2008 Annual ITEA International Symposium

Verification and Validation Standards to Test and Evaluate New Complex Systems for the National Airspace System

John Frederick
Air Traffic Organization NextGen and Operations Planning Services,
Federal Aviation Administration
Catherine Jaggard
Operational System 2nd Level Engineering, Apptis Corporation
Mike Paglione
Simulation & Analysis Group, Federal Aviation Administration
Clifton Baldwin
Systems Engineering, Federal Aviation Administration

Contact Data: FAA William J. Hughes Technical Center
Atlantic City, New Jersey 08405
Phone: 609-485-5259
Fax: 609-485-5019
Email: John.Frederick@faa.gov

Section 1: The Evolving T&E Environment and Culture

Any level of research into the test and evaluation (T&E) of complex systems will show that T&E is a daunting task. Scoping and planning a T&E program in advance of development could be considered an art of analysis and prediction. Conducting and reporting a T&E program could be considered an art of balance and management of the test sufficiency, cost, schedule, and risks. A test manager or director who has been involved with complex system programs will attest that the test program is in most cases, undercut or overwhelmed by system development efforts or system integration complications. Additionally, to stay within cost and schedule, programs make adjustments to operational procedures, training, deployment schedules, and to the planned capabilities for implementation.

As the Federal Aviation Administration (FAA) makes dramatic strides to improve capacity, efficiency, security, and safety through the implementation of Next Generation Air Transportation System (NextGen) capabilities, the test program will depend more that ever on the ability to provide effective planning, test measures, test conduct, and reporting. NextGen will be a system of systems with new operational concepts that expand the bounds of the current National Airspace System (NAS) into the cockpit. With these new challenges, the art of T&E

will be to perform in accordance with all required standards and best practices without being undercut or overwhelmed.

With NextGen, the need for more effective, efficient, and comprehensive T&E services will be greater. FAA's focus will shift from air traffic control to air traffic management. NextGen will move towards using precise navigation to allow for greater flexibility with airports and permit more precise departure and arrival paths. To implement these NextGen capabilities with tighter safety thresholds, programs will depend on T&E services that provide thorough analysis and comprehensive test coverage with decision support reporting methods. The NextGen environment will culturally change the FAA's service domain functional view. Capabilities implemented through service domains such as En Route, Terminal, Communications, Surveillance, Navigation, and Weather will share common information and be more integrated into the NAS rather than operate as separate and distinct services that operate autonomously. With NextGen, the NAS will operate as a system of systems and use available communication paths to be fully network enabled to access and exchange the right information at the right time. Since NextGen capabilities will be so closely tied together and share common data, T&E managers and test practitioners will be required to have a greater knowledge of the interactions between systems across the NAS.

Increasingly, sound and practical T&E practices are viewed as a critical link to quality in the technology industry. Therefore, T&E involvement is needed in all major phases of the system or product lifecycle. In government acquisitions, the T&E focus is expanding from rigid requirements based testing to include flexible, capability-based testing. Testing and reporting on operational capabilities provide decision makers with more relevant information on the state of the system or product. Specifications and detailed requirements support product design and development, but have been found to be sometimes inadequate as a sole means for verification and validation (V&V) of system or product. Combining requirements and capability-based testing that follow standardized V&V processes which are continuously improved upon, is an optimum approach that is consistent with industry best practices and quality T&E.

As budgets get tighter, government agencies need to be more cost-effective and efficient in their acquisitions without sacrificing quality. Therefore, T&E process improvements are geared towards finding, reporting, and resolving defects and suitability issues as early in the product lifecycle as possible, where the costs, schedule and technical risks have the least impact. Program costs and impacts increase with each phase of the product development and lifecycle. In today's economic environment, it is in the program's best interest to optimize their program by taking both minor steps and major steps that contribute towards detecting, reporting, and resolving defects in the earlier phases of the product lifecycle.

This white paper will explore the V&V of complex systems in the FAA by discussing the following topics in the following three sections:

- A Study on the Systems Thinking approach for V&V.
- Benefiting from V&V throughout the product lifecycle.
- Capability based T&E and the effective use of measures.

Section 2: A Study on the Systems Thinking Approach for V&V

According to legendary systems thinker Peter Senge (2006), a shared vision is capable of uniting a company or agency of distinct divisions. NextGen is that shared vision for FAA and its partners. Therefore, the traditional practice of developing a system and then handing it "over the wall" to the test group is no longer appropriate. As Senge would argue, a systems thinking approach is needed. Supporting this claim, Carlock, Decker, and Fenton (1999) analyzed the National Airspace System (NAS). They concluded that the NAS is a system of systems (SoS) and, therefore, requires a multi-level process methodology. This multi-level process includes the enterprise level, system level, and implementation level. Considering each of these levels together constitutes a systems thinking approach. Yet, what is a system of systems and how does it apply to NextGen? Furthermore, what is a systems thinking approach?

Maier (1998) claims a SoS is different than traditional systems. His definition of a SoS is a collection of component systems with two additional properties. Each component system must have its own purpose independent of the other systems, and the component systems must maintain their independence. Lane and Boehm (2008) describe a SoS as the integration of new and existing independent systems. Boardman et al. (2006) reviewed 41 papers related to SoS and extracted commonalities from the definitions. Similar to Maier, they divide the various traits into five descriptive characteristics, which they call the system of systems characteristics. Boardman and Sauser (2008) claim these characteristics define a SoS as well as differentiate it from other systems. These "essential characteristics" are given the names autonomy, belonging, connectivity, diversity, and emergence. The meanings of these characteristics are presented at a high level, but they can be applied to NextGen to determine its status as a SoS. Autonomy exists in the NextGen SoS since each interface and system is designed for a purpose and will be expected at some level to discretely fulfill its intended function or service. The collection of systems will be dynamically connected and belong to the SoS by contributing to the overall goal of NextGen. Although each system has its individual goal, they also contribute to the overall goal. Obviously these systems will be diverse since their functions range from tracking aircraft to providing a common weather picture. The final characteristic of emergence is evident in the final product. By combining multiple independent systems, the Next Generation Air Transportation System emerges with the ability to meet future demands and increased capacity to avoid gridlock in the sky and at airports. Hence, NextGen is clearly a system of systems according to the stated characteristics. As confirmed by Carlock, Decker, and Fenton, a SoS requires more than traditional systems engineering, it requires a systems-thinking approach.

Systems thinking is a way of approaching a problem or situation, and therefore there is no single methodology of systems thinking. Nonetheless, there is varying guidance on how to use systems thinking. Senge attempts to identify systems thinking for management purposes. Frank (2000) makes an attempt to apply it to engineering by listing 30 laws of systems thinking. Boardman and Sauser present system thinking concepts that can apply to management or engineering. All authors argue that the whole system must be considered as well as the interaction between elements, which is different than traditional reductionism of science. As will be presented in later sections, systems thinking is implemented for the T&E program via the V&V Lifecycle approach.

Part of Senge's systems-thinking description is the relation of cause and effect (Senge 2006). In linear thinking, an effect is thought to follow a cause. Although this linear flow of events is common, it is not mandatory. In many situations the effect of a particular cause takes time to manifest. An example of an application of this philosophy is untestable requirements, such as the user-provided requirement "the system shall be user friendly." This requirement is desirable yet vague. A user-friendly system might mean something different to the tech-savvy computer specialist than to a computer-phobic manager. Therefore the requirement of user-friendly may not have an impact until the system is fielded, which may be years after the requirement was identified. Hence, a system-thinking approach in this case provides a means to identify and mitigate the cause before the undesirable effect can occur. To implement this tactic for NextGen, T&E services must get involved as early as possible for all projects.

Several of Frank's systems thinking-based laws are directly relevant from a T&E perspective. His 7th law directs the systems thinker to always consider testability as well as other T&E concepts. The 12th law addresses testing by directing the systems thinker to look for patterns and repeated structures when problems arise. Complementing this statement is law 22, which indicates that no one person or group can understand fully a complex system (Frank 2000). Taken together, these two laws indicate a need for effective reporting procedures so that the appropriate people have the right information to see any patterns. Including testers in all phases of development allows an opportunity to catch T&E problems, identify patterns, and contribute to the comprehension of the complex system.

Finally, Boardman and Sauser (2008) argue the need to consider paradoxes or to think paradoxically. For example, to save money on a project, more money needs to be spent up front. When a paradox is encountered on a project, an opportunity to try a different tactic may avoid the problem while presenting unforeseen opportunities. Paradoxical thinking includes using T&E expertise in areas not traditionally involved with T&E. That is not to say that T&E can perform non-T&E services, but rather considering T&E in non-T&E areas may provide more effective solutions that would otherwise go unnoticed. Furthermore, this line of reasoning helps field personnel and system developers play a larger part in the testing and evaluating of future systems. This type of paradoxical thinking was used in developing the V&V philosophy, and it needs to continue as the T&E programs approach a sound T&E culture.

<u>Section 3:</u> Benefiting from V&V Throughout the Product Lifecycle

To have an effective and efficient T&E program, an organization must have a sound T&E culture that is embraced by all program stakeholders and is supported by organizational policies. This sound T&E culture, seeks to verify and validate work products continuously from the beginning of the system or product lifecycle until it goes into operations. Organizations that embrace this culture are using T&E as a powerful decision making tool that supports development decisions and clearly defines system limitations and capabilities for deployment decisions. This will provide a value added benefit by adopting quality T&E practices to the fullest extent. During every major stage of a system or product lifecycle, critical decisions are made on the major programmatic elements such as:

- Operational mission
- Operational concepts
- Available technologies
- Solution sets
- Requirements
- Design
- System performance
- Implementation plans
- Criteria for system deployment.

A program needs credible and useful information that will verify and validate these programmatic elements and effectively supports the decision making process. Without this information, programmatic decisions could be flawed. Incremental V&V should be used thorough the lifecycle to avoid detrimental influences and to shield program decision processes from being impacted by the many potential flawed sources that can plague programs, as indicated in figure 1.

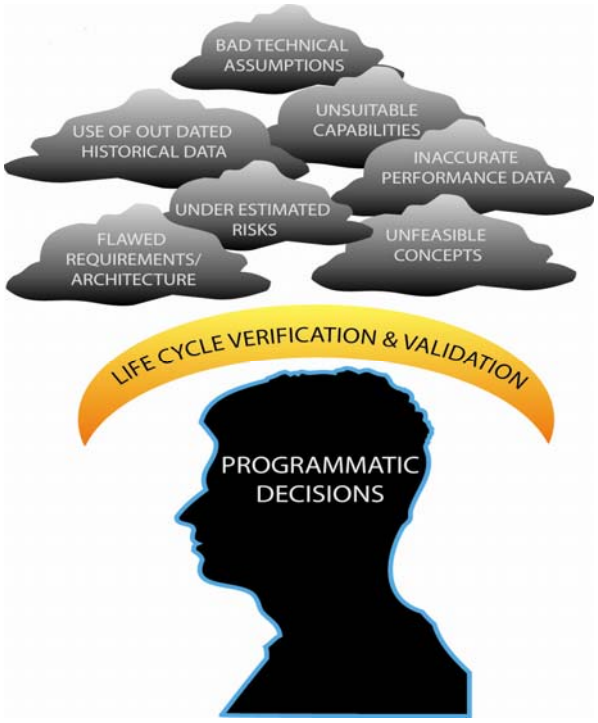

Figure 1: V&V can shield the program from flawed decisions

Best practices defined in the CMMI® models call for V&V to be done on program work products, product components, and integrated products incrementally throughout the lifecycle. The practices can be performed on items such as concept/requirements development products, proposed technical solutions, contract products, specifications, design products, developed products, integrated products, and modified products.

Using this lifecycle, V&V philosophy can move the FAA into a position that will progress programs efficiently and effectively from the concept phase into the implementation and deployment of NextGen capabilities. FAA is implementing common T&E standards and practices to support the V&V of NextGen capabilities. NextGen programs will implement capabilities through new acquisitions and by modifying or incrementally enhancing existing operational systems. This movement to a lifecycle T&E continuum will ensure the success and efficiency of each program and will facilitate the delivery of NextGen capabilities that are critical the future of a safe and capable Air Transportation System. By promoting and establishing policies that reinforce and define the role of T&E to verify and validate work products during each stage of a lifecycle, the program can achieve continuous evaluation as the program progresses rather than limited evaluations during traditional formal test periods. This will result in the following program benefits:

- Greater probability in meeting operational requirements
- Removal of defects from the product early in the acquisition
- Greater probability in meeting user and mission needs
- Improved product and process quality
- More efficient and effective transition into operations
- Improved productivity and performance

Planning for a test program should also address strategies that include a broader scope than just the T&E activities that occur during the solution implementation phase of a program. Planning documents, such as the Test and Evaluation Master Plan (TEMP), should take the whole product lifecycle into account and evolve with the product. The TEMP can continue to be updated and used by the program post deployment to document the T&E plans for transitioning a system into operational service, system maintenance, for testing of in-service modifications, and for planning of subsequent acquisitions.

V&V best practices for effective T&E reporting start in the test planning phase. The test approach, test resources, test schedule, test objectives, test criteria, and test conduct must be planned with reporting in mind. Effective T&E reporting:

- Is provided frequently and on time
- Is presented simply so that the results can easily be extrapolated for decision making
- Provides high level summaries backed up by detailed T&E results and analysis
- Provides unbiased recommendations and assessments
- Is used to track progress and to establish milestone completions
- Characterizes performance levels and provide status on specified requirements or parameters in accordance with planned objectives
- Identifies necessary future testing and recommends re-test and regression testing

If a government decision authority is making a major decision based on T&E results, it is most likely an implementation, contractual, deployment, production, or in-service milestone decision. These are important decisions and carry cost, schedule, political, safety, and/or security risks. Programs implementing NextGen will certainly carry all of these risks because the concepts are so new and the systems are so complex. T&E reporting should not be taken lightly, the program's success hinges on it. The decision points that T&E data must support should be clearly defined in the early plans. Those plans should also define the criteria and reporting mechanisms that T&E will use to support the decisions. These criteria and reporting mechanisms will be integrated into the test program as a reporting system that will be active throughout the entire test program. Program managers and executive level decision makers will need to use the reporting system as a "health check" of the program and to support risk based decision making. The NextGen Program Managers will need to learn about these reporting systems, use them effectively, and improve upon them to ensure that they meet their organizational performance goals defined in the FAA Flight Plan.

Section 3: Capability Based T&E and the Effective Use of Measures

Testing of complex systems is an iterative process that is required throughout all phases of development (from initial design through deployment). A common T&E question for complex systems is "How much testing is enough and when can you end the testing phase?". This is not easy to answer. The well known Dutch computer scientist and physicist, Dr. Edsger Wybe Dijkstra, is credited with the quote, "Program testing can be used to show the presence of bugs, but never to show their absence."[1] Testing is decomposed into the two components: verification testing and validation testing (Paglione 2006).

- Verification testing is the testing that ensures that the product meets the requirements specified by the customer. Verification testing is usually characterized by the question, "Are we building the product right?"

- Validation testing, on the other hand, is characterized by the question, "Are we building the right product?" Validation testing is the testing that ensures that the product fulfills its intended use when placed in its intended environment. This testing often includes performing systematic evaluations of the product in increasingly complex cases under of real-world conditions.

Both verification testing and validation testing assess the correctness and completeness of a system or product; but are also concerned with different evaluation criteria. In other words, verification testing establishes whether a system performs in accordance with specification, while validation testing is defined in terms of evaluating the system against the operational mission, baseline, or requirements.

[1]Dr. Edsger Wybe Dijkstra (1930-2002) was a Dutch computer scientist and physicist.

Another important aspect of testing is its effect on project cost. Barry Boehm in his book titled "Software Engineering Economics" evaluated a number of software projects and estimated how the relative cost of fixing an error significantly increases as the project progresses in phase (Boehm 1981). Figure 2 summarizes the results of this study. Errors are about 15-50 times more costly to fix when found in the testing phases compared to errors that are detected when the requirements are defined. However, after the system has been deployed, it is roughly a hundred times more costly to fix an error. This heavily cited reference may even be underestimating the cost.

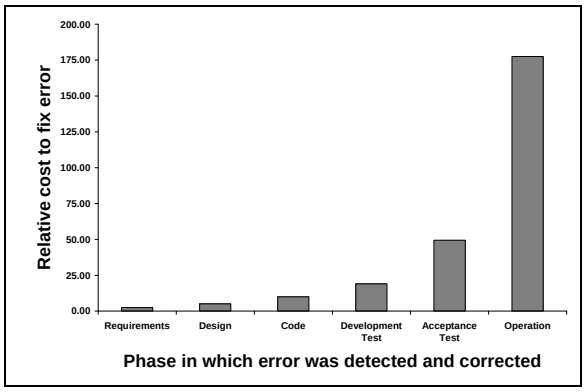

Figure 2. Increase in cost-to-fix through the software development cycle[2]

In addition, when dealing with safety critical systems such as our nation's air traffic control systems, safety is definitely the bigger issue. In a recent article in the Wall Street Journal, which cited examples of software errors found in numerous deployed onboard computer systems, Michaels and Pasztor state that software errors, "while extremely rare, are emerging as a top safety challenge in the air," (Michaels 2006). Although this article focuses on aircraft onboard computer software, the authors feel that the safety challenge can also be appropriately directed toward the ground-based air traffic control systems.

The en route air traffic control computer system is considered the heart, brain, and backbone of the National Airspace System (NAS). En Route Modernization Program (ERAM) replaces the software for the Host Computer System (HCS) and its backup. The HCS software and hardware are based on a 1960's operational concept. ERAM is a major program that will completely replace the software and hardware architecture that has been safely supporting air traffic control services for over 40 years. The computer system processes flight radar data, provides communications support, and generates display data to air traffic controllers. FAA's ERAM Test Group, located at the FAA William J. Hughes Technical Center, formed the Automation Metrics Test Working Group (AMTWG) in 2004 to focus on reducing the cost of program risks by measuring and reporting on ERAM system performance from a different perspective. This is a cross-functional team consists of members from a half dozen organizations located at the WJHTC. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. The targeted system functions are the Surveillance Data Processing (SDP), the Flight Data Processing (FDP), the Conflict Probe Tool (CPT), and the Display System (DS) modules. The focus of the AMTWG is to go beyond requirement-based verification testing by using metrics that directly link to air traffic control services provided by NAS. Whenever appropriate, metrics were designed to measure not only the performance of ERAM, but also to measure the performance of the existing Host Computer System (HCS), which will allow comparison of the functionality in ERAM to the same functionality in the legacy HCS. For logistical purposes, the AMTWG categorized the metrics based on the targeted ERAM subsystems.

[2]Adapted from *Software Engineering Economics* by Barry W. Boehm.[6]

The AWTWG focused on testing ERAM capabilities linked to user services and stressed validation over verification. A capability is the ability to perform an action. Some refer to capabilities as a set of functionalities within a product that are verified. In contrast, the capability is defined in terms of a collection of requirements. A requirement is a "singular documented need of what a particular product or service should be or do," (Wikipedia 2007). Each requirement defines a piece of functionality, attribute, characteristic, or quality of a system under which the capability is performed. AMTWG emphasis on developing metrics to compare the capabilities that ERAM replaced in the legacy system serves to lower FAA risk by catching problems as early as possible.

The AMTWG divided the project into three key phases: a metrics identification phase, an implementation-planning phase, and a data collection and analysis phase.

- In the metrics identification phase, the AMTWG generated a list of approximately one hundred metrics that map to the services and capabilities found in the "Blueprint for the National Airspace System Modernization 2002 Update," (FAA 2002). These initial metrics were published in the "ERAM Automation Metrics: Progress Report of the Automation Metrics Test Working Group," (FAA AMTWG 2004).

- During the implementation-planning phase, initial metrics were prioritized for more detailed refinement, as documented in the "En Route Automation Modernization: Automation Metrics and Preliminary Test Implementation Plan," (FAA AMTWG 2005). It lists the metrics, gives the rational for selecting them, and provides a high level description on how the highest priority metrics will be measured. The paper provides each metric's traceability to the basic controller decisions, ERAM Critical Operational Issues (COIs), and the development contractor's Technical Performance Measurements (TPMs). The categories of high priority metrics are: (1) SDP radar tracking metrics, (2) SDP tactical alert processing metrics, (3) FDP flight plan route expansion metrics, (4) FDP aircraft trajectory generation metrics, (5) CPT strategic aircraft-to-aircraft conflict prediction metrics, (6) CPT aircraft-to-airspace conflict prediction metrics, (7) additional system level metrics, and (8) DS human interface metrics.

- The final phase is the data collection and analysis phase, during which the AMTWG documented the validation, further refinement, and application of these metrics on the current legacy systems in a series of Metric Reports. The AMTWG delivered many metric reports covering each of the ERAM modules: SDP, FDP, CPT, and DS. The AMTWG published these reports in multiple drops to provide the ERAM Test Team with timely information. The drops coincide with the approaches used to implement the metrics. This phase will continue until ERAM is fully deployed and testing phase complete. The metrics defined by the AMTWG are either absolute or comparative in nature, with the comparative metrics being applied first to the current NAS automation systems and then later to ERAM. Metrics are also used iteratively, with frequent output. The metrics were applied first to the legacy NAS to flush out and establish their credibility, then repeated and compared to the ERAM replacement subsystem as they become available. The following subsections, A and B, provide examples of these activities.

A. Example of Metrics Study for ERAM Surveillance Data Processing:
One sample application of the test metrics approach was the implementation of metrics for evaluating the ERAM surveillance tracking algorithm. At the highest level, the FAA's air traffic control system relies directly on aircraft locations provided by the long range en route surveillance radars. The accuracy of the radars is an important factor in determining the overall performance of the system. To support the planned modernization of the air traffic control system, a metrics study was conducted to measure the accuracy of the radar tracking function of the legacy HCS. This was first done by comparing aircraft radar tracks produced by the existing system with the tracks for the same aircraft produced by the Global Positioning Satellite System (GPS) position reports. It was assumed that the GPS data was the ground truth. The GPS data was available from the FAA's Reduced Vertical Separation Minimum (RVSM) Certification Program. The Host Air Traffic Management Data Distribution System (HADDS) at each Air Route Traffic Control Center (ARTCC) captures the radar track data. These data are then archived at the William J. Hughes Technical Center (WJHTC) for a period of time. Radar tracks for 265 flights were compared to their GPS "tracks." Three distance metrics were used: horizontal track error and its two components: cross track error and along track error. A total of 54,170 pairs of position reports were compared. The distributions of the errors were plotted, and basic descriptive statistics were determined.

The average horizontal error was 0.69 nautical miles (nm), the root mean square value of the horizontal error was 0.78 nm, the average cross track error was 0.12, and the average along track error was 0.67 nm (Ryan 2005). The complete results are summarized in table 1.

Table 1: HCS Tracker Results from (Ryan 2006)

		Horizontal Error		Cross Track Error		Along Track Error	
Type	Sample Size	Mean	RMS	Mean	RMS	Mean	RMS
Signed	54170	0.69	0.78	0.00	0.16	-0.67	0.77
Unsigned				0.12		0.67	

Once the GPS analysis was conducted, a second study repeated this approach but used a simulation environment that could be repeated later on the analogous ERAM tracking functions. The metrics study began by recording approximately four hours of air traffic data from Washington DC. ARTCC (ZDC) on March 17, 2005. The Automation Metrics Test Working Group supplied this data to the Integration & Interoperability Facility that ran a simulation using the FAA's Graphical Simulation Generation Tool (GSGT) simulator. The simulation produced the HCS track reports and GSGT positions. The GSGT positions were considered the actual path the aircraft flew in which the HCS track positions were measured against. Data was processed to compare GSGT positions to the HCS track reports. Four error metrics were applied including horizontal error, its orthogonal components cross and along track error, and altitude error. For this study, the mean horizontal error was 0.85 nautical miles. The cross track error distribution is symmetrical about zero nautical miles and a root mean square value of 0.14 nautical miles. However, the along track error distribution is strongly skewed in the negative direction with an average error of -0.83 nautical miles. This represents an uncompensated time error, confirming the previous result using GPS aircraft positions (Paglione 2005).

The methodology then required a repetition of the experiment comparing ERAM functions to the legacy experiment results. A repeat run of the same scenario as described above was adapted for

running into ERAM. The cross track error is reasonably comparable to the HCS.[3] The experiment of 1100 flights from the same ZDC data produced an average unsigned cross track error of 0.04 nautical miles and a root mean square error of 0.0025 nautical miles. The cross track error in ERAM is a third the size of the HCS results. As illustrated in the histograms in figure 3 and 4, the ERAM tracker clearly improved upon the tracking problem in the lateral or cross track dimension, (Ryan 2005) and (Paglione 2008).

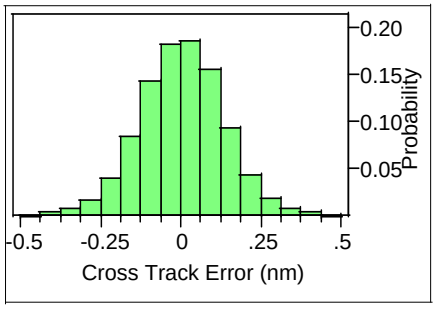

Figure 3: HCS Cross Track Error (Ryan 2005)

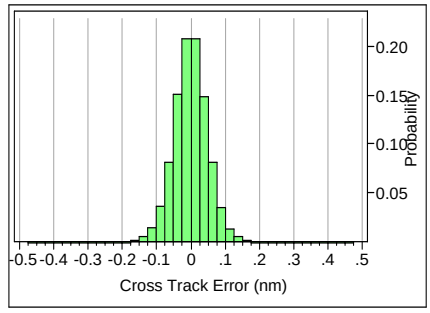

Figure 4: ERAM Cross Track Error (Paglione 2008)

B. Example of Metrics Study for ERAM Flight Data Processing – Converted Route Logic:
A key function of most flight data processing systems, and certainly for the ERAM system, is the processing of the flight plan. The flight plan represents the current air traffic control's cleared 2-dimensional path of the aircraft's flight path. It represents the best intent information available to the automation and is used for many critical air traffic functions from hand-off coordination and flight strip printing to aircraft trajectory prediction. The flight plan is converted to a 2-dimensional set of positions, typically referred to as the converted route. Thus, a metric study was performed first on the legacy automation and then repeated on the ERAM replacement to measure this critical function. The primary metric is defined as the lateral deviation or distance from the current aircraft position to the converted route. This is illustrated in figure 5 below.

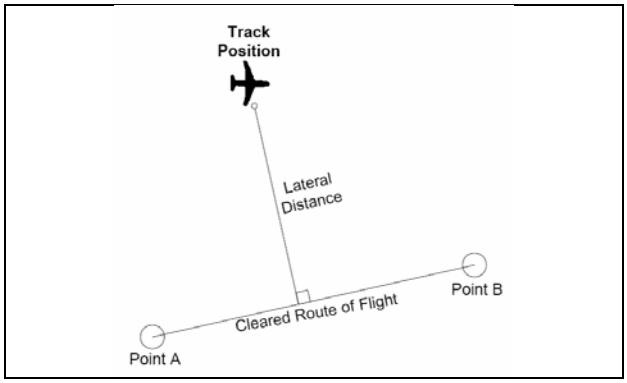

Figure 5: Lateral Distance, adapted from (Baldwin 2005)

In the initial study performed in (Baldwin 2005), data from ZDC were collected and applied to the legacy systems of URET and the HCS, the strategy was applied to determine which system

[3] Due to a problem with one of the interfaces still being repaired at the time of this publication, only the cross track error can be compared in this paper as presented in (Paglione 2008).

had superior performance. The test results indicate that URET performance exceeds the HCS performance with better than a 99.9% probability. The study was later repeated and presented in (Paglione 2008) using ERAM results. The average lateral deviations for the HCS, URET, and ERAM systems were 1.7, 0.9, and 0.9 nautical miles, respectively. The distributions of these deviations are illustrated in figures 6, 7, and 8 for each of the systems HCS, URET, and ERAM, respectively. The HCS is significantly skewed to the right consistent with its mean value, while the URET and ERAM distributions are practically identical. This is as expected since the logic for both URET and ERAM flight plan route conversion was the same. Thus, the test metrics analysis confirmed that the ERAM system would provide enhanced performance over the legacy HCS system and matched the URET system's accuracy.

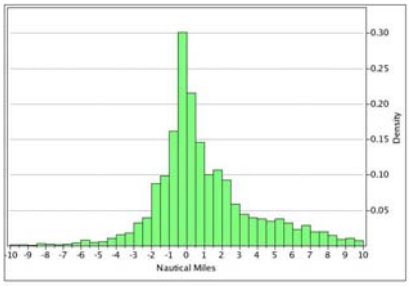

Figure 6: HCS Lateral Distances[4]

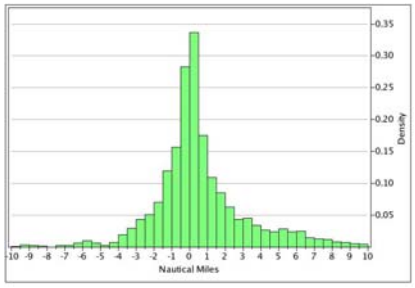

Figure 7: URET Lateral Distances[4]

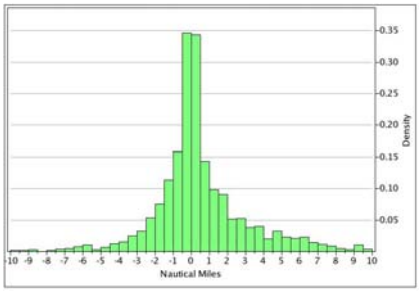

Figure 8: ERAM Lateral Distances[4]

Benefits of Metrics Approach:

Overall, this activity provided several significant benefits to the ERAM program. It provided useful data points for the test program that helped develop test cases and measurements to supplement and enhance the requirements based verification. It served to reduce risk for the entire program by evaluating key subsystems and functions for effectiveness in providing NAS air traffic control services. The metrics provided the ERAM Program Office with supporting data on ERAM benefits. Though these metric results were made available and used extensively by the ERAM prime contractor, it was not integrated into the contract or the Development Test (DT) formal conduct. If integrated into the program's prime contractor and vendor contracts, these test methods could greatly buy down risks to FAA programs.

The activity allowed key ERAM Test Team personnel to receive in-depth experience on ERAM subsystems and similar functions in the legacy automation. This experience increased their effectiveness in reviewing contractor test plans and procedures. The use of a cross organizational metrics team with varied skilled sets and expertise had proven to be very constructive as well. It allowed test personnel to bring an increase of FAA resources to focus on complex problems within their program. Finally, the tools, metrics, and air traffic scenarios were incorporated, where appropriate, into the formal ERAM Test Program. These simulations, test cases, and measures can be continuously updated and used throughout the lifecycle of ERAM to test enhancements and re-baseline the system performance prior to transitioning into operations. With the lessons learned on ERAM, the improved V&V techniques of the Metrics-Based Approach can certainly be further standardized and be applied with even greater success to other FAA programs.

[4] Figures 7, 8, and 9 adapted from (Paglione, 2008).

Section 4: Summary

The art of T&E is a systems engineering discipline that is evolving to be an active process throughout the entire lifecycle. This evolution will put the T&E services into a critical role of a continuum of supporting activities for the programs. An effective reporting system that is used often at critical points in the program will bring to bear the benefits of standardized T&E processes and quality T&E practices at the program management levels. For the Next Generation Air Transportation Systems, systems thinking methodologies will be an important means to understand the essential characteristics of a system and derive usable test cases. Furthermore, Next Generation Air Transportation System T&E practitioners will need to proactively refine methodologies and metrics that support the derivation of testable parameters that have validated measures and go beyond strictly requirements based testing. Applying the Metrics-Based Approach can provide very useful measures of the performance on key functions needed at critical decision points that would not other wise be available. These standard V&V methods all strive to create greater efficiencies in programs by employing T&E practices that reveal the true state of the system in a manner and order that is most useful to planners and decision makers.

References:

Baldwin, W. C. 2005. C*omparison of converted route processing by existing versus future en route automation. DOT/FAA/CT-TN05/29*, Federal Aviation Administration, William J. Hughes Technical Center, August.

Baldwin, W. C., M. M. Paglione, J. E. Rhodes, S. R. Larimer. 2007. A*nalysis of handoffs for future en route automation. DOT/FAA/TC-TN07/15*, Federal Aviation Administration, William J. Hughes Technical Center, February 28.

Boehm, Barry W. 1981. S*oftware engineering economics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Boardman, John, and Brian Sauser. 2008. S*ystems thinking: Coping with 21st century problems*. Boca Raton: CRC Press.

Boardman, John, Spiros Pallas, Mike Pennotti, Brian Sauser, and Dinesh Verma. 2006. Report on System of Systems Engineering, Final Report for the Office of Secretary of Defense, Stevens Institute of Technology, Hoboken, NJ.

Carlock, Paul G., Steven C. Decker, and Robert E. Fenton. 1999. Agency-level systems engineering for "systems of systems.*" Systems and Information Technology Review Journal*. (Spring/Summer): 99–110.

Frank, Moti. 2000. Engineering systems thinking and systems thinking. S*ystems Engineering* 3 no. 3:163-168.

Lane, Jo Ann and Barry Boehm. 2008. System of systems lead system integrators: Where do they spend their time and what makes them more or less efficient? *Systems Engineering* 11 no. 1 (Spring): 81-91.

Maier, Mark W. 1998. Architecting principles for systems-of-systems. S*ystems Engineering* 1 no 4:267 - 284.

Michaels, Daniel, and Andy Pasztor. 2006. Incidents prompt new scrutiny of airplane software glitches: As programs grow complex, bugs are hard to detect; A jet's roller-coaster ride; Teaching Pilots to Get Control. T*he Wall Street Journal* (May 30): A1.

Paglione, M. 2008. En Route Automation Modernization Testing Program: Minimum operational performance standards from test metrics activity. Internal Presentation to the Federal Aviation Administration, William J. Hughes Technical Center, July 11.

Paglione, M., R. Oaks, H. Ryan. 2006. Metrics-based approach for evaluating air traffic control automation of the future. Paper presented at the 51st Air Traffic Control Association Annual Conference Proceedings, October 29.

Paglione, M., W. C. Baldwin, S. Putney. 2005. H*ost Radar Tracking Simulation and Performance Analysis*. *DOT/FAA/CT-TN05/31*, Federal Aviation Administration, William J. Hughes Technical Center, August.

Ryan, H. And M. Paglione. 2005. C*omparison of Host Radar Tracks to Aircraft Positions from the Global Positioning Satellite System*. *DOT/FAA/TC-TN05/30*, Federal Aviation Administration, William J. Hughes Technical Center, August.

Senge, Peter M. 1990. T*he fifth discipline. N*ew York: Doubleday.

U.S. Federal Aviation Administration. 2002. B*lueprint for NAS Modernization: 2002 Update*. http://www.faa.gov/nasarchitecture/Blueprint2002.htm [accessed October, 2002].

U.S. Federal Aviation Administration. Automation Metrics Test Working Group. 2004. *ERAM Automation Metrics: Progress Report of the Automation Metrics Test Working Group* Prepared by the ERAM & ECG Group (ACB-550) and the Simulation and Analysis Group (ACB-330). Atlantic City International Airport, NJ: William J. Hughes Technical Center, June 16.

U.S. Federal Aviation Administration. Automation Metrics Test Working Group. 2005.*En Route Automation Modernization: Automation Metrics and Preliminary Test Implementation Plan, Version 2.6*. Prepared by the Simulation and Analysis Group (ACB-330). Atlantic City International Airport, NJ: William J. Hughes Technical Center, June 16.

Wikipedia contributors. Requirement. *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/wiki/Requirement (accessed October 2007).