LARRY J. LEVY and DAVID W. PORTER

LARGE-SCALE SYSTEM PERFORMANCE PREDICTION WITH CONFIDENCE FROM LIMITED FIELD TESTING USING PARAMETER IDENTIFICATION

Limited system field testing, suitably combined with other factory, subsystem, and system tests, can be used to develop a model that will produce confident operational performance predictions. The approach is to fit a fundamental parametric state-space model to the test data, where the test and operational mission diversities are accounted for in the known model structure (developed from first principles) and the unknown parameters are generally common from test to test. New maximum likelihood parameter identification techniques have been developed that optimally and efficiently estimate the unknown parameters in large-scale system models and provide an identification uncertainty to quantify model confidence. The latter attribute allows test planning trade-offs to be conducted in terms of sizing, instrumentation requirements, and types of tests to achieve a desired numerical confidence. Application of the technology to a simplified example of missile inertial guidance test and evaluation is shown.

INTRODUCTION

High costs and implementation difficulties have tended to limit the use of field testing in large-scale system evaluation and performance prediction. System performance prediction has relied on either traditional empirical methods (e.g., sampling of target errors from many tests for accuracy performance) based on extensive repetitive field testing over the entire range of possible operational mission conditions, or validation of complex engineering models based on a limited set of realistic field tests. The former is usually too costly and yields minimal insight into system understanding. The latter gives some assurance over the limited test conditions but cannot predict system performance over untested conditions with quantified confidence. In addition, test planning and requirements are difficult to predict from system performance requirements. Both approaches neglect the rich source of detailed model information provided by each system test.

For example, inertial navigation system (INS) accuracy prediction requires statistical INS error models that are valid over the entire set of operational mission conditions. Traditionally, these error models have been developed by fitting well-known (from first principles) structural models to factory component and subsystem test data. At times the factory tests cannot sufficiently emulate all the operational mission conditions, resulting in poorly determined models. Field testing is then used to validate that the factory-derived model is "good" over the set of field test conditions (not necessarily good over untested mission conditions). When mismatches between the model and data occur, attempts are made to modify the factory model to match the field test data. Generally, however, no systematic approach is used to extract the error model parameter information from the field tests for comparison and/or combination with the factory-derived model. The traditional methodologies could not handle the nonrepeatable and complex nature of these tests, neglecting the rich source of model information contained therein. Also, it was difficult to quantify the test requirements (type, size, instrumentation) and numerically relate them to some high-level system performance requirement. Consequently, ad hoc reasoning was usually used to justify test requirements.

The methodology we describe in this article provides a computational technique for identifying the system model from all types of test data. It estimates the unknown common fundamental parameters of detailed state-space system models with known structures and supplies its own estimation uncertainty so that quantifiable confidence statements can be attached to model predictions. This latter capability also allows test planning requirements analyses to be conducted before the start of any test program by numerically relating system performance confidence statements to specific test characteristics. We will define the applicable generic state-space model and describe the overall operation of the maximum likelihood identification methodology.[1] (The nonspecialist reader may wish to skip the second, more theoretical section.) Next, we define a simplified missile inertial guidance test and evaluation example and discuss the numerical simulation results, contrasting this method with a more traditional approach.

Other examples of the use of system identification technology include (1) instrumentation and sensor modeling and validation;[2] (2) missile, aircraft, and projectile aerodynamic modeling;[3–5] (3) flexible structures modeling;[6] (4) geomagnetic field modeling;[7] (5) power plant, petrochemical, and other industrial process modeling;[8] (6) image and signal processing;[9] and (7) on-line tuning of controllers.[10]

DEFINITION OF THE IDENTIFICATION PROBLEM

The identification methodology presumes that, for each test, the model for the system and associated test instrumentation consists of a known linear (or linearizable) state-space structure with a set of generally common unknown parameters. The system portion of the structure essentially transforms (via covariance/Monte Carlo simulations) the parameters into system performance such as target accuracy or circular error probable (CEP; radius of the circle centered on the target containing 50% of the probable target errors) for a missile guidance system. We assume that the structure can be known confidently from first principles. With the unknown parameters independent of scenario (i.e., all test and operational mission characteristic dependence is in the structure), diverse tests can be combined for statistical "leverage" in the model identification process.

Mathematically, we assume that each system (or subsystem/component) under test is represented by the following discrete-time linear state-space model for each test (realization) j ($j = 1, \ldots, N$), with unknown parameters in the vector $\boldsymbol{\theta}$,

$$\mathbf{x}_{k+1}^j = \boldsymbol{\Phi}_k^j(\boldsymbol{\theta})\mathbf{x}_k^j + \mathbf{w}_k^j \qquad k = 0, 1, \ldots, n(j) - 1 \,. \tag{1}$$

The state at time k, $\mathbf{x}_k^j$, is observed indirectly by

$$\mathbf{z}_k^j = \mathbf{H}_k^j(\boldsymbol{\theta})\mathbf{x}_k^j + \mathbf{v}_k^j \qquad k = 0, 1, \ldots, n(j) \,, \tag{2}$$

and where $\mathbf{x}_0^j$, $\mathbf{w}_k^j$, and $\mathbf{v}_k^j$ are mutually independent random vectors for all k, j with dimensions $p(j)$, $p(j)$, and $m(j)$, respectively. (Dimensions for the other terms can be inferred from Eqs. 1 and 2.) The initial state $\mathbf{x}_0^j$ and the white noise sequences $\mathbf{w}_k^j$ and $\mathbf{v}_k^j$ have Gaussian distributions with respective means of $\boldsymbol{\mu}_0^j(\boldsymbol{\theta})$, 0, and 0 and respective covariances of $\boldsymbol{\Sigma}_0^j(\boldsymbol{\theta})$, $\mathbf{R}_k^j(\boldsymbol{\theta})$, and $\mathbf{Q}_k^j(\boldsymbol{\theta})$.

Given N realizations (tests) of the random process (system) represented by the model in Equations 1 and 2, our objective is to determine the maximum likelihood estimate (MLE), $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, of $\boldsymbol{\theta}$. The $\boldsymbol{\mu}_0^j(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_0^j(\boldsymbol{\theta})$, $\boldsymbol{\Phi}_k^j(\boldsymbol{\theta})$, $\mathbf{H}_k^j(\boldsymbol{\theta})$, $\mathbf{Q}_k^j(\boldsymbol{\theta})$, and $\mathbf{R}_k^j(\boldsymbol{\theta})$ have known structures that embody the scenario characteristics. The only unknowns are the elements of $\boldsymbol{\theta}$, the q-dimensional parameter vector with elements that are independent of scenario and generally common across most realizations. The elements of $\boldsymbol{\theta}$ need not be in each test, and the dimensions of $\mathbf{x}_0^j$, $\mathbf{w}_k^j$, and $\mathbf{v}_k^j$ can differ from test to test. Thus, we can combine system, subsystem, and component tests that are measured by different instrumentation (different state-space models with possibly some different elements of $\boldsymbol{\theta}$ in each test type). This methodology can apply to nonlinear state-space models as long as they can be linearized about a nominal solution to obtain the form of Equations 1 and 2. Essentially, a number of realizations will be needed to estimate elements in $\boldsymbol{\mu}_0^j(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_0^j(\boldsymbol{\theta})$, but only one realization may be sufficient for some elements in $\boldsymbol{\Phi}_k^j(\boldsymbol{\theta})$, $\mathbf{H}_k^j(\boldsymbol{\theta})$, $\mathbf{Q}_k^j(\boldsymbol{\theta})$, and $\mathbf{R}_k^j(\boldsymbol{\theta})$. The state-space models in Equations 1 and 2 are quite broad, including linear regression and autoregressive moving average models, each with their own specialized identification techniques. We will concentrate on the more difficult general problem, however.

For example, in INS testing, the inertial model must be linearized about a nominal trajectory of the system to fit the form of Equations 1 and 2. Typical INS elements in $\boldsymbol{\theta}$ will be the 1σ values in $\boldsymbol{\Sigma}_0^j(\boldsymbol{\theta})$ of the accelerometer, gyro, and misalignment biases; time constants in $\boldsymbol{\Phi}_k^j(\boldsymbol{\theta})$ for the Markov processes; spectral density values in $\mathbf{Q}_k^j(\boldsymbol{\theta})$ for the Markov and random-walk processes; and residual mean errors (in the factory calibrations) in $\boldsymbol{\mu}_0^j(\boldsymbol{\theta})$. (The simplified problem at the end of this article illustrates these types of random processes.) Occasionally, instrumentation measurement noise covariances may be unknown in $\mathbf{R}_k^j(\boldsymbol{\theta})$. For detailed INS models one could easily expect the number of unknown elements to be in the thirties or more (some possibly in the hundreds). In other applications (e.g., evaluation of aircraft aerodynamic characteristics), the model must be linearized about an estimated aircraft trajectory. Here the unknown elements are stability and control coefficients, along with parameters characterizing atmospheric turbulence.

The method of maximum likelihood essentially uses the known structure of the model defined in Equations 1 and 2 to calculate the analytical form of how the unknown parameters, $\boldsymbol{\theta}$, influence the probability density of observing a set of measurement data from an ensemble of N tests. When a specific set of measurement data is obtained from the tests, the best estimate, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, is the value of $\boldsymbol{\theta}$ that maximizes the probability density of obtaining those data over some allowable set of values for $\boldsymbol{\theta}$. Thus, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is the value that makes the actual measurements "most likely" (or the value giving the greatest chance of getting the data you actually obtain). As more and more data are gathered, the estimate will approach the true value, with an estimation error that is approximately Gaussian-distributed with a computable covariance (uncertainty) that is the inverse of the so-called Fisher information matrix.[11] Thus, for a missile INS, the estimate $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ can be transformed (via the identified Monte Carlo or covariance simulation) to target impact to obtain an accuracy estimate, $\widehat{\mathrm{CEP}}$, whereas the inverse Fisher information matrix, $\mathcal{F}_{\theta\theta^t}^{-1}$, can be transformed to target impact to calculate confidence intervals about $\widehat{\mathrm{CEP}}$. The maximum likelihood method was chosen over other criteria for estimating $\boldsymbol{\theta}$ for the following reasons:

1. The most likely model (most likely value of $\boldsymbol{\theta}$ to have caused the observed test measurements) is an intuitively pleasing criterion.
2. Generally, it is also asymptotically (as the number of test samples gets large enough) an unbiased and minimum mean-square-error estimate.
3. It provides a computationally practical estimate, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, an asymptotic estimation covariance matrix, $\mathcal{F}_{\theta\theta^t}^{-1}$, and an asymptotic normal distribution to calculate confidence intervals.
4. The estimation uncertainty $\mathcal{F}_{\theta\theta^t}^{-1}$ depends on the quality (observability) and quantity (sample size) of the test data. If needed, *a priori* information based on engineering

judgment can be blended in to reduce the estimation uncertainty.

MAXIMUM LIKELIHOOD IDENTIFICATION ALGORITHMS

The MLE is obtained by maximizing the conditional probability density function of the data given $\boldsymbol{\theta}$ or equivalently the log likelihood function,

$$\log L(\boldsymbol{\theta}) \triangleq \log p(\mathbf{Z}^1, \ldots, \mathbf{Z}^N | \boldsymbol{\theta}) \ ,$$

where $\mathbf{Z}^j \triangleq (\mathbf{z}_0^{jt}, \ldots, \mathbf{z}_{n(j)}^{jt})^t$ are the data from each realization j, the superscript t denotes the matrix transpose, and $p(\cdot|\boldsymbol{\theta})$ is the conditional probability density function of the data given $\boldsymbol{\theta}$ ($\triangleq$ means defined as). The traditional MLE approach[11,12] to the foregoing state-space problem reformulates the log likelihood function in terms of the Kalman filter innovations; the resulting Scoring (Gauss–Newton) algorithm requires a Kalman filter plus q "differentiated Kalman filters" for each realization per iteration to calculate the log likelihood gradient and Fisher information matrix. This exploding (as q becomes large in large-scale systems) computational burden plus the known slow and erratic numerical convergence character of Scoring under poor starting values have limited its practical application.[13] Initial work at APL concentrated only on the simpler problem of estimating the elements of $\boldsymbol{\mu}_0^j(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_0^j(\boldsymbol{\theta})$, which did not require q-differentiated Kalman filters.[14–16]

These problems were alleviated[17] by reformulating the MLE around the log likelihood of the "complete" data,

$$\log L^c(\boldsymbol{\theta}) \triangleq \log p(\mathbf{Z}^1, \mathbf{X}^1, \ldots, \mathbf{Z}^N, \mathbf{X}^N | \boldsymbol{\theta}) \, ,$$

where $\mathbf{X}^j \triangleq (\mathbf{x}_0^{jt}, \ldots, \mathbf{x}_{n(j)}^{jt})^t$ and $\mathbf{Z}^j$, $\mathbf{X}^j$ are termed the complete data from the realization j. This resulted in an expectation-maximization (EM) algorithm, where the $i+1$ iteration is given by the expectation step

$$G(\boldsymbol{\theta}; \boldsymbol{\theta}^i) \triangleq E[\log L^c(\boldsymbol{\theta}) | \mathbf{Z}^1, \ldots, \mathbf{Z}^N, \boldsymbol{\theta}^i] \tag{3}$$

and the maximization step

$$\boldsymbol{\theta}^{i+1} = \max_{\boldsymbol{\theta}} G(\boldsymbol{\theta}; \boldsymbol{\theta}^i) \, , \tag{4}$$

where $E[\ ]$ is the expectation operator. The EM idea is to maximize the complete data log likelihood, $\log L^c(\boldsymbol{\theta})$, which results in relatively simple solutions for the maximum. Since the complete data are not available, however, the next best thing is to maximize the average value, given the measurements. It is proved in Ref. 13 that whenever a value of $\boldsymbol{\theta}$ satisfies $G(\boldsymbol{\theta}; \boldsymbol{\theta}^i) > G(\boldsymbol{\theta}^i; \boldsymbol{\theta}^i)$, then $\log L(\boldsymbol{\theta}) > \log L(\boldsymbol{\theta}^i)$ and the algorithm will converge to a stationary point of the $\log L(\boldsymbol{\theta})$ under mild regularity conditions.[18] The expectation step required only a single Kalman filter and fixed interval smoother for each realization per iteration regardless of the size of q. The maximization step is achieved by setting $\partial G(\boldsymbol{\theta}; \boldsymbol{\theta}^i) | \partial \boldsymbol{\theta} = 0$ and solving for $\boldsymbol{\theta}$. (This does not always guarantee a maximum, so that the condition $G(\boldsymbol{\theta}^{i+1}; \boldsymbol{\theta}^i) > G(\boldsymbol{\theta}^i; \boldsymbol{\theta}^i)$ should be checked after each iteration.) When $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, $\mathbf{Q}$, $\mathbf{R}$, $\boldsymbol{\Phi}$, and $\mathbf{H}$ are all unknown and constant (across k and j), simple analytical solutions result for $\boldsymbol{\theta}^{i+1}$. The more realistic partially known model cases were solved[1] by using the Kroneker product and matrix calculus to calculate the gradient of $G(\boldsymbol{\theta}; \boldsymbol{\theta}^i)$. Quasi-analytical solutions for the zeroes of the gradients resulted in general expressions for the maximization step. The amount of computation for this EM algorithm is about equal to three "Kalman filters" per realization per iteration, independent of the value of q as compared with the previous traditional approach requiring $q+1$ Kalman filters per realization per iteration. The EM algorithm has good starting characteristics but slows down as the iterate nears a stationary point of the $\log L(\boldsymbol{\theta})$.[17] Also, no Fisher information calculation is required.

The well-known Scoring algorithm is another numerical technique for finding the maximum of $\log L(\boldsymbol{\theta})$ and is given by

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i + \mathscr{F}_{\boldsymbol{\theta}\boldsymbol{\theta}^t}(\boldsymbol{\theta}^i)^{-1} \left. \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^i} , \tag{5}$$

and the Fisher information matrix is given by

$$\mathscr{F}_{\boldsymbol{\theta}\boldsymbol{\theta}^t}(\boldsymbol{\theta}^i) = E\left\{ \left[\left. \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^i} \right] \left[\left. \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^i} \right]^t \middle| \boldsymbol{\theta}^i \right\} . \tag{6}$$

Again, the traditional approach to computing the gradient and the Fisher information matrix resulted in the exploding dimensionality problem of q.[11,12] It was shown in Ref. 19 that Scoring is related to EM by Fisher's Identity,

$$\left. \frac{\partial G(\boldsymbol{\theta}; \boldsymbol{\theta}^i)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^i} = \left. \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^i} . \tag{7}$$

Thus, the same Kalman filters and fixed interval smoothers used in EM in Ref. 17 could also be used in Scoring,[19] obviating the exploding dimensionality problem of q. The practical algorithm to accomplish this was developed by the authors,[1] again by using the Kroneker product and matrix calculus to determine general expressions for the gradient and Fisher information matrix of the log likelihood. As in the EM algorithm, the computation for the gradient is equivalent to about three Kalman filters per realization per iteration. The Fisher information matrix, however, requires considerable additional computation (i.e., smoother error cross covariances between all pairs of measurement times). We developed a new, simpler method to compute these cross covariances.[1] Experience has shown that only the cross covariances between "closely spaced" measurement times are needed for an adequate Fisher approximation. So far the extra computational burden for an approximate Fisher has been equivalent to about three to five Kalman filters per realization per iteration.

The complementary characteristics of the EM and Scoring algorithms (EM having good starting convergence, poor finish, and no Fisher; Scoring having poor starting convergence, good finish, with Fisher calculation) plus their common Kalman filter/fixed interval smoother computations provided a unified algorithm that starts out with EM and finishes with Scoring, exploiting the desirable convergence characteristics of each procedure. Consequently, computational reliability and efficiency are dramatically improved. A diagram of the overall procedure is shown in Figure 1. When the algorithm is in the EM mode, the bank of Kalman filters/smoothers provides the estimation step, whereas the maximization step is accomplished in the model identifier. The model uncertainty, $\mathscr{F}_{\theta\theta^t}^{-1}$, is calculated only when the algorithm is in the Scoring mode. After the identification algorithm has converged, the identified model and its uncertainty may be transformed into some system performance measure estimate with an associated confidence interval.

The asymptotic distribution of the estimate $\hat{\theta}_{ML}$ is Gaussian, with a mean of θ_0 and covariance of $\mathscr{F}_{\theta\theta^t}(\theta_0)^{-1}$, where θ_0 is the true value of θ. Actually, for finite data, $\mathscr{F}_{\theta\theta^t}^{-1}$ is an approximation to the estimation error covariance. Consequently, care must be exercised when using this quantity in data-poor situations or when the identifiability of a particular parameter or group of parameters is poor from the set of tests. Identifiability of the model in Equations 1 and 2 is discussed in Ref. 11, where it is related to the nonsingularity of the Fisher information matrix. Experience[20] has shown that even if some parameters are poorly identifiable (i.e., not well estimated), their combined projection into some higher-level (e.g., lower-dimensional) performance quantity such as trajectory uncertainty or target accuracy (CEP) may still be reasonably approximated by the asymptotic property. Nevertheless, the diagonal elements of $\mathscr{F}_{\theta\theta^t}^{-1}$ (or its inverse) will indicate which components are poorly identifiable so that the tests can be modified (more data, better instrumentation, different test types). One can calculate $\mathscr{F}_{\theta\theta^t}^{-1}$ from covariance simulations of the tests (before any real testing), thus obtaining the basis for test planning and design. Required confidence intervals in a high-level system performance criterion can be translated into test requirements by iteratively simulating various combinations of test types, test sizes, and instrumentation and then calculating $\mathscr{F}_{\theta\theta^t}^{-1}$ and the resulting confidence intervals until the confidence interval criterion is satisfied.

A portion of the elements in θ may not be identifiable within the practical constraints on the test instrumentation and test program. In this case some *a priori* knowledge (from subjective judgment or other tests) about these elements may be used.[21] First, the identifiable parameters are estimated, assuming the unidentifiable elements are known. Then their Fisher information matrix is adjusted to account for the uncertainty in the unidentifiable elements. The Fisher calculation in our technique[1] could be used to calculate the required adjustments.

SIMPLIFIED MISSILE INERTIAL GUIDANCE EXAMPLE

To illustrate the efficacy of the identification technique, a very simplified missile inertial system test and evaluation problem will be considered. The performance of the inertial system is based on its error model, which is derived from first principles and is shown in Figure 2. The inertial instruments are on a local-level, stabilized platform; vehicle motion is constrained in two dimensions (north-vertical) on a nonrotating Earth using a very good altimeter and where

An, Av	=	north and vertical specific force (ft/s^2)
Vn	=	north velocity (ft/s)
R_E	=	Earth radius
g	=	Earth gravity magnitude
$\delta Vn, \delta Rn$	=	errors in north velocity (ft/s) and position (ft)
ψe	=	"computer-to-platform" misalignment about east (μrad)
de	=	east gyro drift random walk (mrad/h)
fde	=	east gyro drift rate white noise
sde	=	east gyro torque scale factor bias (ppm)
an	=	north accelerometer offset Markov process (μg)
fan	=	north accelerometer offset rate white noise
$1/\beta$	=	north accelerometer offset Markov time constant (s)
san	=	north accelerometer scale factor bias (ppm)

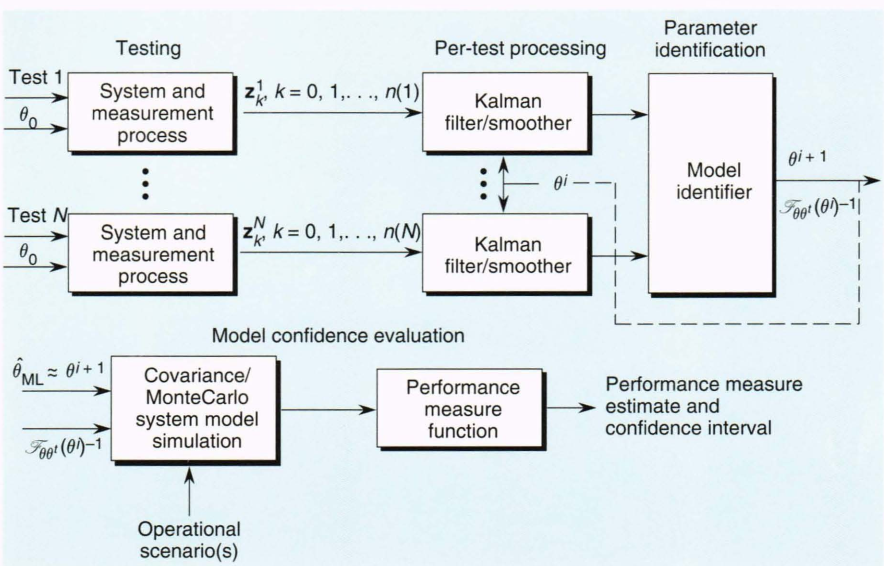


Figure 1. System identification flow.

An estimation of the statistical parameters in this inertial model from factory and system test data is desirable to validate the predicted performance from the manufacturer. The initial values of δVn and δRn are assumed to be zero with no uncertainty since the missile is to be launched from a fixed and surveyed land site. The unknowns in the system model are the initial means and variances of de, sde, san, and an; the spectral amplitude of the white noises fde and fan; and the time constant $1/\beta$. The initial value in ψe is assumed to be $-an/g$ because of the leveling process before launch.

The operational missile scenario (trajectory into actual targets) is an 8000-nm flight with a 150-s constant 5.5-g boost at a thrust angle of 29° from the local horizon. The inertial system is used to control a reentry body deployment (with negligible deployment error) at 600 s. This operational scenario can never be tested, so a combination of other system tests will be used to extract the scenario-independent parameters (the unknowns described above), which will then be projected via a covariance simulation of the operational scenario into an estimate of operational system performance (impact error RMS). Three types of tests will be used:

1. Factory gyro and accelerometer test-stand data.
2. High-performance aircraft testing of the missile INS on a test pallet using a Global Positioning System (GPS) precision-integrated Doppler track of sampled measurements of position error at 10-s intervals for 1/4 h with 0.1-ft (1σ) uncorrelated measurement noise. The assumed trajectory of the aircraft is a horizontal "figure eight" in the north-vertical plane with 25 nm between foci and each turn of radius 30,000 ft at a constant speed of 1000 kt.
3. Missile flight tests over a 6000-nm range with a 150-s constant 5.5-g boost at a thrust angle of 42°. A GPS precision postflight track of the same quality and sample rate as in the aircraft tests is assumed for 600 s of tracking.

The model in Figure 2 for each system test and the operational scenario can be expressed as a linear, continuous dynamics, state-space model given by (suppressing the j superscript and with the subscript t denoting continuous time)

$$\dot{\mathbf{x}}_t = \mathbf{A}_t \mathbf{x}_t + \boldsymbol{\omega}_t \qquad 0 \le t \le t_n , \tag{8}$$

where $\boldsymbol{\omega}_t$ is a seven-dimensional white noise vector having a spectral density matrix $\boldsymbol{\Omega}_t$. The measurements from the system tests are discrete time as in Equation 2. The appropriate EM and Scoring algorithms can be easily obtained by running the discrete-time Kalman filter and smoother at a small enough Δt (possibly smaller than the time between measurements) so that

$$\boldsymbol{\Phi}_k \cong \mathbf{I} + \mathbf{A}_k \Delta t_k \tag{9}$$

and

$$\mathbf{Q}_k \cong \boldsymbol{\Omega}_k \Delta t_k , \tag{10}$$

where $\mathbf{I}$ is the identity matrix.

In this case,

$$\mathbf{x}_t = \begin{bmatrix} \psi e \\ \delta Vn \\ \delta Rn \\ de \\ sde \\ an \\ san \end{bmatrix}; \tag{11a}$$

$$A_t = \begin{bmatrix} 0 & 0 & 0 & 1 & -Vn/R_E & 0 & 0 \\ Av & 0 & -g/R_E & 0 & 0 & 1 & An \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\beta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \tag{11b}$$

$$\boldsymbol{\omega}_t = \begin{bmatrix} 0 \\ 0 \\ 0 \\ fde \\ 0 \\ fan \\ 0 \end{bmatrix}; \tag{12a}$$

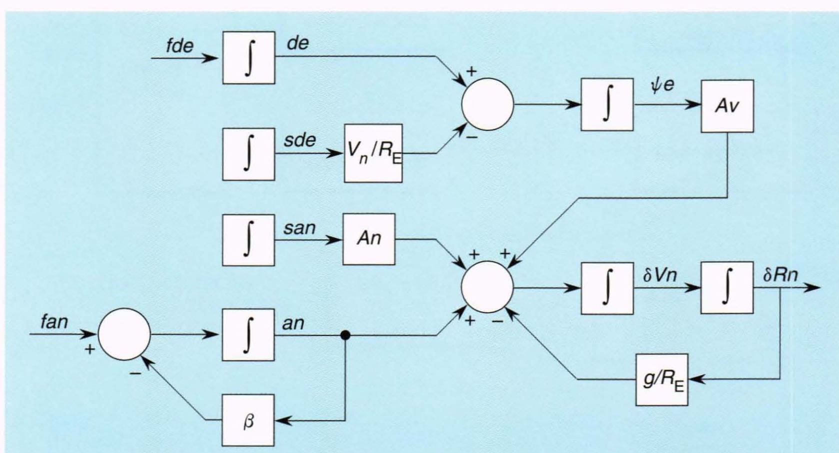


Figure 2. Simplified inertial system error model.

$$\mathbf{\Omega}_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{fde} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{fan} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \qquad (12b)$$

$$\boldsymbol{\mu}_0 = \begin{bmatrix} -\mu_{an} \\ 0 \\ 0 \\ \mu_{de} \\ \mu_{sde} \\ \mu_{an} \\ \mu_{san} \end{bmatrix}; \qquad (13a)$$

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \sigma^2_{an} & 0 & 0 & 0 & 0 & -\sigma^2_{an} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{de} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{sde} & 0 & 0 \\ -\sigma^2_{an} & 0 & 0 & 0 & 0 & \sigma^2_{an} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{san} \end{bmatrix}; \qquad (13b)$$

where the associated discrete-time measurement is given by Equation 2 wherein

$$\mathbf{H}_k = [\,0\ 0\ 1\ 0\ 0\ 0\ 0\,], \qquad (14)$$

and measurement noise covariance $\mathbf{R}_k = (0.1\ \text{ft})^2$ for aircraft and missile testing.

In this example, $\boldsymbol{\theta}$ is the vector of unknown parameters,

$$\boldsymbol{\theta} = \left[\mu_{de}\ \mu_{sde}\ \mu_{an}\ \mu_{san}\ \sigma^2_{de}\ \sigma^2_{sde}\ \sigma^2_{an}\ \sigma^2_{san}\ \beta\ \sigma^2_{fde}\ \sigma^2_{fan}\right]^t. \qquad (15)$$

The assumed true values of $\boldsymbol{\theta}$, shown in Table 1, cause (via simulation) an operational impact error of 1200 ft RMS. The objective is to process combinations of factory, aircraft, and missile test flights to estimate $\boldsymbol{\theta}$ and transform (via covariance simulations as in Fig. 1) the model in Equations 8 through 13 with $\hat{\boldsymbol{\theta}}_{\text{ML}}$ into an operational impact RMS along with a confidence interval derived from the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{ML}}$. The nominal values of $\boldsymbol{\theta}$ (our best guess from engineering judgment) are also shown in Table 1 and cause an operational impact error of 1900 ft RMS. Models for the two system tests, aircraft and missile, have the same dimension and are structurally similar except for differing acceleration profiles.

The simplified gyro factory tests feature mounting the gyro on a precision test turntable with the gyro input axis orthogonal to the Earth rotation vector and the gyro signal generator output driving the turntable motor. With no input to the gyro torquer, the turntable would stay fixed relative to the Earth except for gyro drift effects. A known torquer input, T_g, will cause the table to rotate at a nominal rate of T_g. Each gyro test will involve torquing the gyro ±100°/h for 1 h each; the turntable orientation is

Table 1. Test and evaluation results for missile guidance example (the 90% confidence interval = $\hat{\boldsymbol{\theta}}_{\text{ML}} \pm \alpha$).

$\boldsymbol{\theta}$	Units	$\boldsymbol{\theta}$ (nominal)	$\boldsymbol{\theta}$ (true)	$\hat{\boldsymbol{\theta}}_{\text{ML}}$	α (±)
μ_{de}	mrad/h	0	0.4	0.31	0.09
μ_{sde}	ppm	0	0	–151	106
μ_{an}	μg	0	0	0.27	3.2
μ_{san}	ppm	0	0	–0.41	1.4
σ^2_{de}	$(\text{mrad/h})^2$	0.16	0.09	0.089	0.04
σ^2_{sde}	$(\text{ppm})^2$	40,000	90,000	120,540	55,800
σ^2_{an}	$(\mu g)^2$	400	100	118	48.3
σ^2_{san}	$(\text{ppm})^2$	50	16	20.7	10.6
β	1/s	0.001	0.0025	0.0044	0.0008
σ^2_{fde}	$(\text{mrad/h})^2/\text{s}$	0.0009	0.0001	0.00004	0.000015
σ^2_{fan}	$(\mu g)^2/\text{s}$	1.0	0.25	0.248	0.019
Total impact error RMS (ft)		1900	1200	1290	193 (±15%)

sampled every 2 min with an angular noise error of 30 μrad (1σ). The accelerometer is tested on the precision turntable by rotating its input axis 180° from local horizontal to vertical to local horizontal in 20 min, measuring the integrated accelerometer output every 20 s with a noise error of 0.001 ft/s (1σ). The error models for these tests are shown in Figure 3, where n_g and n_a represent the measurement noise in the angular and velocity measurements, respectively.

As before, these error models can be expressed by the generic form in Equations 2 and 8 through 10, where the gyro test is given by

$$\mathbf{x}_t = \begin{bmatrix} \psi \\ de \\ sde \end{bmatrix};\ \mathbf{A}_t = \begin{bmatrix} 0 & 1 & -T_g \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \qquad (16a)$$

$$\boldsymbol{\omega}_t = \begin{bmatrix} 0 \\ fde \\ 0 \end{bmatrix};\ \mathbf{\Omega}_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma^2_{fde} & 0 \\ 0 & 0 & 0 \end{bmatrix}; \qquad (16b)$$

and

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0 \\ \mu_{de} \\ \mu_{sde} \end{bmatrix};\ \mathbf{\Sigma}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma^2_{de} & 0 \\ 0 & 0 & \sigma^2_{sde} \end{bmatrix}; \qquad (17a)$$

$$\mathbf{H}_k = [\,1\ \ 0\ \ 0\,];\ \mathbf{R}_k = (30\ \mu\text{rad})^2. \qquad (17b)$$

The accelerometer test is given by

$$\mathbf{x}_t = \begin{bmatrix} \delta V \\ an \\ san \end{bmatrix};\ \mathbf{A}_t = \begin{bmatrix} 0 & 1 & A \\ 0 & -\beta & 0 \\ 0 & 0 & 0 \end{bmatrix}; \qquad (18a)$$

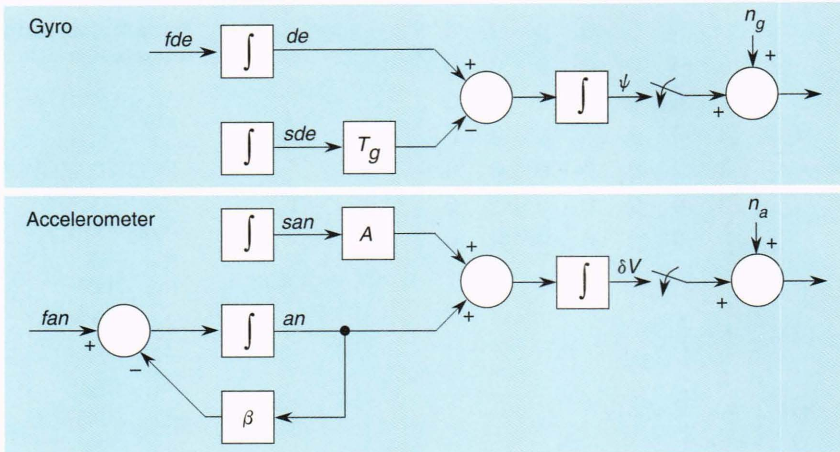


Figure 3. Simplified gyro and accelerometer error models for factory testing (n_g and n_a represent measurement noise in angular and velocity measurements, respectively; $A = g \sin \omega_a t$; $\omega_a = \pi/1200$ rad/s).

$$\omega_k = \begin{bmatrix} 0 \\ fan \\ 0 \end{bmatrix}; \ \Omega_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma^2_{fan} & 0 \\ 0 & 0 & 0 \end{bmatrix}; \tag{18b}$$

and

$$\mu_0 = \begin{bmatrix} 0 \\ \mu_{an} \\ \mu_{san} \end{bmatrix}; \ \Sigma_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma^2_{an} & 0 \\ 0 & 0 & \sigma^2_{san} \end{bmatrix}; \tag{19a}$$

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}; \ \mathbf{R}_k = (0.001 \text{ ft/s})^2 . \tag{19b}$$

Note that the gyro and accelerometer factory tests are structurally and dimensionally very different from the system tests. Combined, however, they will yield the same type of model information as the system tests.

NUMERICAL RESULTS

We conducted a "test planning" analysis to determine the appropriate combinations and numbers of different tests to achieve a system confidence requirement; that is, the true operational RMS was to be within ±15% of the estimated operational RMS with 90% confidence. To achieve this confidence by only traditional impact scoring (sampling target errors from many tests) of test trajectories (not an operational estimate), we would need about thirty-five repeatable test flights with perfect impact location instrumentation. Obviously, we hope to require significantly fewer missile tests by extracting more information per test and by using information from other test types (e.g., aircraft or factory). The test planning analysis follows from Figure 1 where a Kalman filter/smoother covariance analysis is run for each generic type of test, and $\boldsymbol{\theta}$ is set to the nominal model (our best guess). The model identifier is then run in a "covariance only" mode to produce a Fisher information matrix for each type of test. Various combinations and numbers of tests can be merged by simply adding the appropriate Fisher information matrices. The inverse of the combined Fisher matrix along with the nominal $\boldsymbol{\theta}$ defines a Gaussian distribution from which sampled values of $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ can be transformed into impact means and covariances via an operational inertial model covariance simulation as shown in Figure 1. A distribution of RMS's is then obtained from the distribution of impact means and covariances, yielding the appropriate confidence intervals.

The analysis showed that twenty-nine test missiles would be needed to achieve the operational confidence requirement if the model was truly at the nominal values. The aircraft tests provided more information per test as only eighteen tests yielded acceptable confidence intervals. By combining the two types of system tests and recognizing the relatively high cost of missile testing, we found that a set of four missile tests (to exercise and test other missile subsystems) and fifteen aircraft tests satisfied the operational requirement. Obviously, many other combinations could be used. In fact, the tests themselves could be modified to produce more information per test. For example, by lengthening the aircraft tests to one hour, a combination of four missile and seven aircraft tests satisfied the requirement.

Factory tests provided less information than either of the other two types of tests, requiring 100 accelerometer/gyro tests to yield acceptable confidence intervals if the model is truly at the nominal values. When combined with a few aircraft tests, however, a significant reduction in the required number of factory tests occurs. For example, the following combinations of aircraft and factory tests, respectively, met system test requirements; 0, 100; 6, 20; 12, 6; 18, 0. The factory tests alone have combinations of parameters in $\boldsymbol{\theta}$ that are poorly observable, resulting in high uncertainty in the tactical impact RMS projections. The aircraft tests provide significantly better observability of these same combinations, resulting in a complementary combination.

In test planning analysis, one has to guess conservatively at the range of possible "true" values that the model could take on to provide enough tests of adequate information. As the true values get smaller with constant test instrumentation quality and number of tests, the percentage confidence intervals will grow. In other words, achieving the confidence requirement on a better guidance system will require more or better test resources. We therefore chose a conservative set of four missile tests, fifteen aircraft tests (1/4-h long), and fifteen factory tests that would have produced ±11% confidence intervals at the nominal model. This test program was then simulated to exercise the parameter estimation software shown in Figure 1. The results are shown in Table 1, with $\boldsymbol{\theta}$ (nominal) being the starting "guessed" value in the parameter estimation iterative process and $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ the final estimate of $\boldsymbol{\theta}$ (true). The 90% confidence intervals were generated from the Fisher information matrix inverse evaluated at $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$. The 90% confidence intervals cover the true values for all parameters except for $\boldsymbol{\mu}_{sde}$, $\boldsymbol{\beta}$, and $\boldsymbol{\sigma}^2_{fde}$. This Fisher information matrix inverse is somewhat optimistic at the individual component level, being a lower bound on the estimation error covariance. Also, the evaluation of the Fisher information at $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is approximate since it is theoretically defined only at the true $\boldsymbol{\theta}$. One must therefore regard these confidence intervals as indications of parameter uncertainty that must be tempered by analyst experience. The overall impact RMS estimate for the operational trajectory is very good and well within the confidence intervals, however.

The convergence of the iteration process is shown in Figures 4, 5, and 6. Figure 4 shows the likelihood function evaluated after each iteration to verify its increase. The first ten iterations use the EM algorithm, which exhibits the classical EM behavior of slowing down in convergence rate as the iterate nears the maximum. The switch (when it "looks" as though EM has sufficiently slowed down) to Scoring for the last ten iterations extracts the last bit of information from the tests. The projection of each iterate to operational impact RMS shows sufficient convergence by the fifth iteration, however. Figure 5 shows similar rapid convergence characteristics in those individual parameter estimates. The benefit of continuing the iteration can be seen in Figure 6, where the individual parameter estimates are markedly improved by the Scoring iterations, yielding more accurate understanding of the detailed model. In any event, one iteration of Scoring would be needed to evaluate the Fisher information matrix for confidence interval calculation.

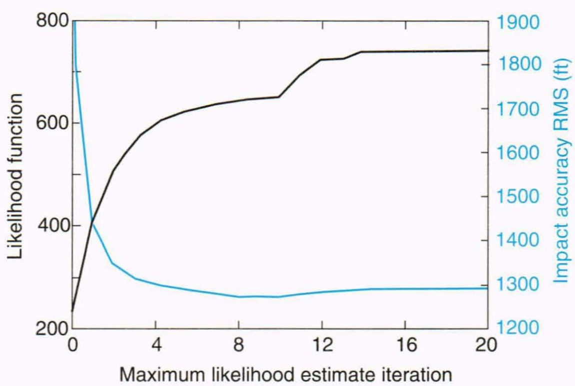


Figure 4. Convergence of likelihood function and projected operational accuracy estimate.

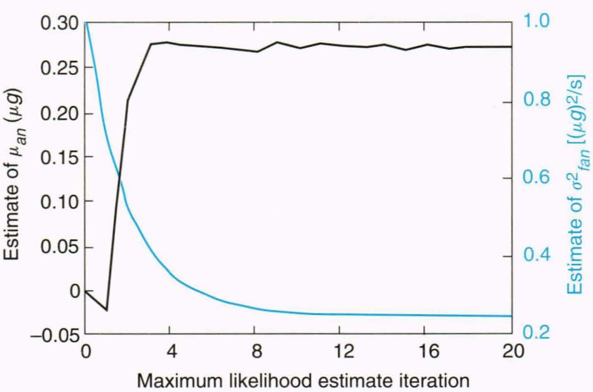


Figure 5. Parameter estimates with rapid convergence.

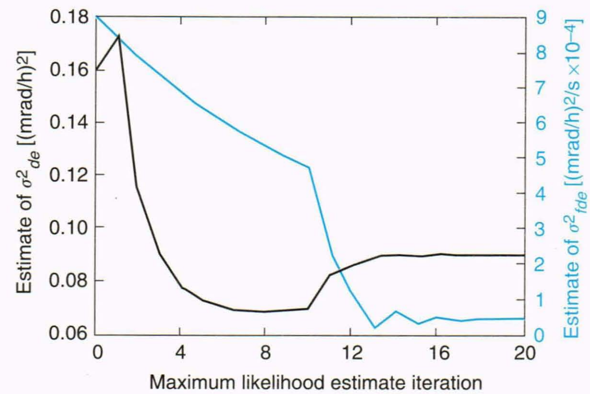


Figure 6. Parameter estimates with slow convergence.

The example we have presented is illustrative, but not truly representative, of the potential of this methodology. Realistic guidance models are much more complex with more dynamic acceleration profiles. These profiles should enable more detailed model information to be extracted per test, producing an even wider disparity between traditional impact scoring and the methodology presented in this article.

CONCLUSIONS

We have presented and demonstrated a new approach to system test and evaluation that uses detailed test data to estimate the underlying ensemble system model, which can then be extrapolated to untested scenarios for tactical system performance estimates. In addition, this methodology provides its own estimation error distribution, which allows confidence statements to be made for the model estimates and extrapolations and enables test planning analysis to be conducted before the test program for the more efficient use of test facilities.

REFERENCES

[1] Levy, L. J., and Porter, D.W., "Unified EM/Scoring Identification Algorithms for Linear State Space Models," in *Proc. 25th Annual Conf. on Information Sciences and Systems*, Baltimore, Md., pp. 91–101 (1991).

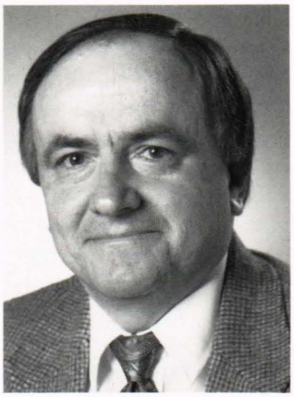
[2] Basseville, M., and Benveniste, A. (eds.), "Detection of Abrupt Changes in Signals and Dynamical Systems," in *Lecture Notes in Control and Information Sciences*, Vol. 77, Springer-Verlag (1980).

[3] Gupta, N. K., and Hall, W. E., "System Identification Technology for Estimating Reentry Vehicle Aerodynamic Coefficients," *J. Guid. Control* **2**(2), 139–146 (1979).

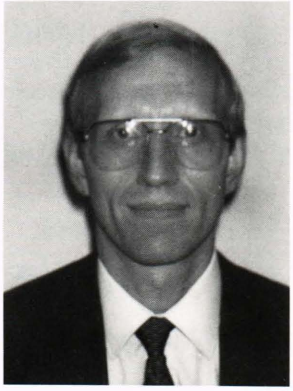
[4] Illiff, K. W., and Maine, R. E., *More Than You May Want to Know About Maximum Likelihood Estimation*, AIAA Paper #84-2070 (1984).

[5] Kain, J. E., "An Evaluation of Aeroballistic Range Projectile Parameter Identification Procedures," in *AIAA Atmospheric Flight Mechanics Conf.* (1979).
[6] Chen, C. W., Huang, J. K., Phan, M., and Juang, J. N., "Integrated System Identification and Model State Estimation for Control of Flexible Space Structures," in *AIAA-90-3470-CP*, pp. 1396–1404 (1990).
[7] Gibbs, B. P., *Geomagnetic Field Modeling Errors and Optimization of Recursive Estimation*, BTS14-87-17, Coleman Research Corp. (Jan 1988).
[8] Nakamura, H., and Akaike, H., "Use of Statistical Identification for Optimal Control of a Supercritical Thermal Power Plant," in *5th IFAC Symp. on Identification and System Parameter Identification* (1979).
[9] Katsaggelos, A. K., and Lay, K. T., "Maximum Likelihood Blur Identification and Image Restoration Using the EM Algorithm," *IEEE Trans. Signal Proc.* **39**(3), 729–733 (1991).
[10] Peloubet, R. P., Haller, R. L., and Bolding, R. M., "On-Line Adaptive Control of Unstable Aircraft Wing Flutter," in *29th IEEE Conf. on Decision and Control* (1990).
[11] Goodrich, R. L., and Caines, P. E., "Linear System Identification for Nonstationary Cross-Sectional Data," *IEEE Trans. Autom. Control* **AC-24**(3), 403–411 (1979).
[12] Gupta, N. K., and Mehra, R. K., "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations," *IEEE Trans. Autom. Control* **AC-19**(6), 774–783 (1974).
[13] Meilijson, I., "A Fast Improvement to the EM Algorithm on Its Own Terms," *J. R. Statist. Soc. B* **51**(1), 127–138 (1989).
[14] Shumway, R. H., Olsen, D. E., and Levy, L. J., "Estimation and Tests of Hypothesis for the Initial Mean and Covariance in the Kalman Filter Model," *Commun. Statist. Theory Methods* **A10**(16), 1625–1641 (1981).
[15] Levy, L. J., Shumway, R. H., Olsen, D. E., and Deal, F. C., "Model Validation from an Ensemble of Kalman Filter Tests," in *21st Midwest Symp. on Circuits and Systems*, Iowa State University, pp. 461–466 (1978).
[16] Smith, R. H., "Maximum Likelihood Identification of Structurally Constrained Covariances for Initial Condition Statistics," in *Proc. American Control Conf.*, pp. 1242–1248 (1986).
[17] Shumway, R. H., and Stoffer, D. S., "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *J. Time Series Anal.* **3**(4), 253–254 (1982).
[18] Wu, C. F. J., "On the Convergence Properties of the EM Algorithm," *Ann. Statist.* **11**(1), 95–103 (1983).
[19] Segal, M., and Weinstein, E., "A New Method for Evaluating the Log-Likelihood Gradient, the Hessian, and the Fisher Information Matrix for Linear Dynamic Systems," *IEEE Trans. Inf. Theory* **35**(3), 682–687 (1989).
[20] Haley, D. R., "Validation of Asymptotic Fisher Theory by Monte Carlo Simulation," in *Proc. 1987 Summer Computer Simulation Conf.*, pp. 185–190 (1987).
[21] Spall, J. C., and Garner, J. P., "Parameter Identification for State-Space Models with Nuisance Parameters," *IEEE Trans. Aerosp. Electron. Sys.* **26**(6), 992–998 (1990).

THE AUTHORS



LARRY J. LEVY is Chief Scientist of APL's Strategic Systems Department. He received his Ph.D. in electrical engineering from Iowa State University in 1971. After joining APL, he became the SATRACK Project Scientist from 1974 to 1980 and the Accuracy Evaluation System Assistant Program Manager from 1979 to 1982, and as such was a principal developer of the technology in each project. From 1981 to 1990, Dr. Levy was the Assistant Supervisor and then Supervisor of the Systems Evaluation Branch in APL's Strategic Systems Department. He has also served as the Exoatmospheric Reentry Vehicle Interceptor Subsystem Program Manager (1989–1990), the Air Force Missile Tracking Program Manager (1990), and a G.W.C. Whiting School of Engineering lecturer (1986–present) on random signal processing and model building for dynamical systems.



DAVID W. PORTER is a Technical Director and the General Manager of the Computational Engineering Division of Coleman Research Corporation in Laurel, Maryland. He received his B.S., M.S., and Ph.D. degrees in electrical engineering from Iowa State University in 1969, 1970, and 1972, respectively. Dr. Porter leads the development of workstation products and services for simulation, state estimation, parameter identification, signal processing, and control for industry and government. He has published widely on stability theory, state estimation, and parameter identification.