The Systems Analysis, Test, and Evaluation of Strategic Systems

Larry J. Levy

The systems analysis, test, and evaluation of strategic systems is a "preeminent technical leadership role" for APL. This activity encompasses the planning, design, development, operation, and performance assessment of the Trident II Weapon System, uniquely providing confidence-based performance assessments over untested trajectories. Physics-based scenario-independent statistical models are cumulatively fit to operational tests by maximum likelihood estimation techniques for maximum extraction of model information. The estimated model propagated into the performance factor domain provides the performance factor estimates and computable estimation error statistics for confidence interval estimation. The flight test restricted environment of the present aging weapon system, new global strike missions, and the ballistic missile defense system will present new technical challenges to providing confidence-based evaluations.

INTRODUCTION

The U.S. national leadership recognized that not understanding how well our strategic deterrent (offensive) systems would perform (i.e., with quantified confidence) would be unacceptable, setting specific quantitative guidelines for testing and evaluating of these systems in classified requirements as early as 1966. The need was not just knowing how well the systems would perform but also how confident we could be in our prediction. Quantified confidence is knowing the system's performance to within a quantified uncertainty (confidence interval). It is statistically knowing what you do not know about the system's performance. Building a weapon system with a good performance estimate (e.g., high reliability) but with a large confidence interval (high uncertainty) about that estimate could be dangerous. This was recognized early on in Submarine Launched Ballistic Missile (SLBM) development and testing, requiring a top-down systems engineering approach to define the test programs. Being able to extrapolate test results to the tactical domain became more important as the accuracy requirement became more stringent. Traditional testing by "shoot and score" could not satisfy the top-level evaluation requirements cost-effectively.

In this article, the approach to SLBM test and evaluation (T&E) is presented from three perspectives: (1) the top-down systems engineering approach that produced the T&E system, (2) a description of the T&E system as applied to Trident II, and (3) a discussion of future technical challenges that could be addressed.

SYSTEMS ENGINEERING APPROACH TO TEST AND EVALUATION

APL's systems engineering approach to T&E is shown in Fig. 1. This was extrapolated from experience with previous weapon systems T&E and especially that of Trident II. The approach is discussed generically here to illustrate its use for other weapon systems as well. The left side of Fig. 1 illustrates the planning steps required to properly design an overall test program to provide adequate evaluation capability at certain milestones in the test program. The right side describes the execution steps in the T&E process. This process can be rather elaborate, as it was for Trident, or simpler, as for nonstrategic systems, depending on the system type, stage in the acquisition process, and APL's role.

The key starting point in the systems engineering approach is specifying the top-level performance evaluation requirements (not how well the weapon system should perform, but how well we should know its performance, i.e., confidence). A few test successes do not guarantee that the system will meet its objectives; it only shows that success is possible. If there are no top-level measures of effectiveness (MOEs) evaluation requirements in terms of confidence, then one can be developed. This would be an iterative process involving developer, evaluator, and user.

The next step is to determine a complete set of lower-level measures of performance (MOPs) with associated confidence requirements over a reference set of scenarios needed to achieve the required MOE and confidence bound. Testable MOPs (or ones that are extrapolated from tests) are sampled from distributions commensurate with assumed confidence bounds, and scenario simulations are used to calculate the resulting MOEs (and confidence bounds). This process is iterated until an optimized set of MOPs (and confidence bounds) is achieved. A possible optimization strategy might be to "balance" the contributions of each MOP confidence contribution to MOE confidence. Other strategies might reflect the difficulty (e.g., cost) in achieving certain MOP confidence such as reliability. Many trade-offs could be evaluated.

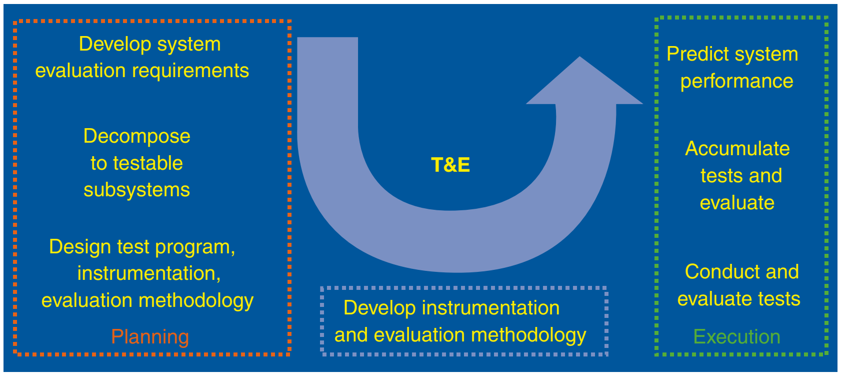

Figure 1. APL's systems engineering approach to T&E.

A test program and analysis methodology are then designed to meet each MOP confidence requirement by hypothesizing various feasible tests (system, subsystem, component), test sizes, instrumentation quality, and evaluation methodologies. Appropriate simulation models (covariance or Monte Carlo) are used to evaluate each hypothesized set until an optimized set is obtained. The results of this phase might require going back to the previous phase to revise the required MOP confidence bounds.

Such a process provides trade-offs while quantifying the implications of decisions to test more (or less), to instrument different functions or systems, or to change the quality of the instruments. As defense spending and costs associated with system development and T&E come under increasing scrutiny, it becomes even more important to be able to quantify the relative benefits of test size and instrumentation quality. Quantifying the confidence with which we will know system performance provides a metric by which we can assess the value of our test programs, instrumentation, and analysis approaches.

To execute the steps of the T&E process (right side of Fig. 1), tests are conducted by traditional testers and evaluators, but with the evaluation outputs complying with the system evaluator's requirements. Test types include system, component, or subsystem tests; monitoring of an in-place system as it awaits operational use; and subsystems assessment "in-the-loop" of a simulation. Detection/isolation of faults on each test is conducted by traditional tester/evaluators, but again with results validated by the system evaluator. Isolated faults are fixed by the developer and removed from the database and models.

The system evaluator calculates a cumulative update of the MOP models, confidence intervals, and estimated distributions. Physics-based models to fit data (system identification) from diverse tests are used where possilbe to gain maximum information from each test. If the model can be broken down into a set of parameters that are independent of scenario, then statistical leverage can be gained by accumulating across all relevant but disparate tests.[1] The associated uncertainty (confidence bound) in the model estimates is calculated from the known observability, instrumentation quality, and number of tests. Prior information and tests from development testing can also be used initially until an adequate number of post-deployment tests can be accumulated. Periodic reassessment of the test program's adequacy to estimate the MOPs and associated confidences may require feedback to the planning stages to reassess the confidence requirements.

Next, the system evaluator predicts the MOE and confidence bounds for the required reference set of scenarios using the force-level simulations to flow up the MOPs (and confidences bounds) to MOEs (and confidence bounds). Model fault isolation follows to determine which MOP is out of specification and its resultant contribution to the MOE. Periodic reassessment of the test program adequacy for current MOE requirements must be done.

Finally, the system evaluator conducts force-level evaluations with the latest estimated models by using force-level simulations to flow up the estimated MOPs (and confidences bounds) to MOEs (and confidence bounds) to evaluate the adequacy of the systems for many different campaigns. This allows trade-offs to be made for optimal planning of the force-level deployment such as in ballistic missile defense.[2] The evaluator can also develop and update a functionalized performance prediction model to be used in the real-time employment of the weapon system against an operational threat.

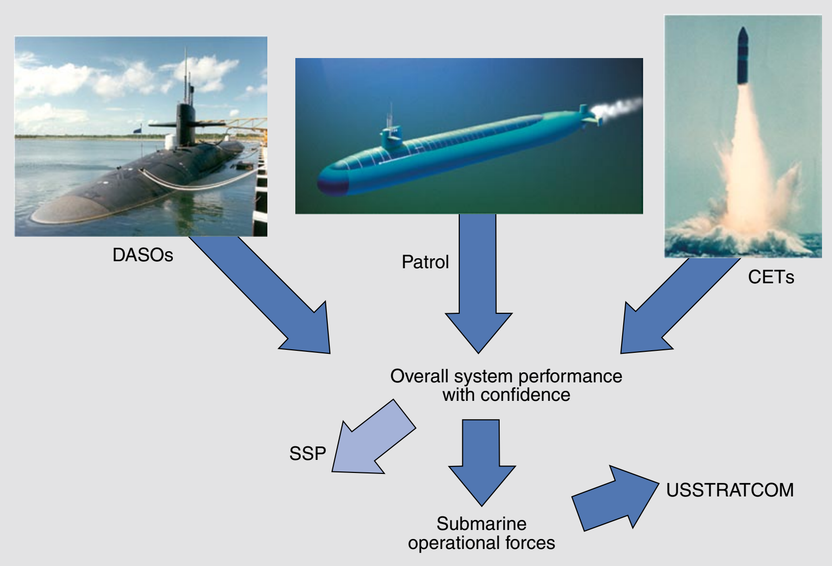

Figure 2. Strategic deterrence systems T&E.

STRATEGIC DETERRENCE TEST AND EVALUATION

Because of the national importance of our strategic deterrent systems, APL instituted a T&E program of the highest caliber that began in the late 1950s for the Navy's Fleet Ballistic Missile Strategic Weapon System, sponsored by Strategic Systems Programs (SSP). The SLBM on its nuclear-powered submarine platform provides a mobile, long-patrol duration, covert, and invulnerable strategic deterrent force. Figure 2 depicts the three major types of system testing of the SLBM: (1) demonstration and shakedown operations (DASOs), i.e., flight testing that is conducted before deployment after either new submarine construction or a shipyard overhaul period; (2) patrol, i.e., recurring nonflight tests conducted during each strategic deterrent patrol; and (3) Commander-in-Chief (CINC) evaluation tests (CETs) or follow-on CETs (FCETs), i.e., end-to-end weapon system tests, including missile flights, conducted with randomly selected missiles periodically throughout the life of the system. The results of the evaluations are provided directly to the Fleet Commands, which then present them to the U.S. Strategic Command (USSTRATCOM) for strategic targeting requirements. In this way APL's T&E is considered "independent" of the developer, SSP.

The scope of these ongoing evaluations encompasses about 220 staff years per year and is the largest concentration of T&E expertise at the Laboratory. SLBM T&E was developed using the full scope of the systems engineering approach described previously. The major S&T innovations—SATRACK, the Accuracy Evaluation System (ACES), and Trident II accuracy—are detailed next.

SATRACK, developed in the late 1970s, uses GPS satellites to precisely track Trident missiles from DASO and CET tests. As illustrated in Fig. 3, the GPS satellite radiates to the test missile containing a GPS translator (instead of a receiver), which relays the raw

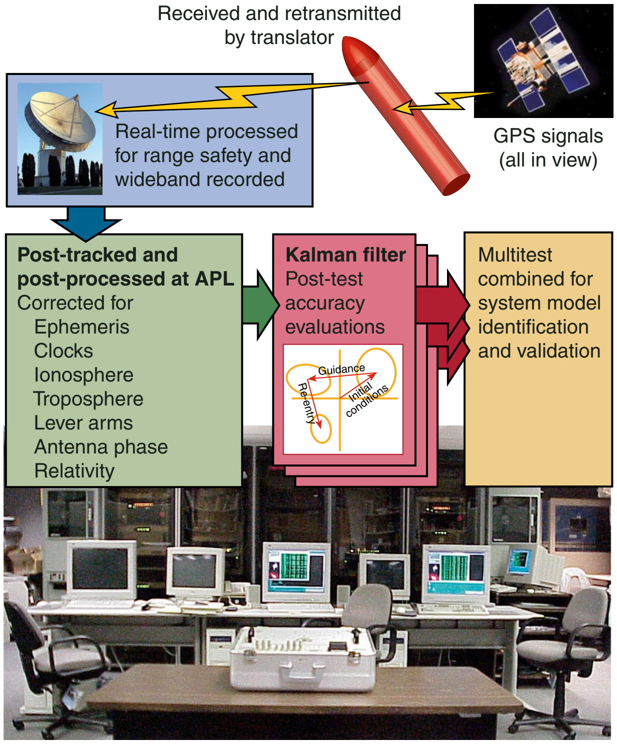

Figure 3. SATRACK for missile accuracy evaluation.

navigation signals to land- and sea-based receiving stations for wideband recording. The recordings are tracked/corrected following the test at the APL SATRACK facility and processed in a large Kalman filter along with missile telemetry for estimation of individual guidance system errors. These estimates can then be propagated to the target point to explain the observed test miss.

Since Trident II was to have a more stringent accuracy requirement, the ACES study, conducted in 1980–1981, used the systems engineering approach to develop system evaluation requirements in terms of accuracy confidence. Instrumentation, test programs, and processing methodology were then determined to satisfy the confidence requirements, resulting in the instrumentation suite shown in Fig. 4. Flight testing then featured an improved SATRACK system for powered flight, inertial instrumentation for deployment and reentry, and improved underwater navigation instrumentation for the prelaunch phase. The major new addition from the ACES study was the cumulative model estimation with confidence, where the per-test results from each test were accumulated via a maximum likelihood method as shown in Fig. 5. Here, a physics-based model of the system, where the unknown parameters are fundamental errors (e.g., gyro drifts) common across all tests, is fit to all the data (even though the test scenarios are different) to estimate the underlying system model and the associated confidence. This results in an estimated model (vs. a validated model) capable of predicting accuracy performance to untested conditions with quantified confidence. The new accuracy modeling, coupled with the traditional reliability modeling, enabled Trident II performance to be predicted with quantified confidence. Starting with Trident I in the late 1970s, more than 180 flights have been processed by SATRACK, with about 100 being Trident II.

CHALLENGES FOR THE FUTURE

SLBM systems will require life extensions, and new missions are being considered such as global flexible response precision strike with low collateral damage. Budget constraints will limit traditional flight testing, requiring new reliability evaluation techniques and

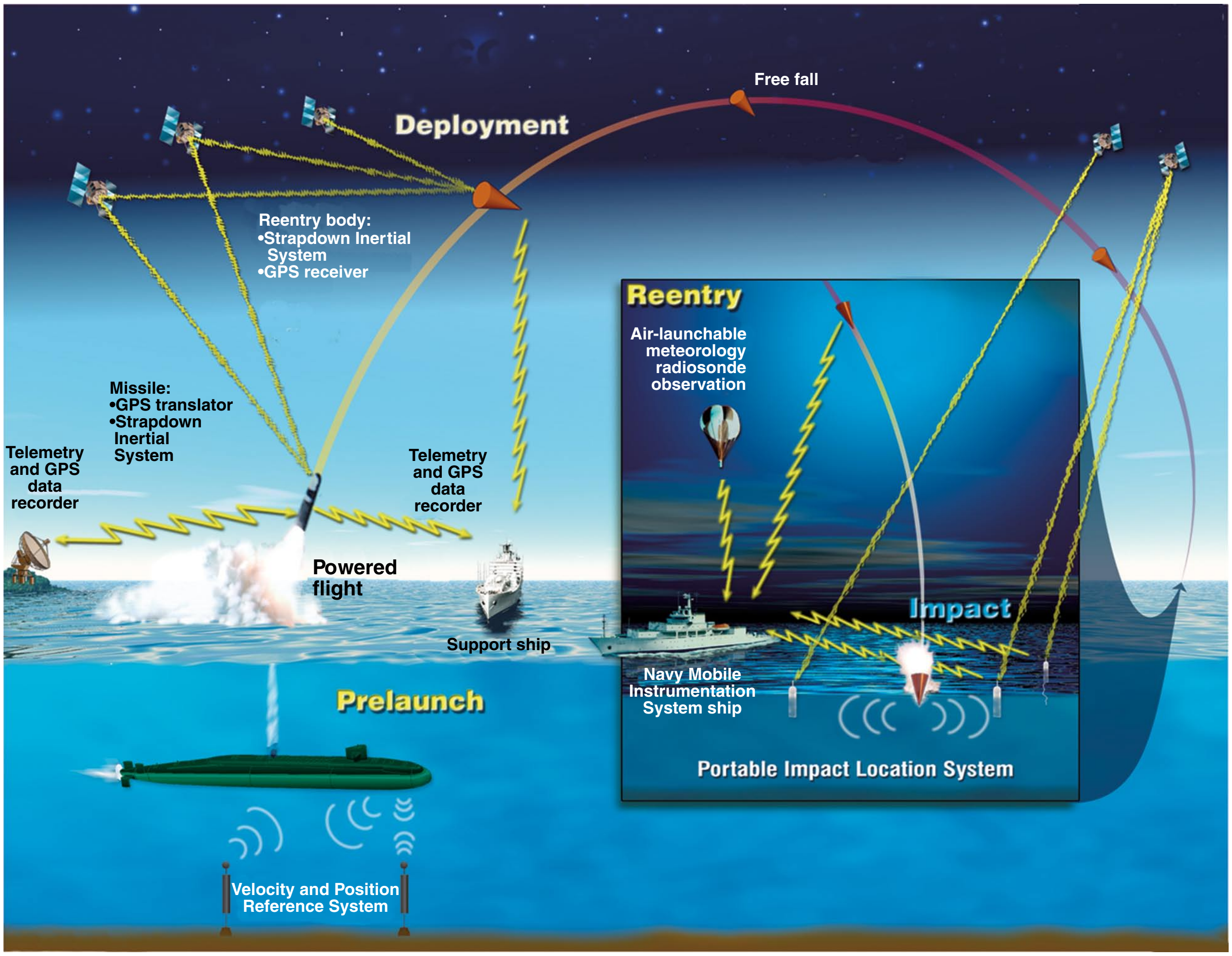

Figure 4. Trident II instrumentation.

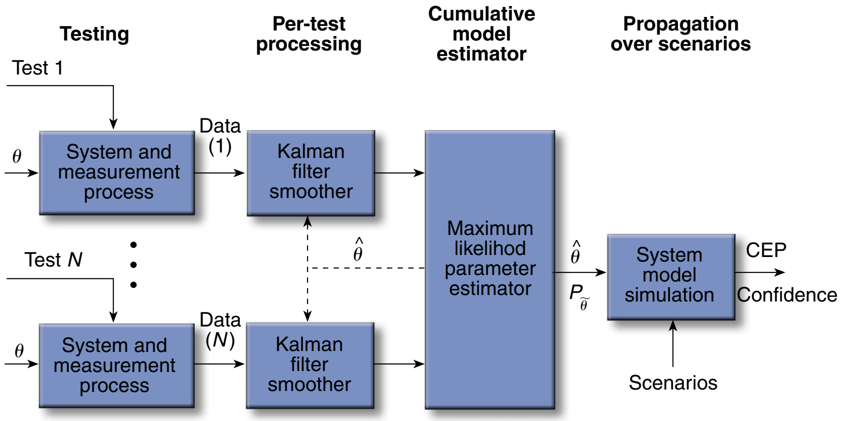

Figure 5. Model estimation for Trident II resulting in the credible performance prediction of a critical system to the government and military system. (θ = true model parameter vector, $\hat{\theta}$ = estimate of θ, $P_{\tilde{\theta}}$ = covariance of estimation error in θ.)

other new testing and instrumentation approaches. Additional test data will be needed to offset the lack of flight testing (there is no "free lunch"). Simulations per se provide no new information. Extensive subsystem ground tests with representative vibration/shock, thermal, and depressurization environments plus centrifuge and high-acceleration aircraft tests can nonsimultaneously replicate the missile environment. New processing methodologies, such as Bayesian Hierarchical Modeling,[3] can be used to appropriately combine ground and aircraft tests with traditional testing. All of these testing and processing methods must be able to provide quantifiable confidence to the performance predictions.

The importance of defending against ballistic missiles with strategic warheads (nuclear and chemical/biological) will require credibility (confidence) in ballistic missile defense performance on the same scale as for our Trident SLBM. This will require a paradigm shift from the traditional defensive systems T&E approach to provide quantified confidence in the performance assessments. The same systems engineering approach to T&E must flow down top-level force-on-force evaluation requirements into detailed subsystem evaluation requirements, followed by appropriate T&E of the subsystems and limited end-to-end tests. All types of testing providing usable performance information will be needed. High-fidelity force-on-force simulations will then propagate scenario-independent parameter estimates and confidences to top-level performance factors.[2] An independent system-of-systems evaluator will be needed to integrate all areas of subsystem T&E with the few available system-of-systems tests.

SUMMARY

Confidence-based performance evaluations of large-scale, complex systems of systems have been demonstrated for the Trident II weapon system, providing a unique approach to systems T&E. It uses detailed physics-based models fit to representative test data to extract maximum information from all relevant tests, providing quantifiable confidence in the model predictions on untested scenarios. Extension of this approach to new critical systems such as ballistic missile defense is possible in principle and necessary to ensure mission success.

REFERENCES

[1] Levy, L. J., and Porter, D. W., "Large-Scale System Performance Prediction with Confidence from Limited Field Testing Using Parameter Identification," *Johns Hopkins APL Tech. Dig.* **13**(2), 300–308 (1992).

[2] Mitchell, B. L., and Spriesterbach, T. P., "Confidence-Based Performance Assessments for the BMDO Family of Systems," in *Proc. AIAA 2nd Biennial Nat. Forum on Weapon System Effectiveness* (2001); http://www.stormingmedia.us/72/7226/A722604.html.

[3] Telford, J. K., *Estimating System Reliability with Confidence When Only Limited System Testing Is Available*, SSD-POR-03-7405, JHU/APL, Laurel, MD (Sep 2003).

THE AUTHOR

Larry J. Levy

Larry J. Levy is the Chief Scientist of APL's Strategic Systems Business Area. He received his Ph.D. in electrical engineering from Iowa State University in 1971, specializing in control theory. Dr. Levy has worked in the areas of navigation system development and system test and evaluation. He has more than 37 years of experience in Kalman filtering and system identification and is the author of numerous papers and presentations on these subjects. Dr. Levy was the co-developer of the GPS translator concept in SATRACK (a GPS-based missile tracking instrumentation system) and was instrumental in developing the end-to-end methodology for Trident II accuracy evaluation. He has developed multiple hypothesis fusion methods for multisensor, multitarget tracking and identification. He teaches graduate courses in Kalman filtering and system identification at the JHU Whiting School of Engineering as well as courses on these subjects for Navtech Seminars. His e-mail address is larry.levy@jhuapl.edu.