Methods for Examining Possible Effects of En Route Automation Modernization (ERAM) on Controller Performance

Kenneth R. Allendoerfer, NAS Human Factors Group

Ben Willems, NAS Human Factors Group

Carolina Zingale, Ph.D., NAS Human Factors Group

Shantanu Pai, L-3 Communications Titan Corporation

May 2006

DOT/FAA/TC-TN06/14

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-TN06/14		

4. Title and Subtitle Methods for Examining Possible Effects of En Route Automation Modernization (ERAM) on Controller Performance	**5. Report Date** May 2006
	6. Performing Organization Code ATO-P
7. Author(s) Kenneth Allendoerfer, NAS Human Factors Group, ATO-P; Ben Willems, NAS Human Factors Group, ATO-P; Carolina Zingale, Ph.D., NAS Human Factors Group, ATO-P; Shantanu Pai, L-3 Communications Titan Corporation	**8. Performing Organization Report No.** DOT/FAA/TC-TN06/14
9. Performing Organization Name and Address Federal Aviation Administration NAS Human Factors Group, ATO-P William J. Hughes Technical Center, Bldg. 28 Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)**
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Research and Engineering Division 800 Independence Ave., S.W. Washington, DC 20591	**13. Type of Report and Period Covered** Technical Note
	14. Sponsoring Agency Code Human Factors Research and Engineering Division, ATO-P

15. Supplementary Notes

16. Abstract

The Federal Aviation Administration is developing the En Route Automation Modernization (ERAM) system to replace the legacy en route air traffic control automation system consisting of the Host Computer System, the Display System Replacement (DSR), and the User Request Evaluation Tool (URET). This technical note provides an analysis of major areas where new ERAM features may affect how controllers do their jobs. We describe test methodologies for examining these effects and corresponding metrics. Our analysis examines the following categories of ERAM changes: (a) backup and redundancy capabilities; (b) Areas of Interest (AOIs) that increase flight data capabilities in ERAM; (c) differences between the legacy system and ERAM user interfaces (UIs); (d) the ERAM tracker; and (e) safety alerts. We also discuss two recommended test activities: a usage characteristics assessment and human-in-the-loop baseline simulations.

17. Key Words	18. Distribution Statement
Air Traffic Control En route Human Factors	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161 A copy is retained for reference by the William J. Hughes Technical Center Library.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	44	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Acknowledgements

We thank Dr. D. Michael McAnulty of the Federal Aviation Administration (FAA) National Airspace System Human Factors Group and Dr. Pamela Della Rocco, Air Traffic Organization En Route and Oceanic Service (ATO-E), who helped develop the technical approach followed in this analysis. We also thank Randy Phillips, Air Traffic Control Supervisor from Cleveland Air Route Traffic Control Center, who again provided us with his air traffic control subject-matter expertise during this analysis. In addition, we thank Scott Ginsburg from ATO-E for his guidance and advice during the execution of this analysis and Dino Piccione from the FAA Human Factors Research and Engineering Division for his helpful comments.

Finally, we thank the FAA Human Factors Research and Engineering Division for sponsoring the project and the En Route Automation Modernization (ERAM) Program Office and ERAM Test Group at the William J. Hughes Technical Center for their assistance and enthusiasm for this work.

Table of Contents

List of Illustrations

Figures Page

Tables Page

Executive Summary

The Federal Aviation Administration (FAA) is developing the En Route Automation Modernization (ERAM) system to replace the legacy en route air traffic control automation system consisting of the Host Computer System, the Display System Replacement (DSR), and the User Request Evaluation Tool (URET). An ERAM Critical Operational Issue (COI) requires that the new system support en route operations with at least the same effectiveness as the legacy system. To allow the FAA to evaluate the new system against this COI, we must measure the effectiveness of the legacy system, including controllers' usage of it. This technical note provides an analysis of major areas where ERAM may affect controller performance. We describe metrics and test methodologies for examining these effects.

Our analysis examines five categories of ERAM changes. First, we examine new backup and redundancy capabilities of ERAM. Second, we examine ERAM Areas of Interest (AOIs) functionality. In ERAM, each facility will have access to flight data for aircraft falling in a designated AOI around the facility. This expansion of the flight database is intended to allow controllers to work traffic near boundaries more effectively. Third, we examine differences between the legacy system and ERAM user interfaces, including changes to the existing lists and toolbars. Fourth, we briefly discuss changes to the ERAM tracker. Because we do not fully understand the effects of the new tracker on controller performance, we do not propose a detailed evaluation method in this case. We do, however, layout an overall approach for how we can examine the effects of the new tracker changes. Fifth, we discuss safety alerts, such as conflict alerts, and propose a method for examining controller expectations for when and if alerts should activate in different situations.

Finally, we discuss how we can incorporate these tests and evaluations into two test activities as part of formal ERAM testing. The first activity, known as a Usage Characteristics Assessment (UCA), is a controlled laboratory study in which researchers collect detailed data on how controllers interact with the legacy system and ERAM. Metrics in the UCA include the time and number of actions needed to complete a command and the amount of attention required. We also propose including an examination of safety alerting algorithms in the UCA. The second group of activities, known as baseline simulations, is composed of realistic simulations in which controllers work traffic as best they can. In the simulations, we can collect broader measures of safety, efficiency, and workload.

1. Introduction

The Federal Aviation Administration (FAA) is developing the En Route Automation Modernization (ERAM) system to replace the legacy en route air traffic control (ATC) automation system consisting of the Host Computer System (HCS), the Display System Replacement (DSR), and the User Request Evaluation Tool (URET). En route controllers use the legacy system to control thousands of flights each day at 20 Air Route Traffic Control Centers (ARTCCs) in the continental United States. Lockheed Martin Corporation is the primary ERAM contractor.

The *Test and Evaluation Master Plan* for ERAM requires that the ERAM Test Program verify critical operational issues (COIs) (FAA, 2003). The first COI requires that ERAM support en route operations with at least the same effectiveness as the legacy system. To allow the FAA to evaluate ERAM against this COI, we must measure the effectiveness of the legacy system, including controllers' usage of it to provide benchmarks for comparison.

1.1 Purpose

ERAM makes numerous changes to the legacy system. In this report, we identify categories of ERAM changes that have potential to affect how controllers do their jobs. We describe general approaches for measuring those effects including data collection, methodologies, and metrics. This report is not a comprehensive examination of ERAM changes. Instead, it examines areas we believe have the most potential to measurably affect controllers' abilities to provide safe and efficient ATC services with a minimum of workload and human error.

ERAM changes are intended to benefit controllers by increasing their efficiency, improving their situation awareness and decision making, reducing their workload, or reducing the frequency or impact of mistakes. However, it is possible that some ERAM changes may have unintended negative effects on controllers. In this report, we describe methods and metrics for evaluating important ERAM changes from a human factors perspective to ensure that the changes provide their intended benefits and do not create significant unintended problems.

This technical note is one of several produced by the Automation Metrics Test Working Group (AMTWG) described in the *ERAM Automation Metrics and Preliminary Test Implementation Plan* (FAA, 2005).

1.2 Background

The FAA ERAM Test Group formed the AMTWG in 2004. The team supports ERAM developmental and operational testing by developing metrics that quantify the effectiveness of key system capabilities in ERAM. The targeted capabilities are the Surveillance Data Processing (SDP), Flight Data Processing (FDP), Conflict Probe Tool (CPT), and the Display System (DS) modules. The team designed the metrics to measure the effectiveness of the legacy system and to allow valid comparisons to ERAM.

The metrics development project contained several phases. First, during 2004, the AMTWG generated a list of approximately 100 metrics and mapped them to the services and capabilities found in the *Blueprint for the National Airspace System Modernization 2002 Update* (FAA, 2002). The AMTWG published the initial metrics in a progress report (FAA, 2004). Second, during 2005, the team prioritized the metrics for further refinement and created an implementation plan (FAA, 2005). The implementation plan lists the selected metrics, gives rationales for their selection, and describes how they identified high-priority metrics. The implementation plan traces each metric to basic controller decisions and tasks, COIs, and the ERAM contractor's technical performance measurements (TPMs). The categories of high priority metrics are

- SDP radar tracking,
- SDP tactical alert processing,
- FDP flight plan route expansion,
- FDP aircraft trajectory generation,
- CPT strategic aircraft-to-aircraft conflict prediction,
- CPT aircraft-to-airspace conflict prediction,
- additional system level metrics, and
- DS human factors and performance metrics.

In the final project phase, the AMTWG will further refine and apply the metrics to the legacy en route system. In 2005, the National Airspace System (NAS) Human Factors Group prepared three reports as members of the AMTWG. The first detailed the frequency of use of controller commands on the legacy system (Allendoerfer, Zingale, Pai, & Willems, 2006). That report described controller usage of the system during routine, "day-in-the-life" operations at Washington ARTCC and identified the most frequently used commands as being most important for testing activities. The second report described how controllers use the legacy system during special situations occurring in the NAS that have the potential to increase controller workload, reduce efficiency, or increase safety risk (Allendoerfer, Pai, & Zingale, 2006). That report identified methods and metrics to examine these effects. The current report discusses the possible effects of several important new ERAM capabilities on controller performance and proposes methods and metrics for examining these effects.

Taken together, the three reports provide guidance on evaluating the effectiveness of ERAM from different perspectives. The first described how the effectiveness of ERAM can be examined for routine operations by comparing the legacy system and ERAM on the most frequently used commands in usability assessments and simulations. The second described how the effectiveness of ERAM can be examined for special situations by simulating the situations on the legacy system and ERAM and comparing how well controllers respond using both systems. The third describes how the effectiveness of ERAM changes can be assessed by presenting controllers with situations where the changes may show differences and determining to what extent the differences between systems actually occur.

1.3 ERAM Changes

Our analysis examines five categories of ERAM changes. First, we examine new backup and redundancy capabilities of ERAM. In the legacy system, a failure of the HCS forces controllers to use a backup system with reduced capabilities, which reduces operational efficiency. ERAM, however, provides a fully redundant architecture in which the backup system is equivalent to the primary. This expansion of the backup capabilities is intended to reduce the negative effects of equipment outages.

Second, we examine the ERAM Areas of Interest (AOIs) functionality. In the legacy system, the flight database maintained at each ARTCC covers only the geographic area for which it is responsible. Flight data for aircraft outside the ARTCC are often unavailable to controllers, requiring manual coordination and transfer of flight plans. The additional coordination significantly increases controller workload, and controllers may take other operational actions instead to avoid it. In ERAM, however, each ARTCC will have access to flight data for aircraft falling in a designated AOI around the ARTCC. This expansion of the flight database is intended to allow controllers to work traffic near ARTCC boundaries more effectively.

Third, we examine major modifications to the user interface (UI) made in ERAM. These modifications include changes to the existing lists and toolbars to provide additional options and make the overall UI more consistent. Some ERAM UI changes expand controller capabilities by allowing controllers to issue multiple commands at once and customize their displays in new ways.

Fourth, we discuss changes resulting from the new ERAM tracker. Because we do not fully understand the effects of the new tracker on controller performance, we do not propose a detailed evaluation method in this case. We do discuss an overall approach for how we could examine the effects of the new tracker on controller performance.

Fifth, we discuss safety alerts. ERAM and the HCS use different algorithms for conflict and terrain alerts. Depending on the trajectories of the aircraft and the characteristics of the situation, when and if alerts will activate may change in ERAM. The intent of the change is to increase the overall accuracy of safety alerts, provide controllers with sufficient time to react appropriately, and reduce nuisance alerts. ERAM also provides airspace alerts, which is a new capability over the legacy system.

1.4 Assessing ERAM Changes

As discussed in our earlier reports, we proposed to conduct two assessments of the effectiveness of the ERAM DS. In the current report, we examine how those assessments should address ERAM changes.

1.4.1 Usage Characteristics Assessment

The first assessment, known as the Usage Characteristics Assessment (UCA), is a tightly controlled laboratory test of how controllers interact with the legacy system and ERAM. It collects data on the speed and accuracy of data entries, the number of steps needed to complete the entry, and the method controllers use to make entries to the system. Depending on resource constraints, the UCA can also examine aspects of controller attention and heads-down time using eye-tracking technology. The UCA methodology is designed to allow precise, detailed measurements. By design, the assessment environment is not realistic and controller participants do not attempt to control traffic. The UCA does not measure ATC safety or efficiency and measures only the aspects of workload associated with interacting with the system.

We discussed the functions we propose to examine during the UCA in our earlier reports (Allendoerfer, Zingale et al., 2006; Allendoerfer, Pai et al., 2006). In particular, we proposed to focus on the most frequent controller commands and the commands needed to deal with uncommon but operationally important situations. This set of commands represents the vast majority of controller interactions with either automation system. If ERAM is found to be at least as effective as the legacy system for this set of commands, we believe that it would be an effective replacement of the legacy system. In Section 3 of this report, we identify cases where ERAM makes changes to the legacy system that would be appropriate to test in the UCA.

1.4.2 Baseline Simulations

The second assessment consists of multiple human-in-the-loop (HITL) baseline simulations. These simulations differ from the UCA in several ways. Most important, the baseline simulations are intended to be as realistic as possible in terms of the tasks that controller participants are asked to complete. We will ask controllers to manage traffic as if they were at an actual facility. We will not make decisions for the controllers or interrupt them while they are working. Controllers will be allowed to change the traffic scenarios in real time by communicating with simulation pilots and issuing clearances.

HITL simulations allow measurement of controller performance in terms of safety, efficiency, and workload. Because of the dynamic, interactive nature of simulations, researchers have reduced control over what occurs during the session. Depending on controller decisions early in the session, the simulation can take many directions. The interactivity of the simulation usually requires that the data collected be broader than a UCA and measure overall controller performance rather than the performance of a particular system command.

The baseline simulations we propose will contain traffic scenarios with defined characteristics. Some scenarios will simulate a regular day in the life at a chosen ARTCC, based on traffic data collected from the field and refined to provide the desired mix of aircraft callsigns, aircraft types, routes, and volume. In past baseline simulations, we have used traffic scenarios that represented heavy but not overwhelming traffic volume (Galushka, Frederick, Mogford, & Krois, 1995; Allendoerfer & Galushka, 1999; Allendoerfer, Galushka, & Mogford, 2000). That is, we used traffic that was heavy and complex enough to present controllers with a challenging simulation but not so high as to deliberately increase the operational error rate. Routine scenarios neither simulate special events, such as outages or emergencies nor introduce flights with special characteristics except when needed to increase realism. For example, a routine scenario would not contain a large number of military flights unless the sector being simulated normally works a similar number of military flights.

We could examine some ERAM changes using routine scenarios, but only changes producing large effects on safety, efficiency, or workload are likely to produce observable effects. Subtle changes, especially those that occur infrequently, may be very difficult to detect. For this reason, we believe that ERAM changes should be evaluated using both routine scenarios and scenarios specifically designed to highlight the positive and negative effects of the changes. For example, the new ERAM backup capabilities are intended to reduce the impact of equipment outages, including the impact on controller workload and efficiency. A scenario targeted at this ERAM change would simulate an equipment outage and ask controllers to respond as they would in the field. In another example, the new ERAM macros are intended to reduce controllers' data entry workload. A scenario targeted at this ERAM change would simulate situations where heavy data entry is necessary, such as implementation of a major reroute, and ask controllers to respond as they would in the field. In a routine scenario, where data entry needs are lower, the potential benefits and drawbacks of the macros might be hidden because there is little reason for controllers to use the macros.

In Section 3, we examine a number of areas where ERAM makes changes to the legacy system. We selected ERAM changes that would be appropriate to test in the baseline simulations. We discuss what metrics could be collected to compare both systems. We also discuss considerations that must be accounted for when designing and executing a simulation examining these changes.

1.5 Previous Research

During the development process for the DSR, the NAS Human Factors Group conducted baseline simulations that compared the HCS using the original Plan View Display (PVD) (Galushka et al., 1995) against the HCS using the DSR (Allendoerfer et al., 2000). Because the legacy system has changed substantially since the simulations were conducted in the 1990s, most notably the addition of URET, and because traffic volumes and procedures have also changed, it is necessary to collect new baseline data for the legacy system for comparison to ERAM.

In addition, the original studies simulated only routine situations with no special operational situations like outages or emergencies (Allendoerfer, Pai et al., 2006). The original studies also did not include any scenarios targeting the areas where ERAM makes significant changes. For these reasons, we need to obtain baseline data for the legacy system using additional specialized scenarios.

2. Method

We conducted an analysis of areas where ERAM makes major changes compared to the legacy system. Because ERAM is still under development by Lockheed Martin and is not yet available at the William J. Hughes Technical Center, we conducted the analysis by examining ERAM requirements documents, specifications, white papers, and briefings. We also discussed the changes with technical, operational, and human factors personnel from the Air Traffic Organization En Route and Oceanic Service (ATO-E) and the ERAM Program Office.

Our analysis team consisted of four human factors specialists from the NAS Human Factors Group and an en route subject-matter expert (SME) detailed to our laboratory. The SME has had significant experience with human factors and served for a time as a member of the ERAM User Team during the requirements development phase.

We assigned each member of the team areas of ERAM changes to research. We compiled and consolidated their findings to identify common themes and overall categories of ERAM changes. We then reviewed the overall areas with personnel from ATO-E and the ERAM Program Office to refine our conclusions and develop possible approaches to testing the changes.

3. Results

The following sections describe several important ERAM changes compared to the legacy system and discuss how the effects of those changes on controllers might be measured. In the tables and accompanying discussions, a "standard metric" is one that is normally collected in HITL baseline simulations (Galushka et al., 1995; Allendoerfer et al., 2000). Appendix A provides a list of standard metrics. Standard metrics include the number of operational errors, the number of aircraft handled, and the number of data entries. In some cases where the ERAM change may have a substantial impact on a standard metric, we include it individually in the table. Where appropriate, we also describe considerations that should be taken into account during the UCA or baseline simulation targeting that change.

3.1 ERAM Backup Architecture

In the legacy system, the HCS is the primary flight data processing computer, and controllers use it to control traffic during normal operations. Depending on the facility, the backup system is the Enhanced Direct Access Radar Channel (EDARC) or the Enhanced Backup Surveillance (EBUS). EDARC has fewer capabilities than the HCS, especially in the flight data processing and safety alerting areas. As a result, when the HCS fails or is taken offline, controller efficiency is reduced and workload is increased. EDARC is normally used only on overnight shifts for testing, software updates, and maintenance. Operational use during normal operations is rare. EBUS is a recent improvement and is still being fielded as of this writing. Though EBUS provides capabilities missing from EDARC, such as alerting functions, it is still not equal to a full HCS in terms of functionality. By the time ERAM is deployed, EBUS will have replaced EDARC at all ARTCCs.

ERAM contains two parallel, equivalent channels known as Channel A and Channel B. Each channel can be configured as Active, Backup, or several other modes used by Technical Operations (TechOps) personnel. The active channel is used by controllers to work traffic. The active channel receives data from all external interfaces and it sends synchronization information to the backup channel. The backup channel receives data directly from some external interfaces and conducts flight data and weather processing independently from the active channel. Other data in the backup channel are received from the active channel through a synchronization process. Under normal conditions, when synchronization is operating correctly, all data are available on both channels. Unlike the legacy system, the backup channel is fully functional. Controllers should experience no loss of functionality or efficiency when switched to the backup channel and no corresponding increase in workload.

If a problem is detected with one of the channels, alerts are generated at the ERAM Monitor and Control (M&C) position. If the active channel fails, TechOps will activate the backup channel by a command at the M&C position. The controller console indicates which channel is being viewed and the status of each channel. Controllers can switch from Channel A to Channel B using a keyboard entry. TechOps, the Air Traffic Supervisors, and the controllers will verbally coordinate channel switchovers as needed.

3.1.1 Proposed Data Collection Method

Conducting HITL simulations that contain simulated equipment outages is the most direct way to examine the effects of the ERAM backup architecture on controllers. On the legacy system, the simulation would fail the HCS and require controllers to switch over to EBUS. After a period of time, the HCS could be restored, and controllers would transition back to the HCS according to procedures and traffic volume. On ERAM, the equivalent failure would be to fail the active channel and require controllers to switch to the backup. In both cases, researchers would use the same traffic scenarios and controller participants to ensure comparability. Safety, efficiency, and workload data would be collected before and after the HCS or channel failure to document the effect of the failure in both systems.

Because the active and backup channels are equivalent in ERAM, the loss of the active channel may lead to no measurable changes in safety, efficiency, or workload compared to normal operations. In that case, there might be a minor disruption in the control room while the switch over is coordinated, but this disruption would be short-lived with no measurable changes to overall performance. If we obtained this result, ERAM would compare favorably against an

HCS failure on the legacy system where controllers need to switch to EBUS and then manage the traffic according to the reduced capabilities. On the other hand, it is possible that a channel failure on ERAM has unanticipated consequences for controllers. Table 1 lists potential measurable differences between an HCS failure on the legacy system and an equivalent channel failure on ERAM.

Table 1. Potential Differences Resulting from Backup Architecture

Construct	**Metric**	**Possible Direction & Magnitude of Differences After Outage Compared to Before Outage**	
		Legacy System	**ERAM**
Safety Risk	Standard metrics	No measurable change	No measurable change
Efficiency	Standard metrics	Large decrease	No measurable change
	Elapsed time from the HCS/channel outage until controller is working at the same efficiency as before the outage.	Reduction in efficiency may last until after HCS is restored.	No reduction in efficiency; returns to baseline efficiency immediately.
	Elapsed time from restoration of HCS/channel until controller is working at the same efficiency as before the outage.	Reduction in efficiency may last long after HCS is restored and traffic levels are restored by Traffic Flow Management (TFM).	No reduction in efficiency; returns to baseline efficiency immediately.
Workload	Standard metrics	Large increase until traffic flow is adjusted to account for the outage.	Small increase while switch over is being coordinated.
	Elapsed time from the HCS/channel outage until controller is working at the same workload level as before the outage	Increase in workload lasts until HCS is restored or traffic flow is adjusted to account for the outage.	Increase in workload lasts until switch over is complete and returns to baseline quickly.

Table 1. Potential Differences Resulting from Backup Architecture (continued)

Construct	Metric	Possible Direction & Magnitude of Differences After Outage Compared to Before Outage	
		Legacy System	**ERAM**
Workload (cont.)	Frequency of commands related to executing the switch over	Small increase	Small increase
	Frequency of ground-ground communications	Moderate increase within affected ARTCC while switch over is coordinated. Large increase between affected ARTCC and adjacent facilities while coordinating reroutes and manual flight plan transfers.	No measurable change or small increase within affected ARTCC while switch over is coordinated. No measurable change after switch over is complete. No measurable change between affected ARTCC and adjacent facilities.
	Frequency of route amendments	Moderate increase while some traffic is rerouted away from affected ARTCC. Moderate increase while traffic is rerouted back to affected ARTCC once HCS is restored.	No measurable change

3.1.2 Considerations for Testing

It is our intention to conduct a realistic, rigorous, and fair comparison of both systems in this situation. However, given resource constraints, there is some question as to whether HITL simulations are warranted here. The main reason for conducting baseline simulations is to determine if ERAM is at least as effective as the legacy system. There is little question that if the ERAM backup and redundancy capability works as designed and does not contain technical problems or bugs, that it would be at least as effective as the current EDARC or EBUS capability. Controllers should notice no reduction in safety or operational efficiency due to the outage. If that is true, why are baseline simulations necessary in this case? Operational Test and

Evaluation (OT&E) will identify the technical problems and bugs, which standard test and regression methodologies will address.

We have two responses to this question. First, it is possible that the backup capability will not work as designed. OT&E may find that the capability meets requirements but that some of its features are suboptimal. Depending on the nature and severity of the problems and the program's schedule, workarounds or other interim solutions may be necessary. The baseline simulations could help ensure that the workarounds do not negatively affect operations or at least not worse than EDARC or EBUS. Second, the ATO-E may wish to demonstrate benefits of ERAM as part of a business case. The baseline comparisons could provide a clear, data-driven demonstration that the new ERAM backup architecture reduces the negative effects of an HCS failure and that no unintended consequences exist. In this case, the simulations would not serve only the ERAM test program but the ERAM program overall.

We have not routinely conducted realistic simulations of equipment outages in ATC human factors research. However, because the new backup architecture provided by ERAM is such an important capability, we believe creating such a simulation is feasible if the issues described in the following four subsections are addressed.

3.1.2.1 Ghost Sectors and Confederates

A total HCS outage affects the entire ARTCC and all the adjacent facilities. Our past baseline studies simulated four sectors with the role of all other sectors and facilities played by one or two "ghost" controllers (i.e., controllers who were confederates of the researchers who communicated with the participant controllers as if they were adjacent sectors or facilities). A realistic simulation of an HCS outage would require a complex mapping of communication frequencies and many more ghost controllers than we normally use. Figure 1 shows an example of a simulation of four sectors surrounded by seven ghost sectors. The controllers staffing the ghost sectors play the role of adjacent sectors (labeled "ADJ SECTORS" in the figure) or other adjacent facilities ("ADJ ARTCC," "ADJ TRACON"). A single ghost controller plays the role of multiple ghost sectors, just the way a simulation pilot plays multiple aircraft, but this requires careful layout of the communication channels and extensive testing and shakedown.

Because we desire high levels of realism in the baseline simulations, staffing the ghost sectors with non-controllers probably would not be adequate. One approach is to rotate the participant controllers through several roles. In some runs, they would play the ghost controllers; in others, they would be participants. We would also need other controllers to play the roles of supervisors and TFM confederates.

3.1.2.2 Training and Familiarity

Though controllers must maintain currency on all procedures and commands, most do not routinely switchover to the backup system during heavy traffic operations. As a result, they may not be quite as skilled on these procedures as they are for activities they perform every day. Before testing, controllers should participate in a refresher simulation to make sure they are current and proficient on the switchover procedures.

In addition, depending on the ERAM testing schedule, some facilities may not be fully operational on EBUS or may have significantly less experience with it than other facilities. Drawing controllers from facilities that recently transitioned from EDARC to EBUS would introduce another experimental confound, and we recommend against doing so.

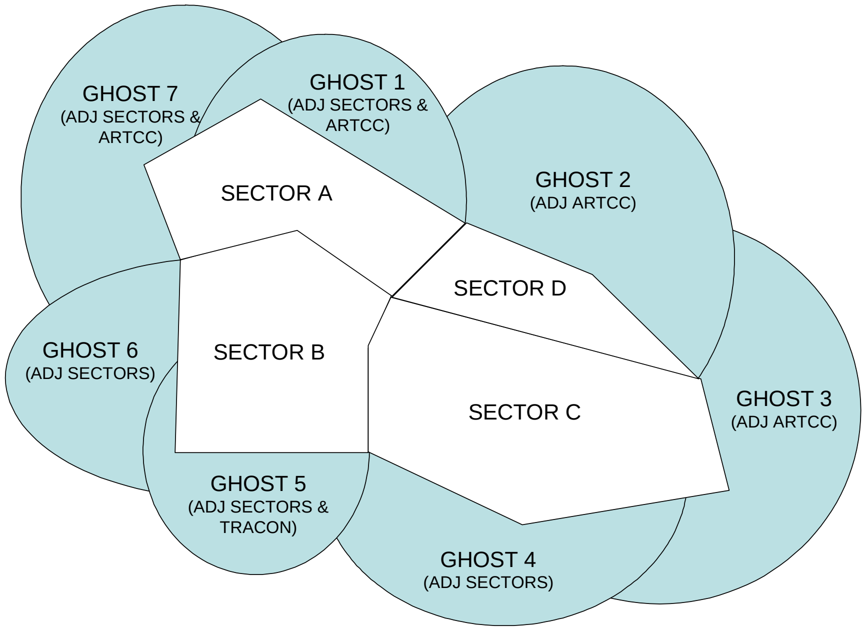

Figure 1. Four sectors with seven surrounding "ghost sectors" playing the roles of all adjacent sectors and facilities.

3.1.2.3 Equivalency of Outages

Creating equivalent outages in the legacy system and ERAM may be difficult due to the differences between the architectures. Is simply disabling one ERAM channel truly equivalent to disabling the HCS? Is this a fair comparison or is the test biased in one direction or another? Researchers must conduct an analysis of both systems in collaboration with the ERAM Program Office and knowledgeable TechOps personnel to develop techniques to create operationally equivalent outages on both systems. Creating equivalent outages requires identifying the specific data or functions we wish to disable and then creating methods for disabling them. Researchers must formalize the disabling method into different simulation scripts for both systems. Following the scripts should produce similar outcomes across systems even when the individual steps within each script are different.

In addition, if creating the outage requires technical skills, such as knowledge of the ERAM M&C position and how to disable its equipment, we need to ensure that simulation personnel obtain these skills or that personnel from the Technical Center, TechOps, or Lockheed Martin are available to help create the outages.

3.1.2.4 Data Collection

Because so much of the intended purpose of the simulation will occur in the few minutes preceding and following the outage, it may be desirable to change the timing or sampling rate of some measurements to correspond to the timing of the outage. For example, we normally collect workload ratings every 2 to 5 minutes. If the effect of the outage is short-lived, such a window of time could miss or dampen some of the effects on controller workload. One solution might be to issue several workload probes immediately before and after the outage to temporarily increase the sampling rate during the period of greatest interest. Increasing the rate of workload probes may also affect workload, but the increase would occur in both the ERAM and legacy conditions equally.

3.2 Areas of Interest

In the legacy system, controllers do not have full access to flight data for an aircraft inbound to the ARTCC until it reaches the facility boundary. A controller staffing a sector receiving handoffs from an adjacent facility is only able to see limited datablocks (altitude and beacon code) for flights in the adjacent facility. Once an interfacility message is sent, the controller will be able to view the full datablock.

However, there are many circumstances where it would be beneficial for controllers to have access to full flight data for aircraft in adjacent facilities to use the quick look, point out, and handoff capabilities more effectively. For example, flights regularly “clip the corners” of sectors or briefly pass through a sector’s altitude. In the legacy system and ERAM, if the affected sectors are within a single ARTCC, controllers can point out the aircraft to the middle sector and avoid a handoff that lasts for only a few minutes, as shown in Figure 2. The point out is beneficial to pilots because only a single frequency change is needed and also reduces communications and data entry for controllers. This kind of flexibility is very useful when the traffic situation is changing dynamically and aircraft are not flying typical routes, such as when they are deviating around weather or are flying direct to a destination rather than following a standard route.

However, if the boundary in question is an ARTCC boundary, these capabilities are unavailable or involve additional workload for controllers. For example, the adjacent ARTCC may not have the flight plan data required for a legal point out. In some cases, local procedures or letters of agreement have been established to address this interfacility coordination, but these would not normally be available when routes are changing dynamically.

In a second example, the geography of the ARTCC boundaries creates operational inefficiencies. For example, Figure 3 shows the boundary between Jacksonville ARTCC (ZJX) and Miami ARTCC (ZMA), the locations of Orlando International Airport (MCO) and Tampa International Airport (TPA), and their associated TRACONs. Both airports are located near the ARTCC boundary, and their TRACONs straddle it. TPA is located within the ZMA airspace, MCO within ZJX. Controllers must hand off northbound departures from TPA first to ZMA even though they will be handed almost immediately to ZJX. In a similar way, southbound departures from MCO are handed off first to ZJX and quickly handed off to ZMA. To avoid extra handoffs and coordination and to increase efficiency, controllers could make flight plan amendments and manually coordinate the flight plans with the other ARTCC. However, this creates substantial extra workload for controllers. To avoid this workload, controllers initially route the aircraft away from the most desirable flight path and into the ARTCC that already has the flight plan. This is an inefficient use of the airspace and increases workload for the pilots. It is used in the legacy system, however, because the alternative is too workload intensive for controllers to use on a regular basis.

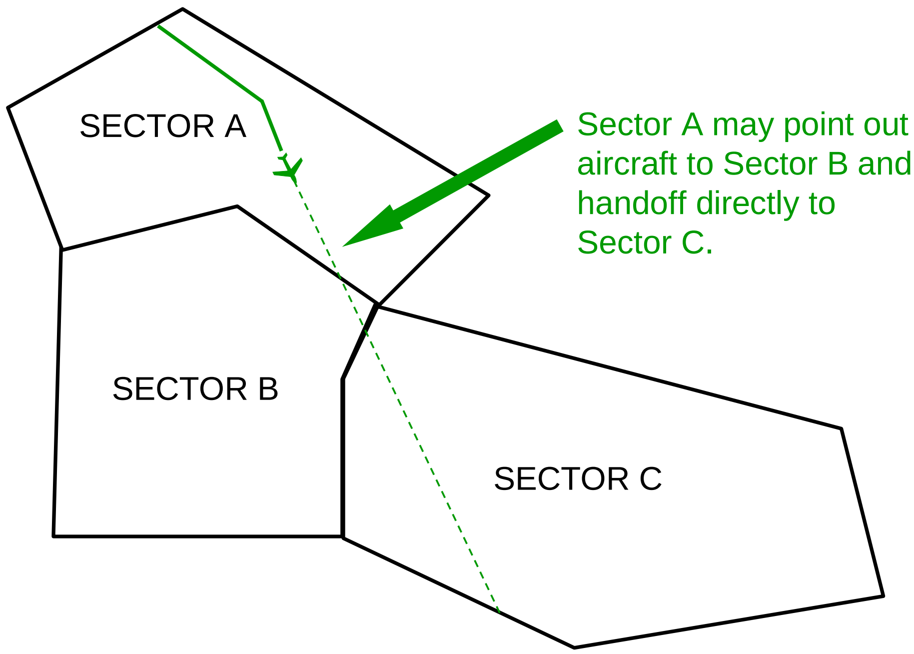

Figure 2. A point out between sectors within one ARTCC.

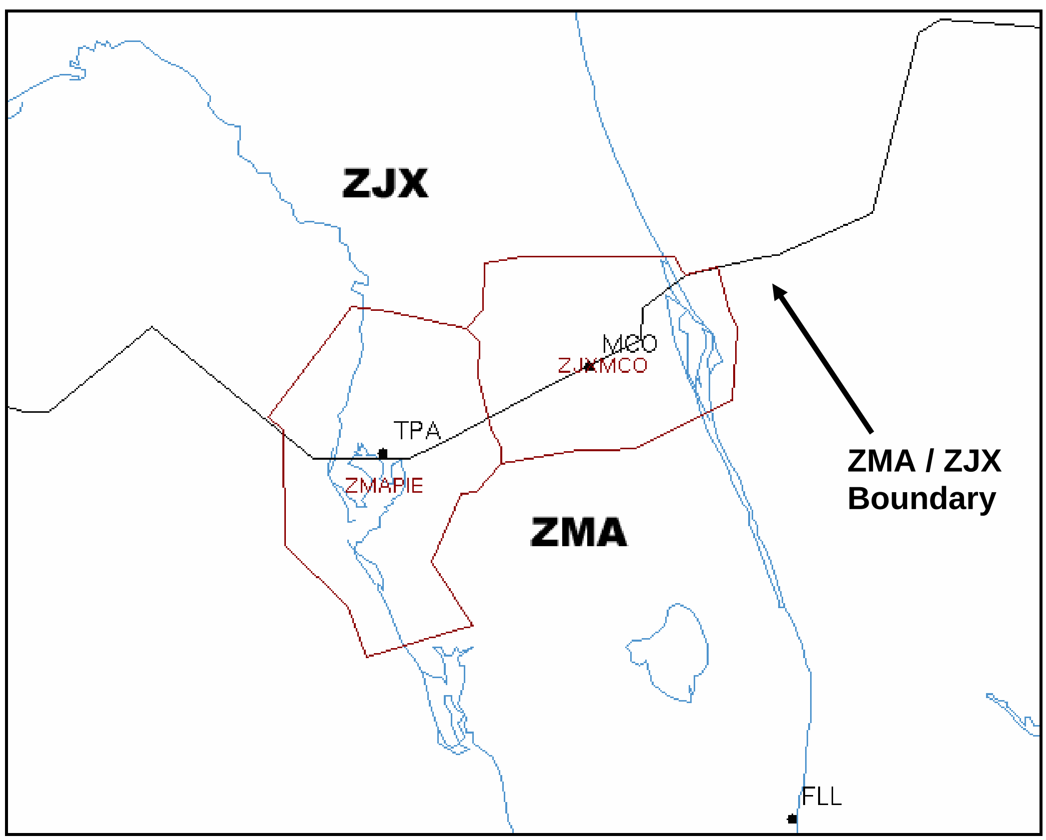

Figure 3. Boundary between ZJX and ZMA and associated TRACONs.

To address issues like these, ERAM adds the concept of AOIs, which are adapted volumes of airspace outside the ARTCC. Aircraft in the AOI are under the control of an adjacent facility. Controllers have access to flight data for all aircraft that penetrate their AOI, regardless of whether the flight will actually enter their ARTCC airspace. These data include flight plans, track updates, handoff status, and altitude assignments. Controllers also will have flight data for aircraft that they have handed off to the next facility but are still within the AOI. In essence, the AOI looks and operates like another sector in the ARTCC.

In the examples discussed previously, AOIs are designed to allow controllers to work traffic near the ARTCC boundary more effectively. Less manual coordination should be necessary, which should decrease controller workload and encourage use of more efficient and flexible procedures and routes.

3.2.1 Proposed Data Collection Method

The most direct way to examine the effects of ERAM AOIs on controllers is to conduct baseline simulations of sectors and scenarios near ARTCC boundaries where AOIs would be useful. With help from the field, we would identify likely facilities and sectors and simulate traffic that briefly crosses an ARTCC boundary or *would* cross it if such a procedure were available. The traffic scenario would be designed to encourage use of interfacility point outs, handoffs, and quick looks. Table 2 lists metrics that could be used to measure the effects of the AOIs on safety risk, efficiency, and workload.

Table 2. Potential Differences Resulting from Areas of Interest

Construct	Metric	Possible Direction & Magnitude of Differences from Legacy System
Safety Risk	Standard metrics	No measurable change
Efficiency	Standard metrics	Moderate increase
	Time in sector (receiving sector/facility)	Small increase because the sector will receive the aircraft sooner
	Number of aircraft handled (intervening sector/facility)	Small decrease because the intervening sectors do not control the affected aircraft.
	Distance flown (affected aircraft only)	Moderate decrease because controllers will be able to provide more direct routes
Workload	Standard metrics	Moderate decrease
	Frequency of point out commands	Moderate increase

Table 2. Potential Differences Resulting from Areas of Interest (continued)

Construct	Metric	Possible Direction & Magnitude of Differences from Legacy System
Workload (cont.)	Frequency of route amendments	Moderate decrease
	Frequency of handoff commands	Moderate decrease
	Frequency of ground-ground communications	Small increase because coordination is needed for point outs
	Frequency of air-ground communications	Small decrease because intervening sectors do not need to communicate with the aircraft

3.2.2 Considerations for Testing

As with the ERAM backup capabilities, there is some question as to whether baseline simulations of ERAM AOIs are needed, given their significant cost. If the AOI capability works as designed and does not contain any bugs, it would almost certainly provide benefit compared to the legacy system. There is little reason to believe that a point out between two facilities would require more workload to execute than a point out between two sectors within a facility. If the ERAM AOI capability allows controllers to use advantageous procedures that are currently unavailable, this would almost certainly indicate that ERAM is at least as effective as the current system in this area.

As we discussed in Section 3.1.2, the baseline simulations could be useful to establish ERAM benefits and to evaluate any workarounds or limitations. In addition, the baseline simulations could allow facilities to test new procedures or letters of agreement that the ERAM AOI capability makes available for the first time. For example, ZJX and ZMA could develop draft procedures for use in the baseline simulations that contained advantageous routes made possible by the AOIs. The simulations could ensure that the draft procedures work well and do not have any unintended consequences.

Like other simulations discussed in this report, we have not attempted a simulation of this scope in previous baseline studies. However, given the possible operational benefits of AOIs, such simulations should be given serious consideration. Simulations targeted at AOIs should take into account the considerations discussed in the following subsections.

3.2.2.1 Scope

This test would require simulating two ARTCCs simultaneously, which we have not yet attempted as part of a baseline simulation. It would require two parallel ERAM systems communicating with each other and connected to a single Target Generation Facility (TGF) simulation. Testing interfacility communication is part of the standard ERAM test program so laboratory configurations that could accommodate this simulation should be available at the Technical Center. Ideally, the two ERAM systems would be located in different rooms, and controllers would communicate over the Voice Switching and Communication System (VSCS).

Simulating two ARTCCs increases laboratory and personnel costs substantially. Technical personnel would need to staff each ERAM system. Scenario development and shakedown would be more complex because researchers would need to create and test traffic for two ARTCCs. Finally, we would need to recruit SMEs from two field facilities for our scenario development, shakedown, and data collection phases.

3.2.2.2 Procedures

Because the AOIs may make routes available that are currently not available in the field, we would need to create procedures for use in the simulation that controllers will follow in the legacy and ERAM conditions. Creating procedures would require consultation with SMEs from the selected facilities and possibly procedures experts from ATO-E.

In cases where the procedures do exist but are rarely used because of the extra workload, we may need to provide controllers with refresher training to ensure that they are familiar and experienced enough to execute the procedures during the simulation.

3.3 User Interface Changes

ERAM provides a variety of new UI capabilities compared to the legacy system. In general, these changes are intended to standardize the UI across components that had been developed independently in the legacy system. For example, the URET windows and views have a somewhat different appearance and interaction style than DSR because both components were independent systems at one time, developed by different teams for different purposes. ERAM is a unified system, and the UI differences between the components are reduced.

From a human factors and usability perspective, UI standardization and consistency is very beneficial. Most important, the likelihood of negative transfer is greatly reduced. Negative transfer occurs when a person applies techniques learned for a previous situation to a new situation, but those techniques are actually inappropriate for the new situation. For example, if two systems each have the same command available but the syntax of the command is different on each, users may unintentionally use the syntax for one system when using the other, leading to errors and frustration. In addition to reducing negative transfer, consistent UIs reduce training requirements by leveraging what users already know when features and capabilities are added.

Some ERAM UI changes are designed to reduce controller workload by reducing the number of steps needed to enter a command. For frequently used commands, this could result in measurable differences in data entry workload. However, these changes are unlikely to have a measurable effect on overall controller efficiency or safety, at least in the short run. In theory, a reduction in controller data entries would free up some mental and physical resources. Newly available resources might allow controllers to spend more time doing separation assurance or optimizing routes, which would in turn improve safety or efficiency. Considered across the whole NAS over a long period, small UI changes may indeed have a measurable impact.

However, the impact is not likely to be detectable during a 60-minute simulation of only a few sectors. In the following subsections, we discuss ERAM UI changes where we believe differences could be measured during testing activities.

3.3.1 Interactive Lists

In the legacy system, a number of tabular lists appear on the situation display including the Hold, Inbound, Departure, Group Suppression, Conflict Alert, and Metering lists. Depending on the list, controllers can toggle the list on and off, move the list, and change its brightness.

In ERAM, lists become interactive. The lists maintain their primary function of displaying tabular information but provide many new capabilities. Some of the new capabilities make the lists appear and behave more like other views in ERAM and generally make them behave more like modern UI windows. Controllers can adjust the size, font, brightness, and transparency of the view using on-screen menus and the trackball cursor. Controllers can also sort lists.

From a UI perspective, one of the most interesting features of interactive lists is a capability for controllers to create "tearoff" sublists. A controller selects a section of a main list, detaches it with the cursor, and positions it in a separate location. Controllers may use the tearoff sublists when the sublist applies only to a certain geographic location on the screen, such as a fix. For example, a controller might tear off a sublist from the Hold List and position the sublist near the fix where the aircraft are holding. In this way, the controller is placing information about the aircraft in hold in close proximity to the aircraft targets and datablocks. Placing information nearby may reduce the amount of time the controller spends searching the screen to obtain information. All the information about the aircraft in hold would be contained in a relatively small region.

Some interactive lists provide even more functionality in which controllers can interact with items in the list to compose commands that are keyboard entries in the legacy system. For example, in the legacy system, to remove UAL110 from the Hold List, a controller makes one of the following entries using the keyboard only or keyboard and trackball.

QH C UAL110	
QH C 172	← 172 is the computer identifier (CID)
QH C 4672	← 4672 is the beacon code
QH C <trackball>	← controller clicks on UAL110 with the cursor

In ERAM, these commands are still available but with an additional option of clicking on UAL110 in the Hold List. The list then provides a confirmation prompt to ensure that the controller actually intended to remove the aircraft from the list, as Figure 4 shows. If the confirmation is made, the aircraft is removed immediately, and the proper command is sent to the flight database. In theory, this will help controllers by providing all the functionality related to hold in a single location. They will not need to search the screen for the preview area or datablocks. While they are interacting with the Hold List, they can accomplish all their functions from one place.

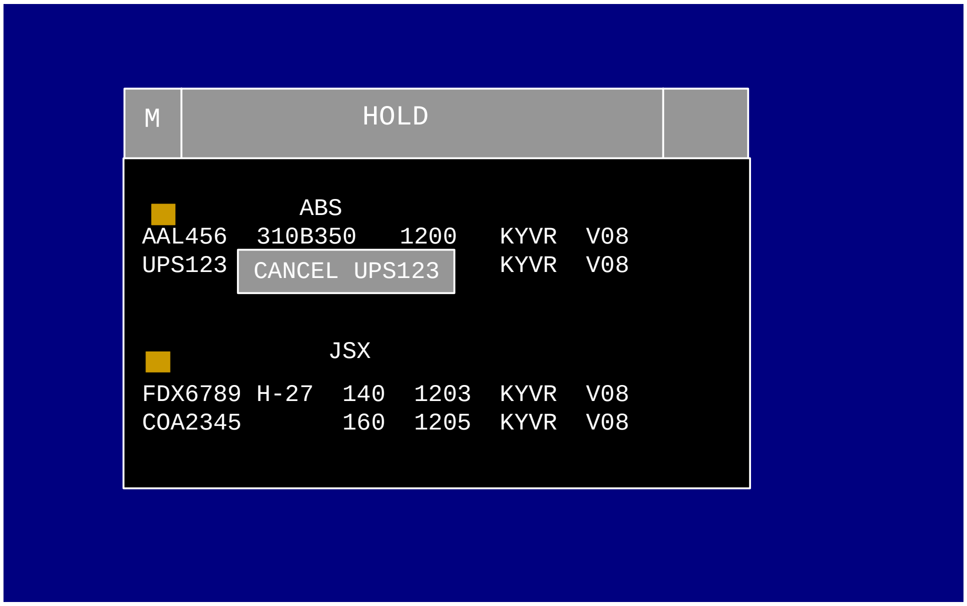

Figure 4. ERAM Hold List with cancel confirmation.

Other capabilities of the interactive Hold List include the ability to change the Expect Further Clearance (EFC) time and the capability to change altitudes directly from the list, as shown in Figure 5. To change the altitude at which an aircraft is holding, the controller can click on the altitude in the Hold List. A menu appears similar to the altitude flyout menu available on datablocks in DSR and ERAM. The controller can scroll to the desired altitude and select it. Other ERAM interactive lists contain similar functionality for parameters applicable to that list.

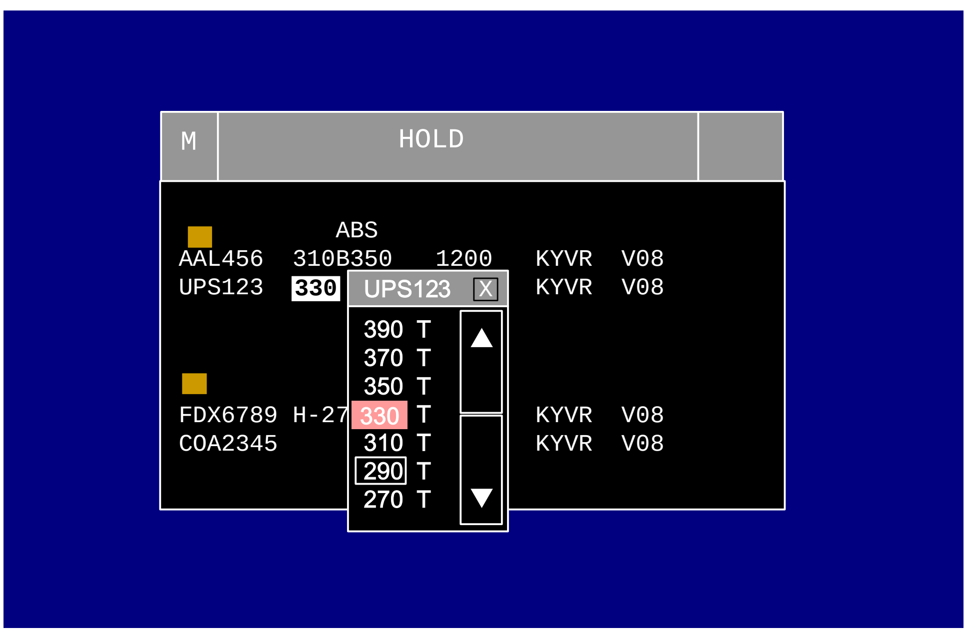

Figure 5. ERAM Hold List with altitude submenu.

3.3.1.1 Proposed Data Collection Method

We propose that the ERAM interactive lists be examined in two ways. The first approach is to include the interactive lists as an additional condition in the UCA. This approach would allow us to compare the interactive lists to other data entry methods available in the legacy system and ERAM.

For example, the interactive Hold List allows controllers to change altitudes for aircraft in hold directly from the list. Controllers can still change the altitude in traditional ways, such as the QZ keyboard entry and the altitude flyout menu on the datablock. For aircraft in hold, using the interactive list may be faster or more convenient. In the UCA, controllers would make a series of altitude changes for aircraft in hold. They would make the same entries with the legacy system and ERAM in several conditions. In one condition, they make the commands using the keyboard and trackball, the second using the flyout menus, and the third, for ERAM only, with the Hold List. In each case, we would measure the controllers' speed, accuracy, and visual attention. These data would allow us to determine which method provides the most effective interaction method and measure the benefit, if any, of changing altitudes using the Hold List. Table 3 provides a list of possible metrics that could be examined in the UCA for interactive lists.

Table 3. Potential Differences Resulting from Interactive Lists
(Usage Characteristics Assessment)

Construct	Metric	Possible Direction & Magnitude of Differences from Legacy System
Safety Risk	Not measured	Not measured
Efficiency	Not measured	Not measured
Workload	Standard metrics	Moderate decrease
	Speed for entering commands associated with lists	Small increase
	Accuracy for entering commands associated with lists	Small increase
	Attention required to complete commands associated with lists	Moderate decrease

The second approach is to examine the interactive lists as part of the baseline simulations. Controllers would make their own decisions regarding when and how to use the interactive lists to respond to the traffic situation. To encourage controllers to use the lists, we would design scenarios where using the lists would be an effective strategy though not necessarily the only one. For example, to encourage use of the interactive Hold List, we could present a traffic scenario with numerous aircraft in hold. Because the simulation is dynamic and interactive, controllers would decide on their own how best to handle the situation. In the legacy system, controllers would handle the situation the way they do now. In the ERAM condition, however, controllers could choose to use the new interactive list if they wished. We would be able to examine how controllers use the lists in a realistic operational context and determine if the lists have a measurable effect on safety, efficiency, or workload. In some cases, the interactive lists may lead to more timely issuance of commands associated with the lists because the lists are located in more convenient locations. Table 4 describes potential measurable differences in the ERAM condition compared to the legacy system.

Table 4. Potential Differences Resulting from Interactive Lists (Baseline Simulation)

Construct	**Metric**	**Possible Direction & Magnitude of Differences from Legacy System**
Safety Risk	Standard metrics	No measurable change
Efficiency	Standard metrics	Small increase
	Timeliness of commands associated with interactive lists	Moderate increase
Workload	Standard metrics	Small decrease
	Frequency of interaction with lists	Moderate increase
	Frequency of interaction with other methods	Moderate decrease

3.3.1.2 Considerations for Testing

Including interactive lists in the UCA and the simulations requires researchers to identify situations where the lists would be useful and develop scenarios around those situations. For the Hold List, this would require including scenarios with an increased number of holds. For the Metering List, this would require including scenarios with metered traffic and simulating sectors where metering is commonly done. Situations must be chosen very early in the scenario development process.

In addition, training controllers to use the interactive lists will be an important issue. In the UCA timeframe, no controllers will be available with any significant experience using the ERAM interactive lists. We will need to provide multiple training sessions and require UCA participants to achieve a performance criterion before data collection can begin. Even so, the participants' experience with the ERAM interactive lists will be drastically less than with the legacy system lists. In the simulations, we assume that controllers will have undergone some training and will have acquired experience with the interactive lists as part of ERAM OT&E. In fact, if we conduct the simulations at the end of the OT&E, the controllers from the OT&E team will be the most experienced users of the ERAM in the world. However, even with this level of experience, the participants will have substantially less experience with the ERAM interactive lists than with the legacy system lists.

Unfortunately, there is no good way around this. Controllers have extensive training and years of experience using the legacy system UI and almost none with ERAM. The most valid approach for comparing the systems would be to let controllers use the new UI operationally for months (or even years) and only then compare performance across systems. That approach, however, is not useful for a test activity that is trying to decide whether the ERAM is ready to be fielded. Instead, we propose to examine the results of the UCA and simulations with SMEs and FAA training specialists to determine if any observed differences between systems can be attributed to lack of experience, and if training or experience could quickly reduce any negative impact. We also propose to conduct a follow-up UCA and simulations after ERAM has been in the field for 9 to 12 months. If any problems from the initial UCA or simulations are still present, we know that there is a genuine usability problem in the ERAM UI that merits redesign or other improvements.

3.3.2 Enhanced Toolbar and Macros

Data entry consumes a significant portion of controllers' attention and effort. It would be beneficial if the automation system could reduce the amount of data entry required to free up controllers' cognitive and physical resources for other tasks. Features such as automatic handoff initiate already exist in the legacy system to help reduce data entry requirements.

ERAM provides several new UI capabilities as part of its Enhanced Toolbar that are intended to reduce data entry requirements. The Enhanced Toolbar allows controllers to organize buttons on the toolbars according to their needs and preferences. For example, if a particular sector frequently uses a button located on a submenu, the controller can tear off the button and place it in a more convenient location. More convenient buttons should reduce the workload and attention needed to navigate the menus. These UI changes make sense from a usability perspective, but it is an empirical question how often or how effectively controllers will choose to reorganize their toolbars.

A second interesting feature of the ERAM Enhanced Toolbar is the ability to create macros and assign them to toolbar buttons. A macro is a series of commands that can be executed by entering a single command, in this case pressing the assigned button. The intention of the ERAM macros is to allow controllers to issue multiple commands more easily.

For example, current procedures and practices may require that when an aircraft reaches a certain fix, the controller assigns a new altitude, hands the aircraft off, and adjusts the datablock offset. In the legacy system, every aircraft reaching the fix would require three keyboard entries. In ERAM, the controller can create a macro containing the three commands and assign it to a button. When each aircraft reaches the fix, the controller simply activates the macro and clicks on the aircraft. A macro has the potential to significantly reduce the controller's data entry workload for certain situations, at least once a macro is properly created.

In addition, because controllers can tear off the toolbar button and place it near the fix, they may spend less time navigating menus and moving the cursor to execute the commands. Figure 6 shows an example of an ERAM macro that a controller has created, torn its associated button off the toolbar, and positioned it in the main situation display. If UPS123, AAL456, and NWA789 all need to receive the commands associated with macro MAC1, the controller can issue these commands quickly, each from the same location.

Another area where macros could be useful is when controllers receive a National Playbook reroute that they must issue manually to multiple aircraft. Depending on the number of fixes listed in the reroute, the data entry requirements for this in the legacy system could be very large. In ERAM, however, controllers could store the reroute commands in a macro and more easily issue them to multiple aircraft.

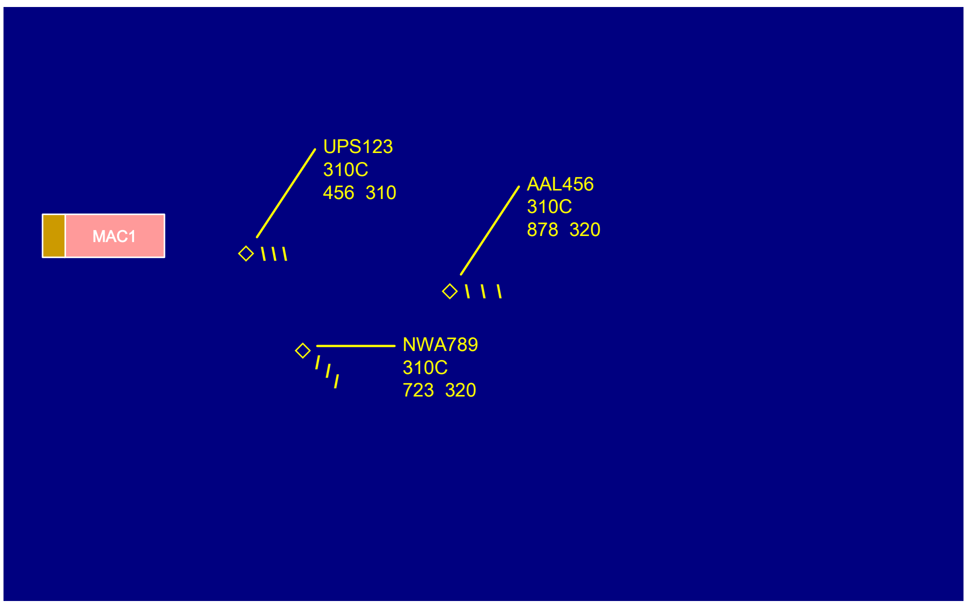

Figure 6. ERAM macro button that has been torn off and positioned on the situation display.

<u>3.3.2.1 Proposed Data Collection Method</u>

We propose to assess the effectiveness of the Enhanced Toolbar and macro capabilities as part of the baseline simulations. We could assess he macros during the UCA, but there is little question that a properly programmed macro would be faster and more accurate than issuing each command manually. Instead, in the UCA, we could measure the effort necessary to create the macro compared to the effort required to issue the commands individually.

More important, however, is the actual use of macros while controlling traffic. How do controllers make use of the functionality under realistic conditions? Does their use of macros lead to improvements in operational performance? We propose to simulate situations where using macros would be beneficial if the controllers choose to use them. The most straightforward example is where controllers execute the same series of commands each time an aircraft reaches a particular location. We would select sectors and build traffic scenarios for the simulation to create this type of situation and allow controllers to choose how to handle it. In the legacy system condition, controllers would need to execute each of the commands manually. In the ERAM condition, controllers could choose to use the macros or continue to execute the commands manually. Table 5 lists potential observable differences in the ERAM condition compared to the legacy system.

Table 5. Potential Differences Resulting from Enhanced Toolbars and Macros

Construct	Metric	Possible Direction & Magnitude of Differences from Legacy System
Safety Risk	Standard metrics	No measurable change
Efficiency	Standard metrics	No measurable change to small increase
	Timely issuance of commands associated with toolbars and macros	Moderate increase because commands are located in more convenient places and/or are part of a macro
Workload	Standard metrics	Moderate decrease
	Frequency of commands associated with macros that are manually entered	Large decrease once the macro has been built and works properly
	Frequency of data entry errors for commands associated with macros	Large decrease once the macro has been built and works properly
	Frequency of interactions with toolbars	Moderate increase

3.3.2.2 Considerations for Testing

As with the interactive lists, testing the Enhanced Toolbar and macros in the baseline simulations is straightforward. We would select sectors and build scenarios where the macros could be advantageous. No additional scripting or special events would be required in most cases, though a National Playbook reroute situation might require extensive development to be realistic. We could run this test as part of the day-in-the-life scenarios.

As with the interactive lists, training participants to use the new ERAM macros and toolbars will be a challenge, especially considering how much experience controllers have with the legacy system versus how little they have with ERAM.

3.4 Tracker

ERAM uses different tracker software and algorithms than the HCS. Tracking is an extremely complex technical subject, and its effects on controller performance are not well understood.

Surveillance data contain inherent error, and all tracking contains some degree of uncertainty. Any algorithm can be more accurate in some situations and less accurate in others.

Because controllers have used the HCS tracker for many years, their expectations for tracker behavior are deeply rooted and influenced by the HCS. Controllers have developed procedures, work practices, and habits in response to its idiosyncrasies. When the ERAM tracker is implemented, even if it is shown to perform better on technical performance tests, there is likely to be some disruption of controllers' work practices. As controllers become experienced with the ERAM tracker, they will eventually reorient their expectations. The duration and severity of the disruption is an empirical question. During the testing and deployment of the Standard Terminal Automation Replacement System (STARS) in the terminal domain, which also provided a new tracker, extra testing and refinement was needed to ensure that field sites were satisfied with the tracker performance before going operational with STARS.

In addition, the ERAM tracker will have its own idiosyncrasies and is likely to perform better in some situations than others. Controllers may need to develop new procedures or work practices in response. During the first few weeks of operational use, the controllers' ERAM training will be relatively fresh in their minds. They will be very aware of the newness of the system, and they are likely to be more vigilant when watching for tracker discrepancies. During this initial period, it is unlikely that tracker discrepancies between systems will have any measurable effect on safety, though it is possible that some effect on efficiency or workload could occur. Developing fully formed expectations and work practices for the ERAM tracker will take weeks or months while controllers experience a broad range of traffic situations.

3.4.1 Proposed Data Collection Method

Without a thorough understanding of the situations where the ERAM and HCS trackers differ, we cannot propose an assessment of the effect of the ERAM tracker on controller performance at this time. However, during the ERAM test and evaluations that will occur throughout the next several years, Lockheed Martin or FAA testers may identify situations where the ERAM and HCS trackers reach significantly different conclusions. This determination may be made based on the techniques and metrics developed by the AMTWG. When such situations are identified, it would be appropriate to develop realistic traffic scenarios of the situations and have controllers work the scenarios in baseline simulations. Depending on the nature of the situations, different human performance metrics could be brought to bear and need to be selected or developed at that time. Predictions about the magnitude and direction of differences cannot be made until the situations are identified.

Because tracking represents a major change between the legacy system and ERAM and is a major focus of the AMTWG, we will continue to look for cases where further research is required and develop methods and metrics as needed.

3.5 Safety Alerts

In addition to using a different tracker, ERAM also uses different safety alerting algorithms from the legacy system. Like tracking, alerting is a complex technical subject. Determining when and if an alert should be activated is not straightforward in a dynamic environment like ATC where inherent uncertainty exists. Nuisance alerts are a constant problem, and the consequences of a missed alert can be very serious. Any alerting algorithm may work well in some situations and not in others.

As with tracking, because controllers have used the HCS alerting algorithms for so long, their expectations are strongly influenced by the HCS. When the ERAM algorithms are implemented, there is likely to be some disruption of controllers' work practices with eventual reorientation to the new algorithms. In the following subsections, we propose a test methodology to examine these work practices and expectations and provide information about how to optimize the ERAM algorithms.

3.5.1 Conflict Alerts

One focus of the AMTWG has been developing techniques for assessing the effectiveness of ERAM alerting algorithms. For example, to assess the ERAM conflict alert algorithm (known as Aircraft-to-Aircraft Safety Alerts in ERAM), the AMTWG has developed a set of conflict geometries. A conflict geometry describes the flight paths of a pair of aircraft and measures the separation between the aircraft at various points, as shown in Figure 7. The aircraft pair can be fed to a conflict alert algorithm to assess if and when the algorithm activates a conflict alert. The AMTWG also has developed automated tools for generating conflict geometries, some of which are extremely difficult to create by hand due to the complex and dynamic nature of the conflict.

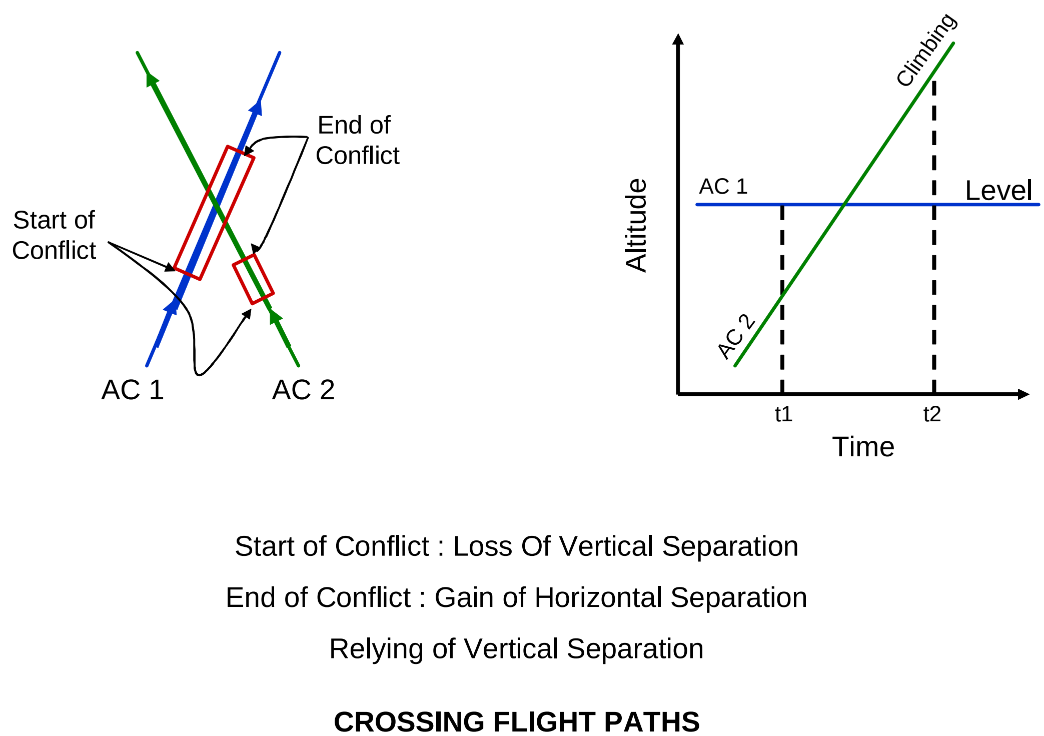

Figure 7. Sample conflict geometry diagram.

However, the evaluations of conflict alert algorithms by the AMTWG do not examine how the algorithms affect controller expectations or performance. For example, it is possible that the ERAM algorithms could activate a conflict alert sooner than the HCS. Controllers could see an early alert as beneficial because the alert is activated with more time to react, or they could see it as a nuisance that draws their attention to a situation that may resolve itself on its own. The ERAM specifications provide requirements for the number of nuisance alerts and the amount of time the alert must activate before the conflict occurs. Though these requirements are reasonable from a system performance and requirement testing perspective, they do not try to optimize the algorithm for specific situations or operational conditions. A more dynamic, user-centered test would help us better understand when controllers expect alerts to occur and when those alerts are nuisances.

3.5.1.1 Proposed Data Collection Method

We propose to include an examination of the effects of the ERAM conflict alert algorithm on controller performance as part of the UCA. We would present controllers with dynamic but noninteractive scenarios. That is, the scenarios would change over time, but the controllers would not be able to change the aircraft trajectories in real time. The scenarios must be noninteractive because controllers naturally try to resolve conflicts before they happen. Changing the geometry even a small amount can dramatically change whether or not there will be a conflict and the characteristics of the situation. For example, in the situation shown in Figure 7, slowing Aircraft 2 down by a small amount could change the geometry from a conflict to a non-conflict.

The scenarios would contain a number of aircraft pairs with the characteristics shown in Table 6. For the test, we would focus on pairs that receive different conflict alerting behavior depending on which system is processing the trajectories.

Table 6. Conditions Comparing Conflict Alerting Algorithms

Actual conflict?	**Alert in Legacy System?**	**Alert in ERAM?**
Yes	Yes	Yes
Yes	No	Yes
Yes	Yes	No
Yes	No	No
No	Yes	Yes
No	No	Yes
No	Yes	No
No	No	No

NOTE: Shaded cells indicate conditions where the legacy system and ERAM provide different alerting behavior.

In a notional methodology, we would present controllers with aircraft pairs and ask them whether a conflict alert is warranted. As the situation develops and it becomes clearer whether a conflict alert should be activated, controllers should become increasingly confident of their judgments. At some point during the scenario, controllers may change their minds in response to aircraft behavior and their own confidence. These data would allow us to examine whether the alerts are activating when controllers expect them. Of particular interest are the situations where the HCS and ERAM reach different conclusions. Is the HCS providing a nuisance alert that ERAM is correcting or is ERAM missing an alert that should be activated?

In cases where the conflict alert does activate for one or both systems, we would ask controllers whether the alert sounded too early, too late, or on time. We could use these data to help optimize parameters for operational deployments. If the ERAM alerts are activating too early and are creating a nuisance, we could adjust the algorithm time parameter. Likewise, if the alerts are activating too late and not allowing sufficient time for controllers to react, we could adjust the algorithm to activate the alert sooner.

Because the ERAM conflict alerting algorithm also will be used during the baseline simulations, the simulations would provide a second forum where we could identify issues with the conflict algorithms. Making predictions about controller performance using these algorithms, however, is not feasible at this time. We do not know where the ERAM conflict alert algorithm produces different alerting outcomes than the HCS. When and if we identify such conditions, we will be able to develop traffic scenarios that include those situations and test controllers' responses. As the AMTWG proceeds in its work and testing of ERAM begins, we hope to identify these areas and develop predictions for how these could affect controller performance.

3.5.1.2 Considerations for Testing

The most difficult problem in running the test described previously is developing enough test cases in each condition. For example, it is possible that there are very few situations where ERAM activates an alert, but the HCS does not. We will need to rely on the work conducted by other members of the AMTWG and the ERAM Test Group to identify these situations and develop traffic scenarios that replicate them. As problematic conflict geometries are identified, we will incorporate them into the UCA test plan.

A second consideration is running enough trials to allow reliable statistical inferences. With a small number of participants, we may need to run a large number of trials to prevent order effects and provide proper counterbalancing. Because controllers quickly learn to identify the same traffic situation when presented multiple times, the scenario development process for this test could be substantial.

A third consideration is how well the participants will be able to separate their own expectations for alerts from expectations created by their long experience with the HCS. Controllers may say they expect an alert simply because that is what the HCS has always done and not because one is actually needed. Additional training or trial runs may be necessary before data collection to ensure that participants are able to make these distinctions.

Finally, controllers are not accustomed to noninteractive scenarios, and controllers normally resolve conflicts before they occur. They do not typically let a conflict "play itself out." As a result, some controllers may never have seen some of the conflict geometries that we could present. Apart from the anxiety that this may cause in the participants, they may have difficulty making judgments about situations they rarely or never encounter in the field.

3.5.1.3 Strategic Conflict Probe

In addition to tactical conflict alerts, ERAM provides conflict probe functionality that replaces URET in the legacy system. As with tactical alerts, the AMTWG is developing methods for assessing the accuracy of the ERAM conflict probe algorithms. The AMTWG or the ERAM

Test Group may identify conflict geometries where the ERAM conflict probe and URET reach substantially different conclusions. If these geometries are identified, we propose to conduct a corresponding user-centered test following the method we proposed for tactical conflict alerts. That is, in the UCA, we would present participants with selected conflict geometries in dynamic, noninteractive scenarios. We would let the situations run until the loss of separation did or did not occur. We would ask participants where and when the conflict probe should have activated to provide the most useful information. In cases where ERAM and URET agree, we would ask whether the alert activated too early, too soon, or on time. The ERAM conflict probe and URET will both be used in the baseline simulations, providing a second opportunity to identify any differences and issues between the systems. Examining conflict probe in these ways carries the same considerations described previously for tactical conflict alerts.

3.5.2 Terrain Alerts

Terrain alerts, also known as minimum safe altitude alerts, occur when an aircraft is projected to be or currently is at an unsafe altitude with regard to the ground. Each ARTCC airspace is divided into polygons, and parameters are set for each that define the minimum safe altitudes within that area.

Because ERAM will use new tracking and alerting algorithms, it is possible that the activation and timing of terrain alerts in ERAM may differ from the legacy system. As with conflict alerts, it would be useful to know when and where controllers expect a terrain alert and compare this to when the legacy system and ERAM actually activate the alert.

3.5.2.1 Proposed Data Collection Method

As for conflict alerts, we propose to include an examination of the effects of the ERAM terrain alert algorithms as part of the UCA. We would present controllers with dynamic but noninteractive scenarios. The scenarios would contain situations with different geometries between the terrain and the aircraft. As for conflict alerts, we would select some situations because the legacy system and ERAM provide a different alerting behavior. We would ask controllers whether the terrain alert activated too early, too late, or on time. We could use these data to help optimize parameters in the ERAM algorithms for operational deployments.

3.5.2.2 Considerations for Testing

The considerations for conducting this test include all those discussed previously for conflict alerts. In addition, some potential participants may be unaccustomed to working terrain alert situations because of the geography of their home ARTCC and the sectors they normally work. For example, a controller who normally works ultra-high sectors may be out of practice making judgments about terrain alerts, especially if the controller works at a facility with few mountains in the en route airspace.

3.5.3 Airspace Alerts

In addition to conflict and terrain alerts, which both exist in the legacy system, ERAM provides a third tactical alert known as airspace alerts. An airspace alert occurs when an aircraft is projected to enter or currently is in protected airspace or active holding airspace. In the legacy system, a similar function is provided as part of URET but with different characteristics and a more strategic purpose.

3.5.3.1 Proposed Data Collection Method

Similar to conflict and terrain alerts, we propose to include an examination of the effects of the ERAM airspace alerts as part of the UCA. Again, controllers would be presented with dynamic but noninteractive scenarios. The scenarios would contain situations with different geometries between the protected airspace and aircraft. Unlike the other alert types, there would be no situations where the legacy system and ERAM disagree because the legacy system does not provide this type of alert. We would ask controllers whether the airspace alert sounded too early, too late, or on time. We could use these data to help optimize parameters in the ERAM algorithms for operational deployments.

Because controllers do not currently have a tactical airspace alert, they probably will not have well-formed expectations for when these alerts should be activated. In some ways, that may be beneficial for the test because the participants should be giving judgments unbiased by previous systems. On the other hand, controller judgments could vary widely, which would bring the utility of the data into question.

In addition to the UCA, we propose that airspace alerts also be examined in the baseline simulations. In our earlier report, we identified two special situations involving protected airspace (Allendoerfer, Pai et al., 2006). In the first, airspace is being protected for a major sporting event but an intruder violates the airspace. Controllers must respond by keeping regular traffic away from the protected airspace and by assisting an intercept by law enforcement or national security aircraft. In the second situation, a moving presidential motorcade is causing arrival restrictions at a nearby airport. Controllers must respond by keeping arrival traffic away from the airport and changing operations in response to changes from the motorcade. Because tactical airspace alerts represent new functionality in ERAM, how controllers handle these two situations in ERAM could be measurably changed. Table 7 provides metrics where we could measure these possible differences.

3.5.3.2 Considerations for Testing

For the UCA, the considerations are similar to those for conflict and terrain alerts. Unlike conflict and terrain alerts, however, ERAM airspace alerts are new functionality. The only experience controllers will have with them is the experience they gained during ERAM training and testing. Their expectations for when the alert should occur may vary widely between controllers and situations.

In the baseline simulations, similar problems due to lack of experience may occur. Controllers will not have used the functionality in the field, and it is possible that controllers may not immediately recognize an airspace alert or may not immediately know how to address the situation. In addition, mock procedures for airspace alerts will be necessary for the simulation.

Finally, because the baseline simulations are interactive, it is possible that controllers will take action early in the simulation to prevent the airspace alerts from ever occurring. We may need to specially script simulation pilot actions to ensure that the airspace alerts occur when and where intended.

Table 7. Potential Differences Resulting from Airspace Alerts

Construct	Metric	Possible Direction & Magnitude of Differences from Legacy System
Safety Risk	Standard metrics	Small decrease
	Number of aircraft penetrating protected airspace	Small decrease
	Timeliness of identification of aircraft approaching protected airspace boundary	Moderate increase
Efficiency	Standard metrics	No measurable change
Workload	Standard metrics	Moderate decrease
	Frequency of air-ground communications	Small decrease because controller can vector aircraft away from protected area before violations occur

4. Discussion

In the previous sections, we have discussed important differences between ERAM and the legacy system that may have measurable effects on how controllers do their jobs. We believe that tests that examine how controllers use both systems are warranted in these cases, following established human factors methodologies. These tests can be accomplished as part of ERAM DS testing to examine the effectiveness of ERAM relative to the legacy system. The tests also can help identify the operational benefits of ERAM. As we have discussed, we propose that this testing occur in two phases: a UCA and several HITL baseline simulations.

4.1 Usage Characteristics Assessment

The UCA will focus on the list of 30 most frequently used commands identified in our first ERAM metrics report (Allendoerfer, Zingale et al., 2006) and the important-but-infrequent commands identified in our second report (Allendoerfer, Pai et al., 2006). The UCA also will present an opportunity to conduct controlled laboratory assessments of specific ERAM changes. In particular, it should include examination of the interactive lists discussed in Section 3.3.1. The UCA should collect information about the time to complete an action, the number of keystrokes or mouse clicks required, the time spent looking at the keyboard or screen, and the error rate. The UCA will most likely be conducted as part of the ERAM Early Operational Evaluation scheduled for mid-2007 at the Integration and Interoperability Facility.

In addition, we propose that the UCA contain an examination of conflict, terrain, and airspace alerts, as discussed in Section 3.5. The alert portion of the UCA would differ in that it would not be concerned with interaction methods but rather with when and how alerts occur. We can construct the scenarios used in the UCA to allow completion of both parts during a single session.

In preparation for the UCA, we will write a test plan that outlines methodological details such as data collection, scenario simulation, and metric collection. We will develop and shakedown the UCA procedures as part of preparations for the activity. As of this writing, the UCA is planned to occur during the ERAM Early Operational Evaluations in early 2007.

4.2 Baseline Simulations

The best method for directly comparing controller usage of the legacy system and ERAM is to conduct baseline simulations on both platforms. In the simulations, we will present controllers with carefully chosen traffic situations and ask them to respond to the situations using both systems. We will calculate the same metrics for both systems and make direct comparisons with a minimum of confounding variables. Interested readers can find discussion of the baseline methodology in the *Air Traffic Control Baseline Methodology Guide* (Allendoerfer & Galushka, 1999) and the reports of baseline simulations conducted for the PVD (Galushka et al., 1995) and the DSR (Allendoerfer et al., 1999).

If the changes in ERAM have direct effects on how controllers do their jobs, these differences should appear in new metrics discussed here. For example, the ERAM AOIs allow facilities full flight plan functionality for some flights in adjacent facilities. The new ERAM capabilities could directly result in a reduction in manual flight plan transfers between facilities, as discussed in section 3.2.

Changes in ERAM also may have indirect results that can be detected in the standard baseline metrics. These metrics include measures of air traffic safety, efficiency, and workload (Allendoerfer & Galushka, 1999). For example, if the ERAM macro capability reduces controller data entry workload, controllers may be able to handle somewhat more traffic, all other parameters equal. Baseline metrics such as the number of aircraft handled or the average time in the sector may show improvements, even though the ERAM change was only intended to directly affect workload.

In preparation for the baseline simulations, we will write a test plan that outlines the situations to be simulated, metrics that will be captured, and other methodological details. The descriptions of the simulated situations will outline requirements for traffic volume and characteristics (e.g., number of aircraft, number of intersecting trajectories) and operational events (e.g., emergencies, outages) that will occur in several scenarios that drive the simulation platform. The simulated situations will allow controllers to exercise all selected functions, and we will design them to elicit latent effects of other ERAM changes, if any. We will develop and shakedown the scenarios as part of preparations for the simulations.

References

Allendoerfer, K. R., & Galushka, J. (1999). *Air traffic control baseline methodology guide* (DOT/FAA/CT-TN99/15). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Allendoerfer, K. R., Galushka, J., & Mogford, R. H. (2000). *Display system replacement baseline research report* (DOT/FAA/CT-TN00/31). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Allendoerfer, K. R., Pai, S., & Zingale, C. (2006). *An analysis of en route air traffic control system usage during special situations* (DOT/FAA/TC-TN06/11). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Allendoerfer, K. R., Zingale, C., Pai, S., & Willems, B. (2006). *En route air traffic controller commands: Frequency of use during routine operations*. (DOT/FAA/TC-TN06/04). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Federal Aviation Administration. (2002). *Blueprint for NAS modernization 2002 update.* Retrieved May 23, 2006 from http://www.faa.gov/nasarchitecture/Blueprint2002.htm

Federal Aviation Administration. (2003, October). *Test and Evaluation Master Plan (TEMP) for en route automation modernization.* Atlantic City International Airport, NJ: Author.

Federal Aviation Administration. (2004). *Progress report of the Automation Metrics Test Working Group (AMTWG).* Atlantic City International Airport, NJ: Author.

Federal Aviation Administration. (2004, October). *National Airspace System en route configuration management document: Computer program functional specifications - message entry and checking* (NAS−MD−311). Atlantic City International Airport, NJ: Author.

Federal Aviation Administration. (2005, June). *En route automation modernization automation metrics and preliminary test implementation plan, version 2.7.* Atlantic City International Airport, NJ: Author.

Galushka, J., Frederick, J., Mogford, R., & Krois, P. (1995). *Plan view display baseline report* (DOT/FAA/CT-TN95/45). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Acronyms

AMTWG	Automation Metrics Test Working Group
AOI	Area of Interest
ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
CID	Computer Identifier
COI	Critical Operational Issues
CPT	Conflict Probe Tool
DS	Display System
DSR	Display System Replacement
EBUS	Enhanced Backup Surveillance
EDARC	Enhanced Direct Access Radar Channel
ERAM	En Route Automation Modernization
FAA	Federal Aviation Administration
FDP	Flight Data Processing
HCS	Host Computer System
HITL	Human In The Loop
M&C	Monitor and Control
MCO	Orlando International Airport
NAS	National Airspace System
OT&E	Operational Test and Evaluation
PVD	Plan View Display
SDP	Surveillance Data Processing
SME	Subject Matter Expert
STARS	Standard Terminal Automation Replacement System
TechOps	Technical Operations
TFM	Traffic Flow Management
TPA	Tampa International Airport
TPM	Technical Performance Measurement
TRACON	Terminal Radar Approach Control
UCA	Usage Characteristics Assessment
UI	User Interface
URET	User Request Evaluation Tool
ZJX	Jacksonville ARTCC
ZMA	Miami ARTCC

Appendix A

Standard Metrics for Human-in-the-Loop Baseline Simulations

The following metrics are adapted from the *Air Traffic Control Baseline Methodology Guide* (Allendoerfer & Galushka, 1999). Formal definitions of these metrics can be found in that document along with advice on successful collection and analysis of these data. The *Methodology Guide* incorporates earlier en route baseline studies (Allendoerfer et al., 2000; Galushka et al., 1995) and other research in metrics of controller performance (Hadley, Guttman, & Stringer, 1999).

Key
R = Radar Controller
D = Data Controller
SME = Subject Matter Expert
ATWIT = Air Traffic Workload Input Technique

Safety Risk

- Number of Operational Errors
- Number of Conflict Alerts
- Number of Halo Initiations
- Descriptions of Other Safety Critical Issues (e.g., reports from participants or SMEs)

Efficiency & Performance

- Number of Aircraft Under Control
- Average Time in Sector
- Number of Altitude, Speed, and Heading Changes
- Post-Run Questionnaire Ratings
  - Quality of ATC services from a controller point of view-R
  - Quality of ATC services from a controller point of view-D
  - Quality of ATC services from a pilot point of view-R
  - Quality of ATC services from a pilot point of view-D
- SME Over-the-Should Rating Form Items
  - Maintaining Safe and Efficient Traffic Flow-R
  - Maintaining Safe and Efficient Traffic Flow-D
  - Maintaining Attention and Situation Awareness-R
  - Maintaining Attention and Situation Awareness-D
  - Prioritizing-R
  - Prioritizing-D
  - Providing Control Information-R
  - Providing Control Information-D
  - Technical Knowledge-R
  - Technical Knowledge-D
  - Communicating-R
  - Communicating-D

Workload/Taskload

- Number of Data Block Offset Actions
- Number of Overall Data Entries- R
- Number of Overall Data Entries-D
- Number of Data Entry Errors-R
- Number of Data Entry Errors-D
- ATWIT Workload-R
- ATWIT Workload-D
- Number of Air-Ground Communications (also called Communication Taskload)
- Number of Ground-Ground Communications (also called Coordination Taskload)
- Post-Run Questionnaire Ratings
    - Post-Run Workload-R
    - Post-Run Workload-D