfully with simulated data by using band-limited noise. The band limit is well defined for the simulated data, and when the Cramér-Rao bounds are corrected for the colored noise, they agree excellently with the scatter.

These results support the conclusion that the discrepancy in the Cramér-Rao bound is adequately explained by the presence of colored noise. To permit the Cramér-Rao bound to reflect flight data scatter accurately, the noise statistics used in computing the Cramér-Rao bound must represent the entire residual error, including the contribution of modeling error (which may be much larger than the actual instrumentation error). For this reason, studies that use Cramér-Rao bounds based solely

on instrumentation characteristics (refs. 10 to 12) are likely to be extremely over-optimistic. Colored noise would also be expected to affect the insensitivities, but should not directly affect the correlations. The correlations might, however, be affected by the fact that the noise contributions from modeling errors in the various signals tend to be correlated.

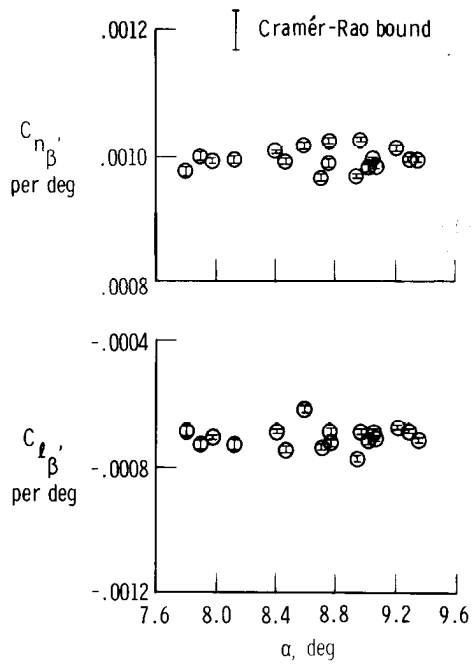


Figure 17. Estimates from simulated data with colored noise.

Suggested Implementations

The previous section explained the reasons for the discrepancies observed in the Cramér-Rao bounds. It remains to discuss the practical implementation of a corrected computation of the bound. Three approaches are considered in this section.

The first approach is to continue to use the fudge factors. The Cramér-Rao bounds computed ignoring noise coloring and then multiplied by a fudge factor of 5 to 10 have proved to be useful in practice. The objection to this approach has been not to its utility, but rather to its ad hoc nature and lack of theoretical justification. Now that the theory has provided an understanding of the need for the extra term, it may be reasonable to use a value for this term based on past experience rather than analytically compute one for each maneuver. It could be looked on as an empirically determined spectral adjustment factor instead of as a mysterious fudge factor. The factor has been observed to be relatively constant over large classes of cases, which tends to justify this approach. The advantages of the approach are simplicity and the fact

that no changes to current programs are required. The disadvantage is that the engineer must watch for changes in the vehicle or in the analysis which might significantly affect the spectral characteristics of the residuals and thus the factor used. If such changes occur, or if discrepancies are noted, it may be necessary to adjust the factor used. The approach may be criticized as arbitrary, but when considered as an aid to an engineering evaluation, instead of as an absolute value of accuracy, it has been in the past and can continue to be useful.

The second approach is to examine, either manually or automatically, the actual spectrum of the residuals. The break frequency can be evaluated, or the spectral density can be used directly. The advantage of this method is that it provides the most information. The spectral characteristics of each signal can be adjusted separately, instead of using a single factor for all of the signals. The entire spectral shape can be examined for such peculiarities as resonant modes. The disadvantages are twofold. First, this approach is the most complex. A Fourier transform routine must be included in the analysis program if the adjustment is to be automatic; appropriate plotting routines will also be desired. It is always good practice to examine at least a few sample residual power spectral density plots, but it is simpler to create the power spectral density plots in a separate program. The second disadvantage is that the value to use for the spectral density or the break frequency is not usually obvious from the plot. Figure 14 illustrates this problem. The spectrum does not exhibit an obvious flat area followed by a well-defined break. Picking a specific value from the plot can be as much of an art as picking a value for the spectral adjustment factor from experience.

The third approach is a compromise between the first two. It obtains information from the actual residuals, but keeps the programming relatively simple. This approach uses the total power of the low-pass filtered residuals. A simple single-pole filter is used with a break frequency two or three times the system natural frequency. The power of the filtered residuals is then divided by the filter break frequency to give the average power spectral density at frequencies near and below the system natural frequencies. This method does not provide the complete spectral information of a power spectral density plot, but it does provide a reasonable estimate of the multiplication factor with very little work. The bulk of the implementation is filtering the residuals, which is done as the residuals are computed.

This method is somewhat approximate and requires picking a value for the filter break frequency. The obvious approximations involved include the assumption that the noise spectrum is flat at least to the filter break frequency, the choice of filter break frequency, and the use of a first-order filter. In practice, noise tends to be closely concentrated around the system natural frequencies. This violates the requirement for the noise bandwidth to be well above the system frequencies. Because of these approximations, the method should not be considered to give a precise measure of the expected scatter; the only claim is that the method is an improvement on previous computations. Furthermore, some of the proposals in the previous section, although incapable of explaining discrepancies of a factor of 10, might contribute measurably to smaller discrepancies. Empirical observation suggests that a multiplication factor of about 2 typically remains necessary, because of the approximations in the proposed approach.

The method might therefore seem to be little improvement over the first approach, since the need for an empirically determined factor remains (although the factor is now smaller). The advantage of this approach is that large changes in the spectral characteristics are noticed and approximately accounted for. If the spectral characteristics of all of the maneuvers are similar, this approach and the first one are equally valid.

Figure 18 shows the flight data from figure 10 with Cramér-Rao bounds computed from the filtered residuals. A break frequency of 0.5 hertz was used for the filters.

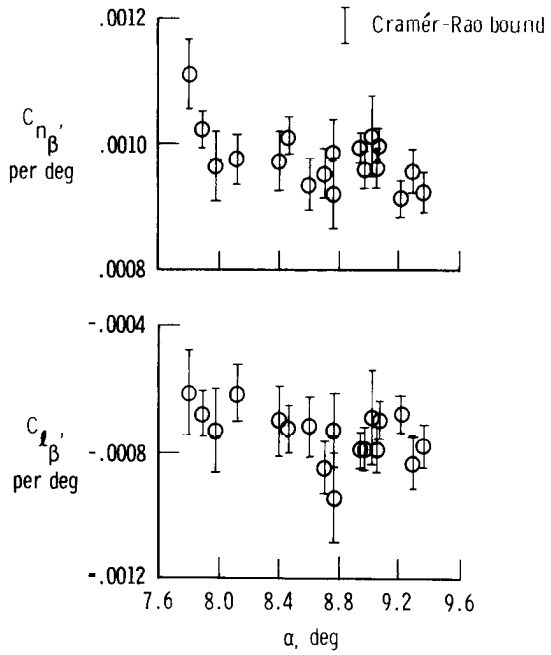


Figure 18. Estimates from flight data corrected using filtered residuals.

The correction factors computed from the filtered residual power were from 3 to 6, and a remaining empirical factor of 2 was used. The magnitudes of the Cramér-Rao bounds on this plot are reasonable and give a good visual indication of estimate accuracy. Although this plot shows no outstandingly good or poor maneuvers, there is a noticeable tendency for the estimates near the center of the scatter band to have smaller Cramér-Rao bounds than the outliers. It is concluded that this approach results in a useful estimate of the accuracy.

EXAMPLES OF APPLICATION

This section describes the application of the Cramér-Rao bounds to several sets of actual flight data. The examples illustrate the kinds of information that can be deduced with the aid of the bounds.

Example 1

The first example uses the data from figures 1 and 2, which are for a PA-30 airplane (fig. 19). Figure 20 duplicates figure 1 except that the Cramér-Rao bounds

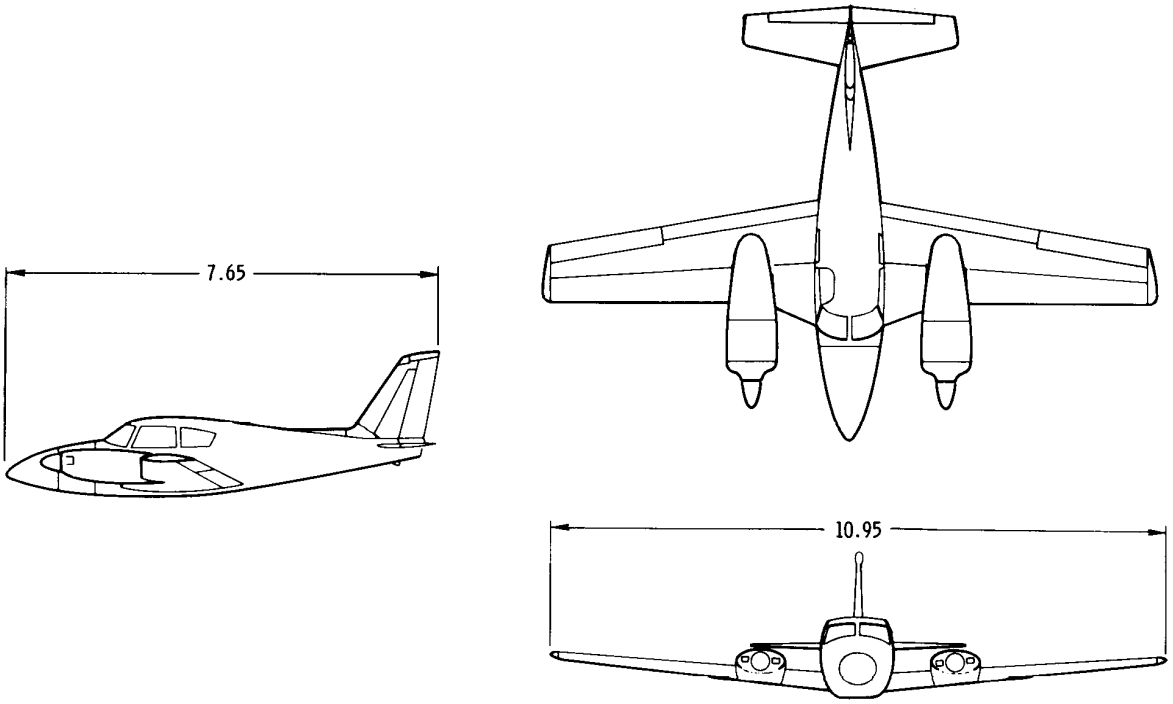


Figure 19. Three-view drawing of PA-30 airplane. Dimensions are in meters.

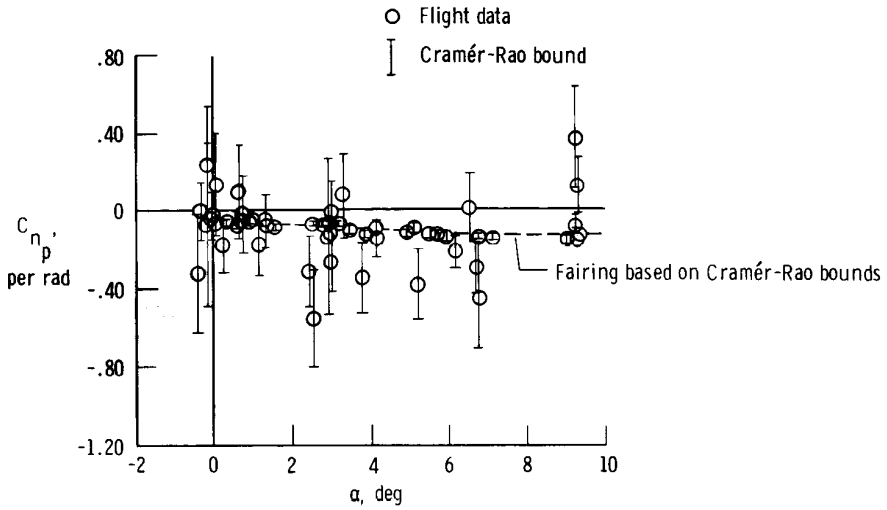


Figure 20. Estimates of C_{n_p} with Cramér-Rao bounds.

have been included. The Cramér-Rao bounds in this figure are multiplied by an empirical factor of 10 to correct for the effects of colored noise. The Cramér-Rao bounds clearly indicate which data represent the best estimates. A fairing of the data can confidently be made on the basis of the Cramér-Rao bounds, even though the total data scatter is large. Figure 21 shows the same data segregated into rudder and aileron maneuvers. Both the scatter and the Cramér-Rao bounds show that the

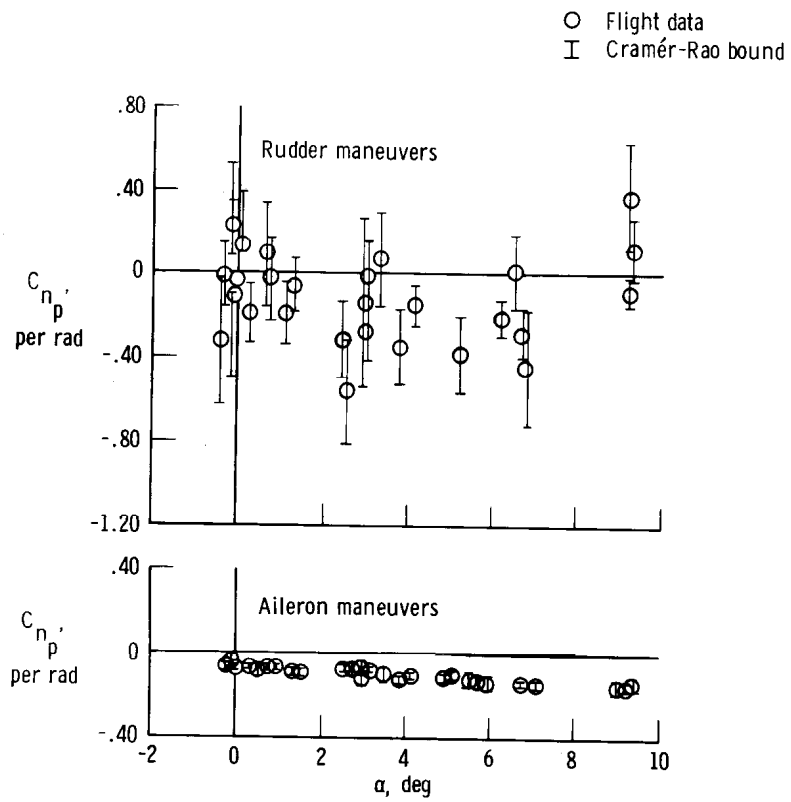


Figure 21. Estimates of C_{n_p} with Cramér-Rao bounds, segregated by input used.

aileron maneuvers give superior estimates of C_{n_p}. The rudder maneuvers do not excite the airplane well enough for the estimation of C_{n_p}. Some results can be obtained from separate rudder and aileron maneuvers, but it is apparent that in order to obtain the best estimates of all of the stability and control derivatives, all the maneuvers should include both rudder and aileron inputs. Similar conclusions have been reached in other studies (ref. 19).

An aileron and a rudder maneuver at the same flight condition can also be analyzed together to give a single set of estimates that is based on the data in both maneuvers. Figure 22 shows the estimates of C_{n_p} obtained by this pairing of the data in figure 20. All of the resulting estimates have small Cramér-Rao bounds and small scatter.

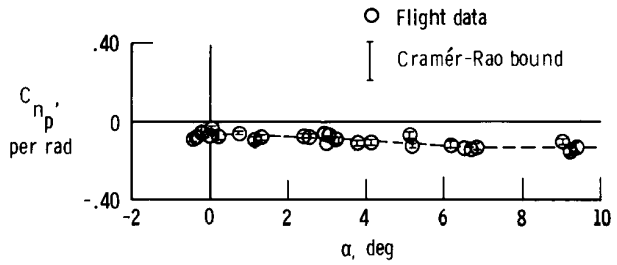


Figure 22. Estimates of C_{n_p} with Cramér-Rao bounds from multiple maneuver analysis.

Example 2

In the second example, data from the remotely piloted oblique wing vehicle (fig. 23) (ref. 30) are used. Cross coupling derivatives between the longitudinal and lateral-directional modes were of interest for this nonsymmetric vehicle. Figure 24 shows

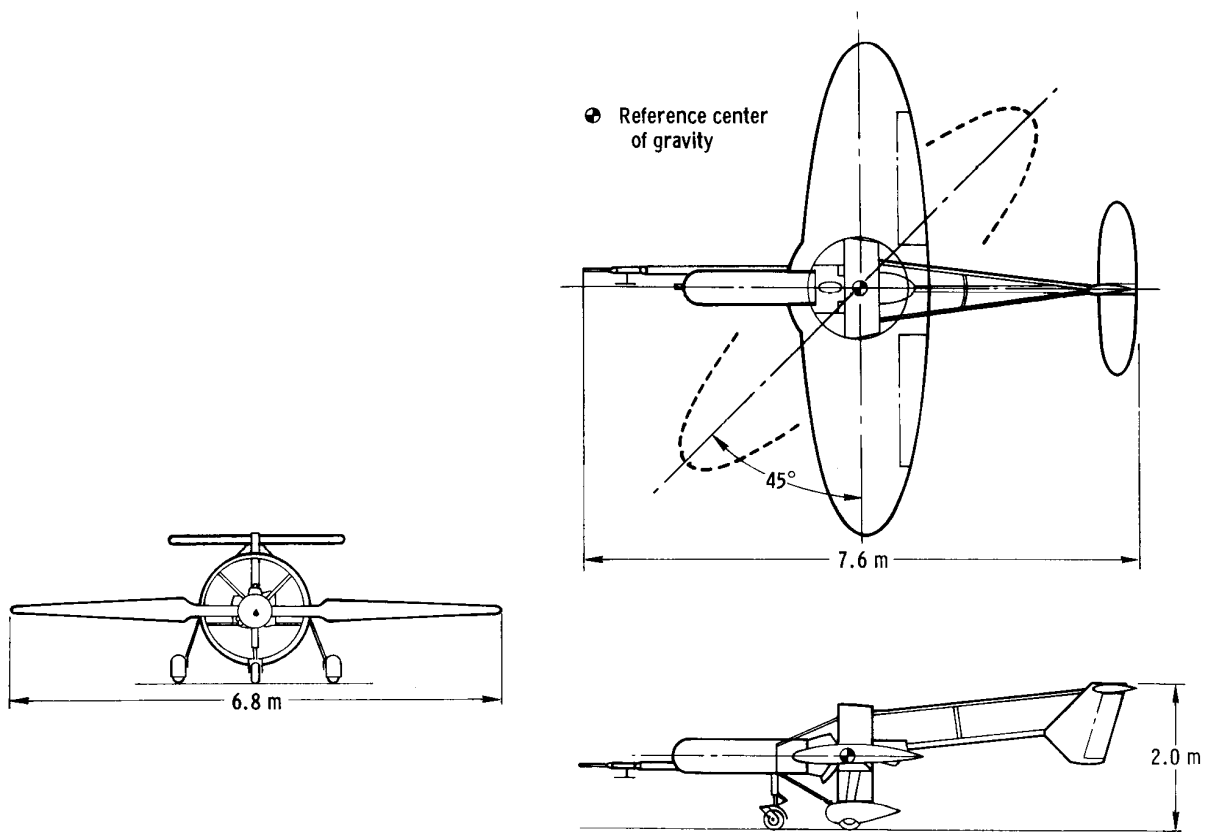


Figure 23. Three-view drawing of oblique wing aircraft.

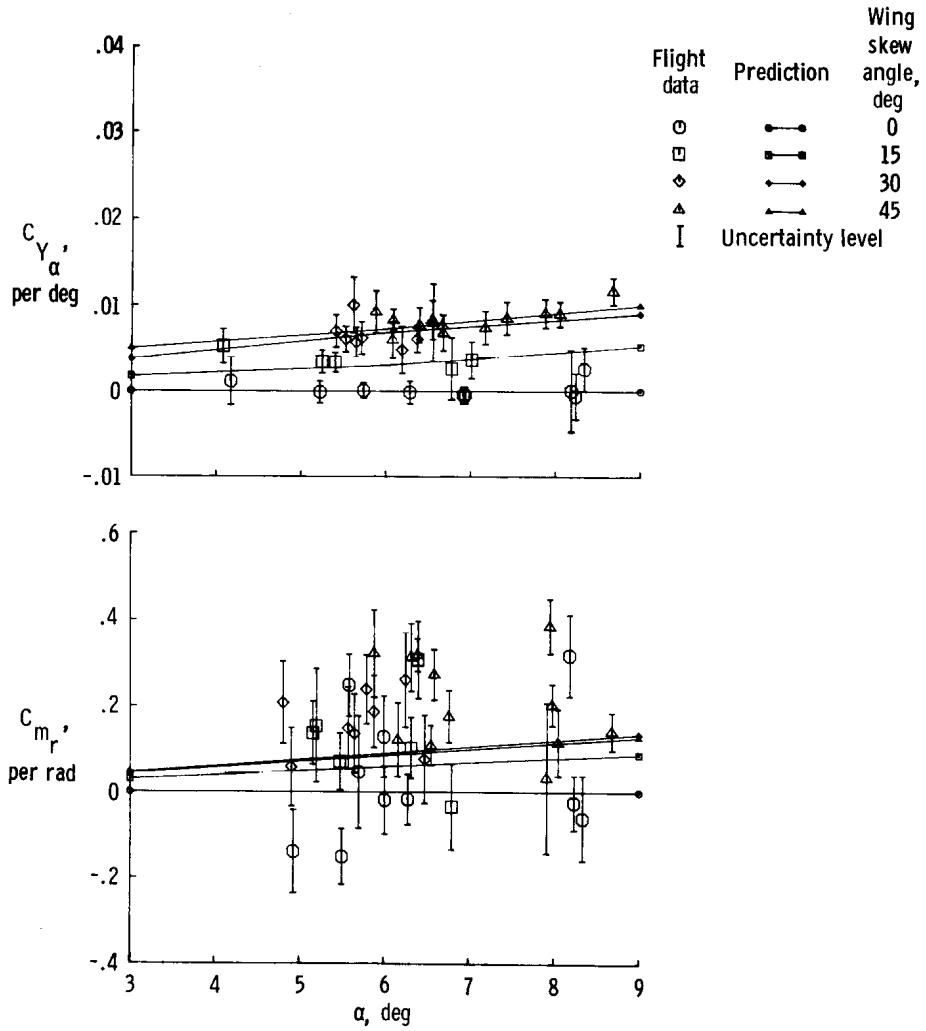


Figure 24. Estimates of C_{Y_α} and C_{m_r} for oblique wing aircraft.

the estimates of two such derivatives, C_{Y_α} and C_{m_r}. An empirically determined spectral adjustment factor of 5 was used in the Cramér-Rao bounds shown. The C_{Y_α} data show relatively small scatter and correspondingly good Cramér-Rao bounds. The effect of this derivative is easy to identify in the flight data, and the derivative estimates agree well with predictions.

For C_{m_r}, on the other hand, the scatter and the Cramér-Rao bounds are of about the same magnitude as the estimates. There is very little information about the value of C_{m_r} in the flight data. With the wing skewed 15° to 45°, the data show that C_{m_r} is positive and probably somewhere in the range from 0.1 to 0.3. It cannot be determined more precisely than this from the data available.

This example illustrates the use of the scatter and the Cramér-Rao bounds in combination to determine which derivatives can be accurately estimated from the data. Confidence in the determination is improved if both scatter and the Cramér-Rao bounds are used.

Example 3

The data for the third example were acquired from a remotely piloted 3/8-scale F-15 airplane (fig. 25, ref. 33). A spectral adjustment factor of 5 was used for these

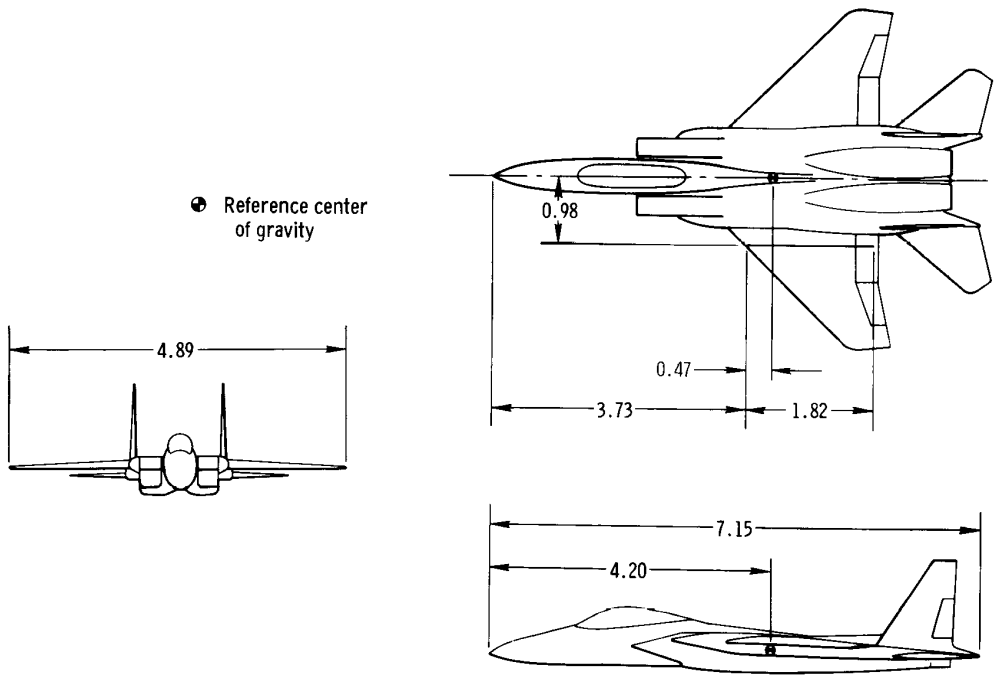


Figure 25. Three-view drawing of 3/8-scale F-15 remotely piloted research vehicle. Dimensions are in meters.

data. The estimates of C_{m_q} are shown in figure 26. The behavior of the estimates around an angle of attack of 25° is of particular interest. It is difficult to justify the fairing shown on the basis of data scatter alone. The fairing was based primarily on the data with small Cramér-Rao bounds. In particular, the data near 25° angle of attack that show values of C_{m_q} near zero have small Cramér-Rao bounds, while the data at the same angle of attack that show C_{m_q} values of -3 and -4 have larger bounds. Therefore, the fairing goes to zero at 25° angle of attack. The estimates of C_{m_α} in

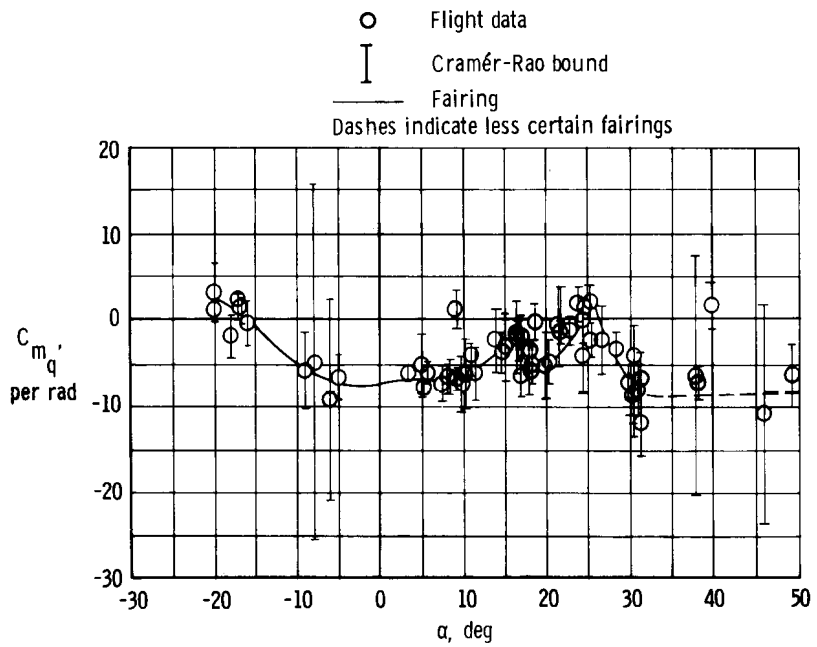


Figure 26. Estimates of C_{m_q} for 3/8-scale model of F-15 airplane.

figure 27 unambiguously show a similarly shaped trend near 25° angle of attack. Subsequent results have tended to verify the validity of the C_{m_q} fairing.

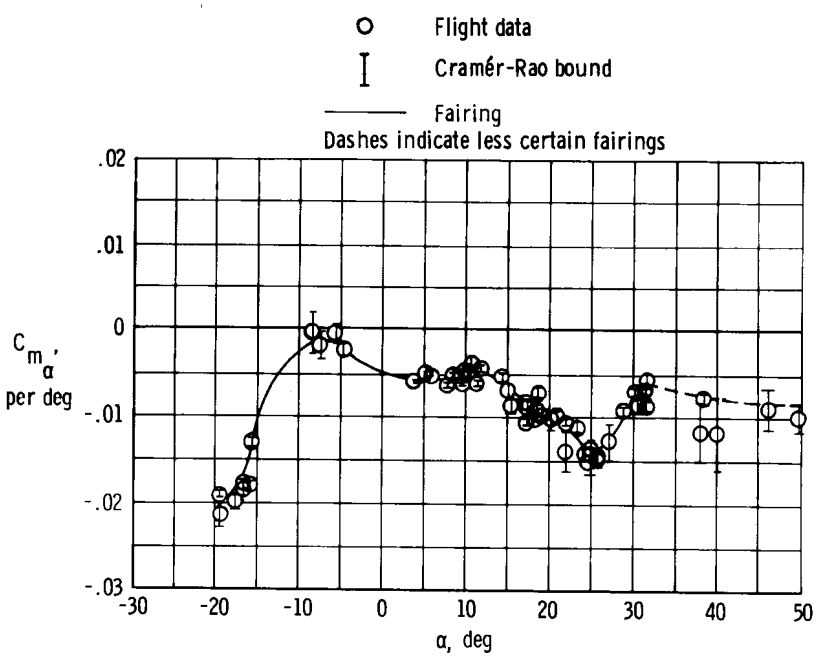


Figure 27. Estimates of C_{m_α} for 3/8-scale model of F-15 airplane.

In this example, the Cramér-Rao bounds provided more information than the scatter alone. The bounds indicated that the accuracy of some data was good, even though the overall scatter was large because of a few poor estimates. The Cramér-Rao bounds also indicate that some of the estimates above 35° angle of attack plus some of those near -10° are quite unreliable, although the scatter is not large. See example 6 for more on this point.

Example 4

Figure 28 shows estimates of C_{n_β} taken from the same 3/8-scale F-15 data as shown in example 3, using the same spectral adjustment factor of 5. The scatter above 35° angle of attack is quite large, as are most of the Cramér-Rao bounds. A few data

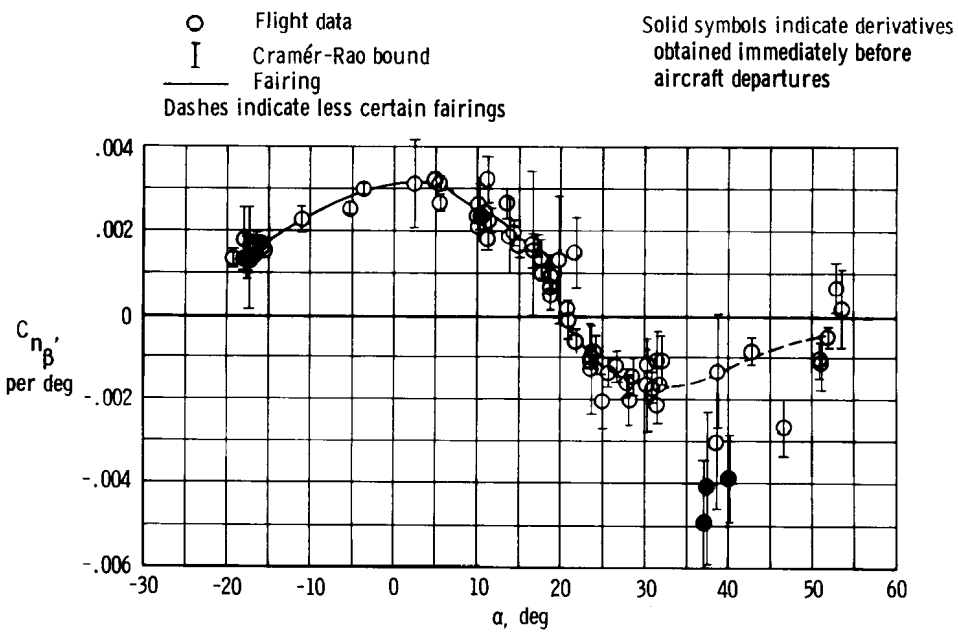


Figure 28. Estimates of C_{n_β} for 3/8-scale model of F-15 airplane.

points with small Cramér-Rao bounds form the basis for the fairing used. It is interesting that the three data points farthest from the fairing were obtained immediately before aircraft departures. These more negative values of C_{n_β} may be an early manifestation of the aerodynamic phenomenon that causes the departures.

Example 5

The fifth example uses data from an F-111A airplane (fig. 29, ref. 27). Figure 30 shows estimates of C_{m_α} for the F-111A airplane at 26° of wing sweep. The Cramér-

Rao bounds were multiplied by an empirically determined factor of 10 to correct for the effects of colored noise. The factor of 10 was chosen to make the Cramér-Rao bounds about the same magnitude as the scatter for most of the derivatives. However, the scatter in C_{m_α} is much larger than the Cramér-Rao bounds, even after the use of the factor. The data for Mach 0.9 are particularly puzzling. Since the same factor resulted in good agreement between the Cramér-Rao bounds and the scatter for the other F-111A derivatives, the C_{m_α} estimates deserve careful study.

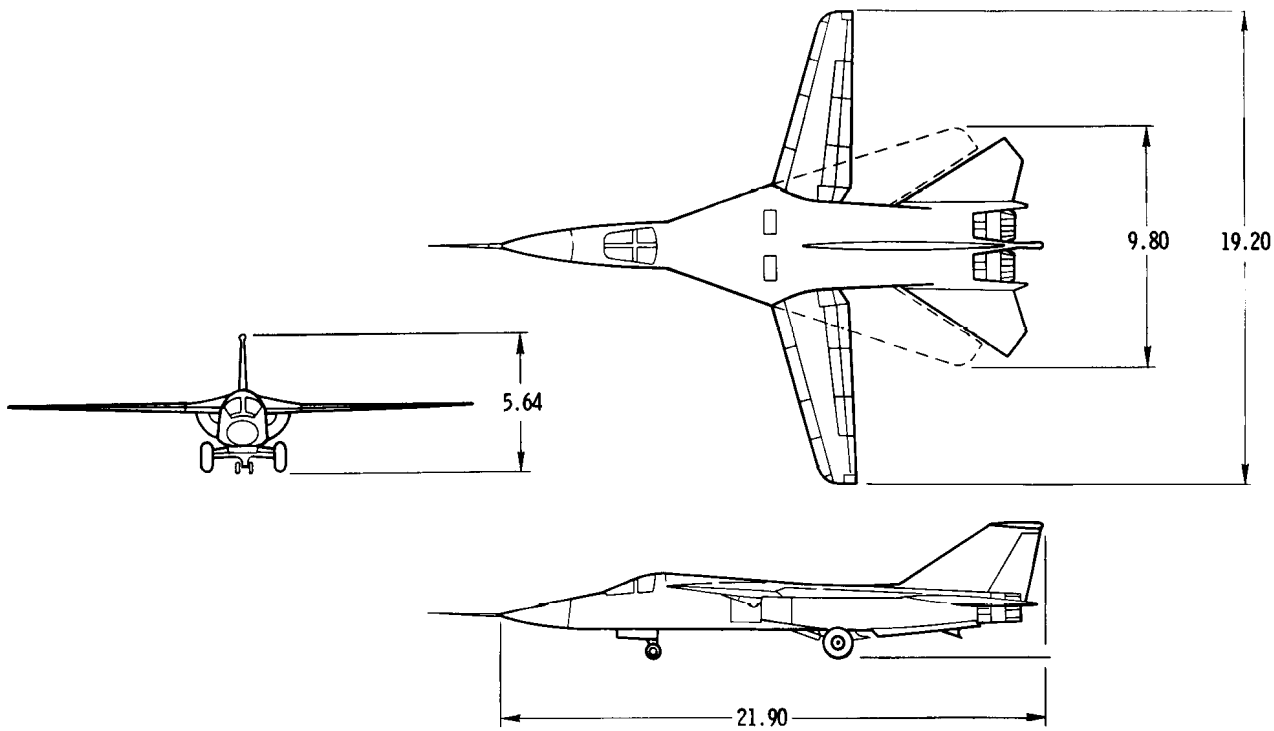


Figure 29. Three-view drawing of F-111A airplane. Dimensions are in meters.

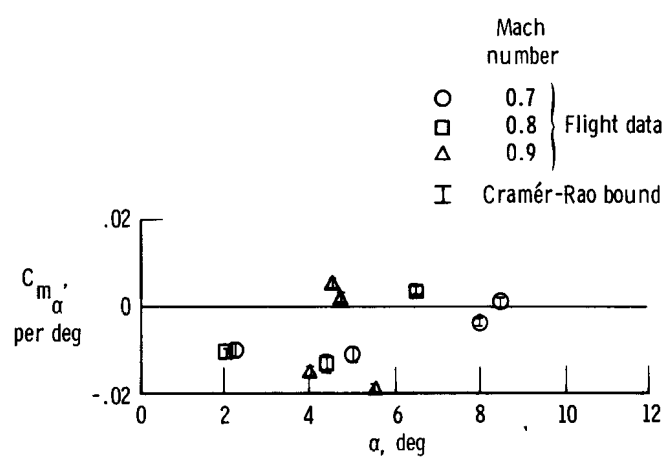


Figure 30. Estimates of C_{m_α} for F-111A airplane at 26° of wing sweep.

In order to study these results, the flight data were replotted as a function of Mach number (fig. 31). The reason for the apparent discrepancies was then evident. As shown by the fairing in figure 31, the value of C_{m_α} is significantly less stable at Mach numbers near 0.86. In the previous figure, Mach number distinctions were made only to the nearest tenth (0.7, 0.8, and 0.9). The data near Mach 0.86 were all automatically rounded to Mach 0.9. This lack of distinction between the C_{m_α} values for Mach 0.86 and 0.9 made the Mach number effect look like scatter. Figure 32 is a duplicate of figure 30 with the points near Mach 0.86 segregated and fairings of the flight data added. The scatter about these fairings is reasonable.

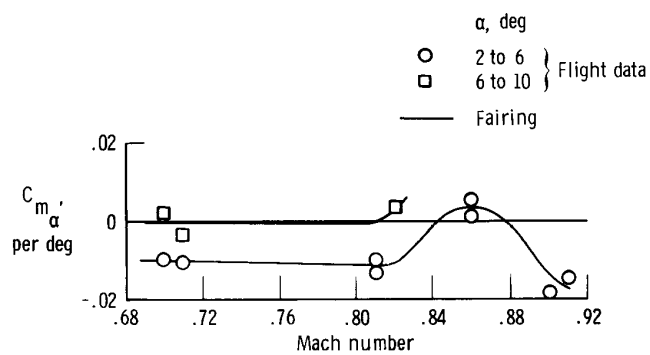


Figure 31. Estimates of C_{m_α} versus Mach number for F-111A airplane at 26° of wing sweep.

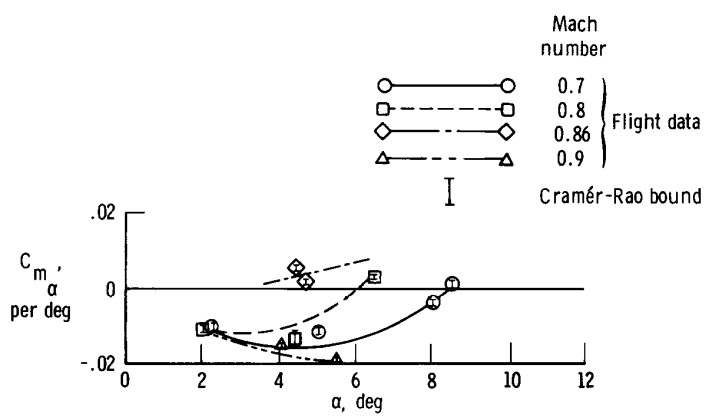


Figure 32. Estimates of C_{m_α} for F-111A airplane at 26° of wing sweep.

In this example, the Cramér-Rao bound did not directly indicate the nature of the problem, but it did draw attention to the fact that some of the data needed more careful analysis. The ensuing study disclosed the unanticipated Mach number effect.

Example 6

Figure 33 shows estimates of C_{ℓ_r} for the F-111A airplane at 35° of wing sweep. A spectral adjustment factor of 10 was used as in example 5. The data in this figure were acquired to look for derivative changes as a function of normal acceleration, which could be attributed to structural deformation. The solid line is a fairing of previous 1g flight data.

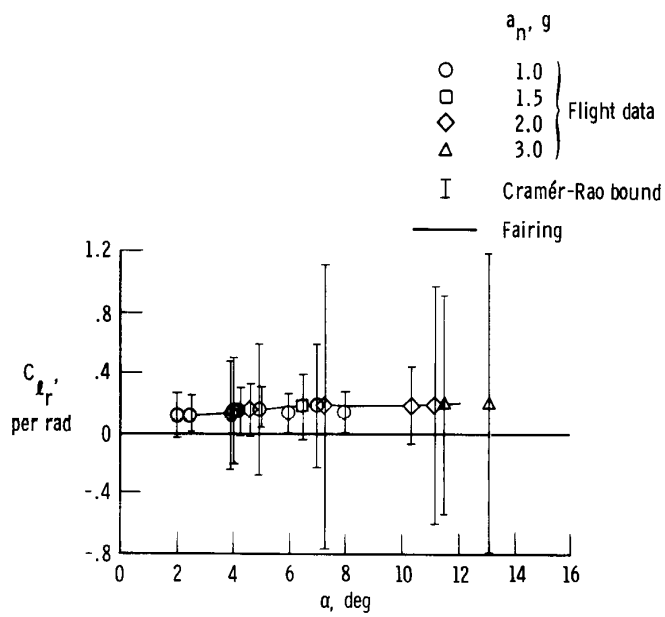


Figure 33. Estimates of C_{ℓ_r} for F-111A airplane at 35° of wing sweep.

This figure has characteristics opposite of those in the previous example. The Cramér-Rao bounds are much larger than the data scatter. In addition, the estimates are all close to the fairing of the previous data. This behavior is typical of a problem that can occur when *a priori* weighting (refs. 8 and 24) is used in the estimation.

Convergence difficulties were encountered during the analysis of several of the elevated g maneuvers, many of which were not well stabilized in flight condition. Therefore, *a priori* weighting was used to improve convergence. The *a priori* values are indicated by the fairing in figure 33.

A priori weighting is a quadratic penalty function that is added to the cost functional and penalizes the estimator for departure from the *a priori* (starting) values. The effect of this penalty is to hold estimates near the *a priori* values unless there is significant evidence that the *a priori* values are incorrect; the level of significance required can be adjusted. *A priori* weighting thus tends to eliminate large deviations in the estimates that are based on minimal information. This moderation can improve convergence if the information content of a manuever is poor.

The Cramér-Rao bounds in figure 33 indicate that the maneuvers contained little information about the derivative C_{ℓ_r}. The smallness of the scatter is probably due almost entirely to the *a priori* weighting rather than to information from the maneuvers. Thus, the data in figure 33 do not verify the previous flight data, a conclusion which might be mistakenly drawn if the Cramér-Rao bounds were not considered. Instead, figure 33 indicates that the new data do not have sufficient information either to contradict or to verify the *a priori* values.

This example also illustrates an important fact about the computation of the Cramér-Rao bound when an *a priori* weighting is used. The Cramér-Rao bounds are computed from H, an approximation to the second gradient of the cost function. When an *a priori* weighting is used, a penalty function is added to the cost funciton. The H-matrix used for the Cramér-Rao bound should reflect only the original cost function; it should not have a term added for the second gradient of the penalty function. The Cramér-Rao bound will thus reflect the amount of information in the maneuver. If the second gradient of the penalty function were added into the computation of the bound, as is sometimes done, the bound would reflect the sum of the information from the maneuver and from the *a priori* weighting. There would then be no way of determining how much of the information was obtained from the maneuver, and the phenomenon in figure 33 would not be observed (the Cramér-Rao bounds in the figure would be much smaller).

Example 7

Figure 34 shows the estimates of $C_{n_{\delta_r}}$ obtained from the same maneuvers as those used in example 6. As in example 6, the data scatter is small and agrees well with the fairing shown of the previous flight data. *A priori* weighting was used for both examples. In this example, however, the Cramér-Rao bounds are small and consistent with the scatter. It can be concluded that the maneuvers contain significant information about the derivative $C_{n_{\delta_r}}$. Therefore, the data shown give positive verification of the fairing. This contrasts with example 6, where it could only be concluded that the new data did not contradict the fairing.

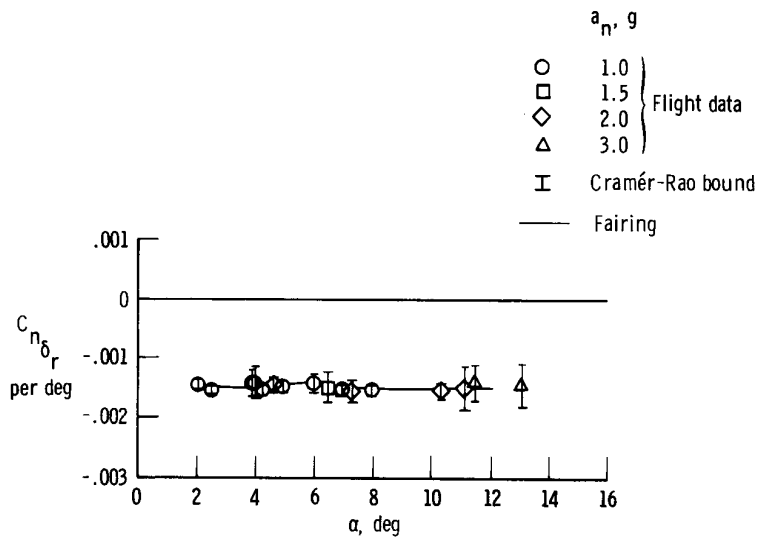


Figure 34. Estimates of $C_{n_{\delta_r}}$ for F-111A airplane at 35° of wing sweep.

CONCLUSIONS

The Cramér-Rao bound is the best of the theoretical measures of accuracy: it accounts completely for all of the sensitivities and correlations in any number of dimensions. Therefore, the sensitivities and estimated correlations cannot provide additional information about the magnitude of the accuracy of an estimated coefficient. The sensitivities and estimated correlations can be used to help determine the cause of poor accuracy, but not to test for the presence of poor accuracy. The uncertainty ellipsoid provides a statistical or a geometric picture of the relationship between the Cramér-Rao bounds, sensitivities, and correlations.

The Cramér-Rao bound is significantly affected by the colored measurement noise and modeling error present in actual flight data. Computations of the bound that ignore these effects are often in error by as much as a factor of 10. Approximate corrections for these effects are easily computed, however, and result in reasonable values for the Cramér-Rao bounds. Cramér-Rao bounds corrected in this manner have proved to be extremely useful in application to real flight data.

Modeling error must be considered to be a component of noise. Theoretical studies based only on instrumentation noise characteristics produce results that are overoptimistic.

It is usually impractical to use the bias as a quantitative theoretical measure of estimate accuracy. The possibility of biases should be considered by the engineer during any evaluation of estimates.

The scatter of the estimates is a useful empirical measurement that should always be examined if sufficient data are available for its definition. A comparison of the scatter and the Cramér-Rao bounds provides a good check of empirical and theoretical measures.

Engineering judgment remains the cornerstone of gaging the accuracy of the estimates. The empirical and theoretical measures should all be taken as tools to aid an engineer's judgment rather than as absolute incontrovertible values. To arrive at the final best estimate of accuracy, the engineer must use his judgment to combine the information from all of the available tools. He must evaluate the applicability of the tools and consider the factors such as modeling error that were ignored in their development.

National Aeronautics and Space Administration
Dryden Flight Research Center
Edwards, Calif., September 23, 1980

REFERENCES

1. Iliff, Kenneth W.; Maine, Richard E.; and Montgomery, T. D.: Important Factors in the Maximum Likelihood Analysis of Flight Test Maneuvers. NASA TP-1459, 1979.

2. Montgomery, R. C.; Mekel, R.; and Nachmias, S.: A Learning Flight Control System for the F8-DFBW Aircraft. Guidance and Control Conf., Palo Alto, Calif., August 7-9, 1978, Technical Papers, Am. Inst. Aeronaut. Astronaut., 1978, pp. 325-331.

3. Goodwin, Graham C.; and Payne, Robert L.: Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, 1977.

4. Balakrishnan, A. V.: Communication Theory. McGraw-Hill Book Co., c.1968, Ch. 3.

5. Cramér, Harald: Mathematical Methods of Statistics. Princeton Univ. Press, 1957.

6. Papoulis, Athanasios: Probability, Random Variables, and Stochastic Processes. McGraw-Hill Book Co., c.1965.

7. Meyer, Paul L.: Introductory Probability and Statistical Applications. Second ed., Addison-Wesley Publishing Co., Inc., 1970.

8. Maine, Richard E.; and Iliff, Kenneth W.: User's Manual for MMLE3, A General FORTRAN Program for Maximum Likelihood Parameter Estimation. NASA TP-1563, 1980.

9. Klein, Vladislav: Identification Evaluation Methods. Parameter Identification, AGARD-LS-104, 1979, pp. 2-1—2-21.

10. Hodge, Ward F.; and Bryant, Wayne H.: Monte Carlo Analysis of Inaccuracies in Estimated Aircraft Parameters Caused by Unmodeled Flight Instrumentation Errors. NASA TN D-7712, 1975.

11. Sorensen, John A.: Analysis of Instrumentation Error Effects on the Identification Accuracy of Aircraft Parameters. NASA CR-112121, 1972.

12. Gupta, Narendra K.; and Hall, W. Earl, Jr.: Design and Evaluation of Sensor Systems for State and Parameter Estimation. AIAA J. Guidance and Control, vol. 1, no. 6, Nov.-Dec. 1978, pp. 397-403.

13. Gupta, Narendra K.; Hall, W. Earl, Jr.; and Trankle, Thomas L.: Advanced Methods of Model Structure Determination From Test Data. AIAA Paper 77-1170, 1977.

14. Fiske, Philip H.; and Price, Charles F.: A New Approach to Model Structure Identification. AIAA Paper 77-1171, 1977.

15. Akaike, Hirotugu: A New Look at the Statistical Model Identification. IEEE Trans. Automat. Contr., vol. AC-19, no. 6, Dec. 1974, pp. 716-723.

16. Caines, P. E.; and Ljung, L.: Asymptotic Normality and Accuracy of Predection [sic] Error Estimators. Proceedings of the 1976 Joint Automatic Control Conference, Am. Soc. Mech. Eng., c.1976, pp. 586-592.

17. Stepner, David E.; and Mehra, Raman K.: Maximum Likelihood Identification and Optimal Input Design for Identifying Aircraft Stability and Control Derivatives. NASA CR-2200, 1973.

18. Williams, James L.; and Suit, William T.: Extraction From Flight Data of Lateral Aerodynamic Coefficients for F-8 Aircraft With Supercritical Wing. NASA TN D-7749, 1974.

19. Cannaday, Robert L.; and Suit, William T.: Effects of Control Inputs on the Estimation of Stability and Control Parameters of a Light Airplane. NASA TP-1043, 1977.

20. Wells, W. R.; and Ramachandran, S.: Flight Test Design for Efficient Extraction of Aircraft Parameters. Proceedings, AIAA 3rd Atmospheric Flight Mechanics Conf., June 1976, pp. 101-107.

21. Wilkinson, J. H.: The Algebraic Eigenvalue Problem. Oxford Univ. Press, c.1965.

22. Dixon, L. C. W.: Nonlinear Optimisation. Crane, Russack & Co., Inc., c.1972.

23. Parrish, Russell V.; and Steinmetz, George G.: Lateral Stability and Control Derivatives of a Jet Fighter Airplane Extracted From Flight Test Data by Utilizing Maximum Likelihood Estimation. NASA TN D-6905, 1972.

24. Iliff, Kenneth W.; and Maine, Richard E.: Practical Aspects of Using a Maximum Likelihood Estimation Method To Extract Stability and Control Derivatives From Flight Data. NASA TN D-8209, 1976.

25. Suit, William T.: Aerodynamic Parameters of the Navion Airplane Extracted From Flight Data. NASA TN D-6643, 1972.

26. Tanner, Russel R.; and Montgomery, Terry D.: Stability and Control Derivative Estimates Obtained From Flight Data for the Beech 99 Aircraft. NASA TM-72863, 1979.

27. Iliff, Kenneth W.; Maine, Richard E.; and Steers, Sandra Thornberry: Flight-Determined Stability and Control Coefficients of the F-111A Airplane. NASA TM-72851, 1978.

28. Powers, Bruce G.: Phugoid Characteristics of a YF-12 Airplane With Variable Geometry Inlets Obtained in Flight Tests at a Mach Number of 2.9. NASA TP-1107, 1977.

29. Sim, Alex G.: A Correlation Between Flight-Determined Derivatives and Wind-Tunnel Data for the X-24B Research Aircraft. NASA SX-3371, 1976.

30. Maine, Richard E.: Aerodynamic Derivatives for an Oblique Wing Aircraft Estimated From Flight Data by Using a Maximum Likelihood Technique. NASA TP-1336, 1978.

31. Iliff, Kenneth W.; and Maine, Richard E.: Further Observations on Maximum Likelihood Estimates of Stability and Control Characteristics Obtained From Flight Data. AIAA Paper 77-1133, 1977.

32. Balakrishnan, A. V.; and Maine, Richard E.: Improvements in Aircraft Extraction Programs. NASA CR-145090 [1976].

33. Iliff, Kenneth W.; Maine, Richard E.; and Shafer, Mary F.: Subsonic Stability and Control Derivatives for an Unpowered, Remotely Piloted 3/8-Scale F-15 Airplane Model Obtained From Flight Test. NASA TN D-8136, 1976.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA RP-1077		

4. Title and Subtitle	5. Report Date
THE THEORY AND PRACTICE OF ESTIMATING THE ACCURACY OF DYNAMIC FLIGHT-DETERMINED COEFFICIENTS	July 1981
	6. Performing Organization Code RTOP 505-43-14
7. Author(s) Richard E. Maine and Kenneth W. Iliff	8. Performing Organization Report No. H-1128
	10. Work Unit No.
9. Performing Organization Name and Address Dryden Flight Research Center P.O. Box 273 Edwards, California 93523	11. Contract or Grant No.
	13. Type of Report and Period Covered Reference Publication
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This paper discusses means of assessing the accuracy of maximum likelihood parameter estimates obtained from dynamic flight data. The most commonly used analytical predictors of accuracy are derived and compared from both statistical and simplified geometric standpoints. The accuracy predictions are evaluated with real and simulated data, with an emphasis on practical considerations, such as modeling error. Improved computations of the Cramér-Rao bound to correct large discrepancies due to colored noise and modeling error are presented. The corrected Cramér-Rao bound is shown to be the best available analytical predictor of accuracy, and several practical examples of the use of the Cramér-Rao bound are given. Engineering judgment, aided by such analytical tools, is the final arbiter of accuracy estimation.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Accuracy Maximum likelihood Parameter estimation Cramér-Rao bounds Flight test Estimation	Unclassified—Unlimited STAR category 65

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	64	A04

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1981