Comparison of Converted Route Processing by Existing Versus Future En Route Automation

W. Clifton Baldwin, System Engineering & Integration Group,
Federal Aviation Administration

August 2005

DOT/FAA/CT-TN05/29

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report. This document does not constitute FAA certification policy.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN05/29		
4. Title and Subtitle Comparison of Existing Versus Future En Route Automation Processing of Converted Route		5. Report Date August 2005 6. Performing Organization Code System Engineering & Integration
7. Author(s) W. Clifton Baldwin, Federal Aviation Administration		8. Performing Organization Report No. DOT/FAA/CT-TN05/29
9. Performing Organization Name and Address U. S. Department of Transportation Federal Aviation Administration, William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address U. S. Department of Transportation ERAM Program Office Washington, DC 20590		13. Type of Report and Period Covered Technical Note 14. Sponsoring Agency Code ATO-E

15. Supplementary Notes

The author identified above represents the following organizations: W. Clifton Baldwin with FAA System Engineering & Integration (ACB-210)

16. Abstract

The Federal Aviation Administration's (FAA's) En Route Automation Modernization (ERAM) Test Group (ACB-550) formed the Automation Metrics Test Working Group (AMTWG) in 2004. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. This technical note outlines a strategy for testing and analyzing the converted route processes of ERAM in contrast to the Host Computer System (HCS). An overview is provided for the HCS and ERAM converted route process. To achieve the ultimate goals, the User Request Evaluation Tool (URET) is used in lieu of ERAM to test the strategy. The strategy is outlined, which relies on using descriptive and inferential statistical techniques. Data was collected from the Washington Center at Leesburg, Virginia (ZDC) and was prepared appropriately for analysis. Using the sample data from URET and the HCS, the strategy was applied to determine which system has the superior performance. The results of the tests indicate that URET performance exceeds the HCS performance with better than a 99.9% probability. Once ERAM becomes available for testing, the same strategy will be applied to ERAM versus the HCS in order to determine whether ERAM meets its measure of suitability for the converted route processing.

17. Key Words	18. Distribution Statement
Aircraft flight paths Converted Route Lateral Adherence Lateral Deviation HCS, Host Computer System ERAM, En Route Automation Modernization URET, User Request Evaluation Tool	This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405. This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	36	

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

Acknowledgements

I would like to thank Kim May from the ERAM Test Group at the William J. Hughes Technical Center (WJHTC) for her expert input on the operation of the Host Computer System (HCS) and the planned operation of ERAM. Also from the ERAM Test Group, I would like to thank Linda Dolka for her comments and suggestions for improving this paper.

I would like to thank Shurong Liu and Mike Paglione of the Simulation and Analysis Group at the WJHTC. Mr. Liu developed and coded the software tools to process the raw data that is used in this study. Mr. Paglione developed the software tools also, but more importantly he provided feedback on the analysis that was applied in this study.

I would like to acknowledge Randy Phillips, Air Traffic Controller on staff at the Human Factors Group at the WJHTC, who provided insights in HCS processing, and Steve Souder of the WJHTC I^2F who provided the idea on where to find the needed data. Also, I would like to acknowledge Scott Ginsburg, ERAM Air Traffic Representative, for his insights in the plan for this study in the briefing on March 17, 2005.

Finally, I would like to thank the FAA's ERAM Program Office for funding this study via the ERAM Test Group at the WJHTC.

Executive Summary

The Federal Aviation Administration's (FAA's) En Route Automation Modernization (ERAM) Test Group (ACB-550) formed the Automation Metrics Test Working Group (AMTWG) in 2004. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. The targeted system functions are Surveillance Data Processing (SDP), Flight Data Processing (FDP), Conflict Probe Tool (CPT), and the Display System (DS) modules. The metrics are designed to measure the performance of ERAM. They are designed also to measure the performance of the legacy En Route automation systems in operation today. When appropriate, they will allow comparison of similar functionality in ERAM to legacy systems.

The project is divided into key phases: first a metrics identification process was performed. A list of approximately one hundred metrics was generated by the AMTWG and mapped to the Air Traffic services and capabilities found in the Blueprint for the National Airspace System Modernization 2002 Update. Initial metrics results were published in June 2004 in the document titled, "ERAM Automation Metrics Progress Report of the Automation Metrics Test Working Group". Next, an implementation-planning phase was performed. In this step, the identified metrics were prioritized for more detailed refinement during 2005. The plan "ERAM Automation Metrics and Preliminary Test Implementation Plan," documents the implementation-planning phase. It lists these metrics, gives the rational for selecting them, and provides a high level description on how the highest priority metrics will be measured.

The final project phase is the data collection and analysis phase. In this step, AMTWG will document the further refinement and application of these metrics on the current legacy systems in a series of Metric Reports. AMTWG is planning the delivery of four Metric Reports for fiscal year 2005 covering several of the ERAM subsystems. This technical note documents a strategy for testing and analyzing the converted route processes within FDP of ERAM compared to the Host Computer System (HCS). An overview is provided for the HCS and ERAM converted route process.

Flight data was collected over a six-hour period from the Washington Center at Leesburg, Virginia (ZDC). Both the HCS and the User Request Evaluation Tool (URET) processed the converted route data. URET was used in lieu of ERAM since ERAM is not yet available for testing. The processed data was checked for reasonableness, interpolated, and the adherence to the flight plan was computed using legacy software tools developed by the Simulation and Analysis Group for evaluating URET. The primary metric for this study was the lateral distance, which is the closest straight-line distance of the aircraft's track position report projected onto the aircraft's current route segment in nautical miles. The lateral adherence is an indirect metric to the converted route based on the assumption that all positions are static between the two systems except for the converted route.

The strategy begins with some descriptive statistics, such as the central tendency and the dispersion of the dataset of lateral distances. Although informative, the descriptive statistics can only describe the current state of the metrics. In order to make predictions about the performance of the two systems, inferential statistic techniques were applied.

The strategy was tested and can be repeated easily with data from ERAM, once available. As for the results of the analysis, URET demonstrated improved performance over the HCS with better than 99.9% probability for every statistical test attempted.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Table of Contents

List of Figures

List of Equations

1. Introduction

1.1 Purpose

The Federal Aviation Administration's (FAA's) En Route Automation Modernization (ERAM) Test Group (ACB-550) formed the Automation Metrics Test Working Group (AMTWG) in 2004. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. The targeted system functions are Surveillance Data Processing (SDP), Flight Data Processing (FDP), Conflict Probe Tool (CPT), and the Display System (DS) modules. The metrics are designed to measure the performance of ERAM. They are designed also to measure the performance of the legacy En Route automation systems in operation today. When appropriate, they will allow comparison of similar functionality in ERAM to legacy systems (e.g. Host Computer System).

The goal of this paper is to outline a strategy for testing and analyzing the converted route processes of ERAM in comparison to the Host Computer System (HCS). In order to validate the proposed methodology, the User Request Evaluation Tool (URET) will be used as a substitute for ERAM. Using a substitute allows the strategy to be tested before ERAM is available. Since ERAM is using components from URET, this approach will have the added benefit of providing statistics on URET.

1.2 Background

It is the responsibility of the FAA in monitoring the development of ERAM to ensure that the system when delivered meets the needs of the FAA. The charter of the AMTWG is to identify metrics that illustrate the effectiveness of ERAM Release 1 in the areas of SDP, FDP, CPT, and as of January 2005 the DS was added as well. The metrics are designed to measure the performance of ERAM. They also are designed to measure the performance of the legacy En Route automation systems in operation today. When appropriate, they will allow comparison of similar functionality in ERAM to legacy systems (e.g. HCS).

The project was divided into key phases: first a metrics identification process was performed. A list of approximately one hundred metrics was generated by the AMTWG and mapped to the Air Traffic services and capabilities found in the Blueprint for the National Airspace System Modernization 2002 Update (FAA 2002). This took place most of fiscal year 2004 and initial metrics results were published in June 2004 in the document, "ERAM Automation Metrics Progress Report of the Automation Metrics Test Working Group" (WJHTC/ACB-550 2004). Next, an implementation-planning phase was performed. In this step, the identified metrics were prioritized for more detailed refinement during 2005. The plan "ERAM Automation Metrics and Preliminary Test Implementation Plan," documents the implementation-planning phase. It lists these metrics, gives the rational for selecting them, and provides a high level description on how the highest priority metrics will be measured. The Implementation Plan provides the metric's traceability to the basic controller decisions, ERAM Critical Operational Issues (COIs), and the development contractor's technical performance measurements (TPMs). The categories of high priority metrics are: (1) SDP radar tracking, (2) SDP tactical alert processing, (3) FDP flight plan route expansion, (4) FDP aircraft trajectory generation, (5) CPT strategic aircraft-to-aircraft conflict prediction, (6) CPT aircraft-to-airspace conflict prediction, (7) additional system level metrics, and (8) DS human factor and performance metrics.

The final project phase is the data collection and analysis phase. In this step, AMTWG will document the further refinement and application of these metrics on the current legacy systems in a series of Metric Reports. AMTWG is planning the delivery of four Metric Reports for fiscal year 2005 with one covering each of the ERAM modules discussed above, SDP, FDP, CPT, and DS respectively. These reports will be published in multiple drops to provide the ERAM Test Team on-time information. The drops will coincide with the approaches used to implement the metrics. This technical note documents the report implementing metrics on FDP's route conversion process. Due to planned improvements in the functionality of the FDP, converted route processing will be performed differently than the legacy system, HCS.

Due to the differences between the current HCS and future ERAM, there exists the following COI:

> *COI 1.0 – Does ERAM support ATC operations, using current ATC procedures and methods to provide safe, orderly, and expeditious flow of traffic with at least the same effectiveness as the current system?*

This paper describes the strategy for comparing the effectiveness of the upgraded and current systems in the area of converted route processing as a part of answering this COI. Specifically does ERAM system performance meet or exceed that of current En Route automation?

1.3 Scope

The focus of this document is to prepare and test a strategy for addressing the COI related to the converted route processing of ERAM. In order to do so, the HCS's processing of converted routes and the URET's processing of converted route are analyzed. The strategy is outlined and the results of applying the strategy to the HCS and URET are analyzed.

1.4 Document Organization

This technical note begins with the introduction and goals. Section 2 discusses the HCS and the ERAM system regarding how each one processes the converted route. The two systems are briefly compared in order to highlight any major differences. With an understanding of the task and the systems, the method of data collection and data reduction are presented in sections 3 and 4, respectively. A strategy is defined for analyzing the performance of the two systems in section 5. Section 6 presents the results of the analysis. Finally the conclusions are stated in section 7.

2. Overview of the En Route Automation Systems

2.1 Host Computer System

The FDP subsystem is responsible for processing flight data of which route conversion is a component. Route conversion expands the route elements contained in the flight plan into the component fixes making up the route. In the legacy HCS, route conversion applies only to the route elements that fall between the first and last converted fixes for the Center.

After a flight plan passes initial format and logic checks, it is accepted by the HCS and the route is converted. First the HCS determines if Standard Instrument Departure (SID) and Standard Terminal Arrival Routes (STAR) need to be applied. If a SID is filed in the route, the fixes adapted for the SID are inserted into the converted route. If a STAR is filed in the route, is active, and applies to the aircraft, the fixes adapted for the STAR are inserted into the converted route.

Next, the HCS applies preferential routes. A preferential route consists of applying routing indicators to a filed route of flights. Preferential routings apply to departure and arrivals at airports and consists of SIDs, STARs, Preferential Arrival Routes (PAR), Preferential Departure Routes (PDR), and Preferential Departure Arrival Route (PDAR).

PARs are defined to provide rigidly controlled fix postings for flights arriving at specified airports. PDRs are defined for rigidly controlled fix postings for flights departing at specified airports. A combination of PDR and PAR, a PDAR, forms an adapted departure route and arrival route for flights going from one airport to another.

If the flight is a departure, the route is processed for application of a PDR or PDAR. If an active PDR or PDAR for the departure airport is found that applies to the flight, the adapted fixes for the PDR or PDAR are inserted into the converted route. The adapted fixes for a PDAR replace all route elements between the departure point and the destination. The adapted fixes for a PDR replace all route elements between the departure point and the PDR transition fix.

If the flight is an arrival, the route is processed for application of a PAR. If an active PAR for the destination airport is found that applies to the flight, the adapted fixes for the PAR are inserted into the converted route. The adapted fixes for the PAR replace all route elements between the PAR transition fix and the destination.

After the preferential routes, HCS expands the adapted routes. Each adapted route segment in the filed route is expanded into the fixes adapted for that route between the entry and exit fixes of that segment.

Since route conversion is specific to a Center, boundary crossings are inserted. If the flight crosses the Center-boundary inbound from an adjacent facility, the inbound crossing point is inserted into the converted route. If the flight crosses the Center boundary outbound to an adjacent facility, the outbound crossing point is inserted into the converted route. If a flight crosses an Air Defense Identification Zone (ADIZ) boundary inbound to United States airspace, the ADIZ crossing point is inserted into the converted route.

Finally, HCS inserts implied fixes. Route junctions, route segment entry fixes, and route segment exit fixes are inserted into the converted route as implied fixes if they are not explicitly filed in the route. See Figure 1 for a flowchart of this process[1].

[1] See Computer Sciences Corporation, (2001) as source information.

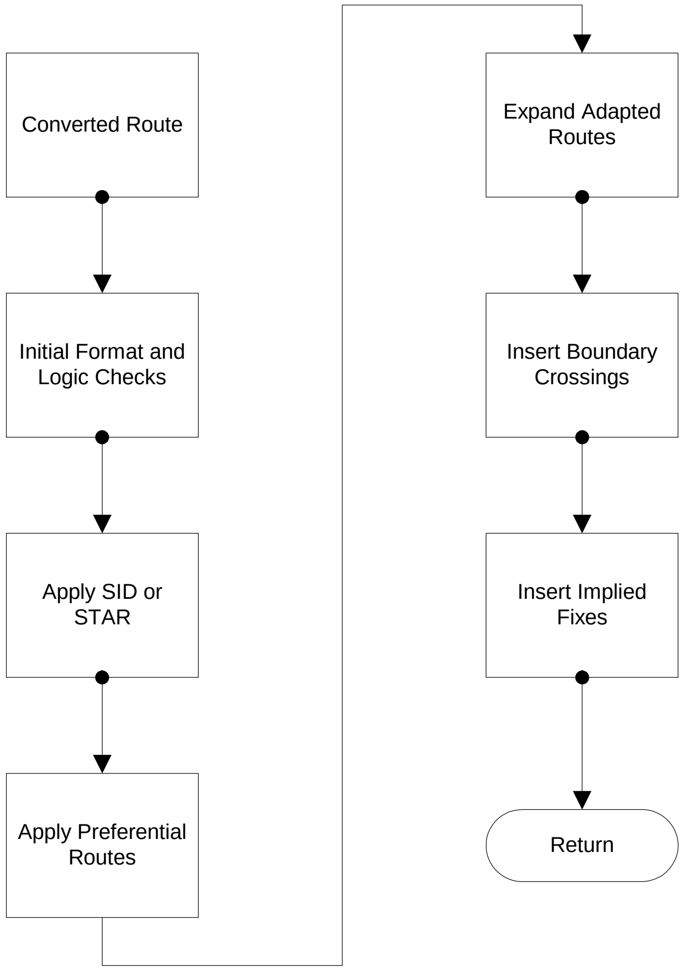


Figure 1: HCS Route Conversion Process.

2.2 En Route Automation Modernization System

Similar to the legacy system, the En Route Automation Modernization (ERAM) system includes the subsystem of Flight Data Processing (FDP), which provides route conversion services. This route conversion function includes conversion of a flight plan route in system plan route string format into a series of fixes, automatic application of adapted routes (if any), and derivation of where a flight returns to its route in support of lateral trajectory re-adherence. Route conversion is performed for the entire route of flight that is within U.S. domestic airspace. System plan route string format is also called the NAS flight plan format, and it is the format used for routes stored with the system plan using a "fix-route-fix" sequence separated by periods[2].

ERAM provides end-to-end route conversion. Each end-point of a direct route segment can be located outside the airspace controlled by the local Air Route Traffic Control Center (ARTCC or Center). All route, fix, ATC Designated Route, and Restriction data is available to each Center. The majority of this shared data resides in national adaptation products gathered from each owning site and distributed from the WJHTC. Some of this data will be dynamically shared in

[2] See Lockheed Martin Corporation (2004b) as source information.

those rare events where an ATC Designated Route needs to be distributed outside of the normal adaptation distribution cycle.

Usually if the flight-plan route departure point is within domestic U.S. airspace, the departure point will be used as the start point for the converted flight plan route. Similarly in most cases if the flight plan route departure point is outside domestic U.S. airspace, the first fix in the trimmed route will be used as the start point for the flight plan converted route.

If the destination point is inside domestic U.S. airspace, the destination point will be used as the final point in the converted route. If the destination point is outside domestic U.S. airspace, the last fix in the trimmed route will be used as the final point in the converted route.

When route conversion commences, route logic checks are performed and the route is parsed into fix and route elements. The intended flight rule transitions are determined and connection of elements is verified. Intersections will be calculated if the adapted sequence of significant points does not share common intersection points.

Next the adaptation data is used for each element to expand the route into a sequence of significant points. Calculated airway and radial route connections are inserted into the route.

Before autoroute processing, the cleared flight plan route is used to determine the list of applicable ATC Preferred Routes (APRs). An APR candidate list will only be created and stored. Later processing will allow the controller to select and apply a specific APR.

When appropriate, an autoroute is chosen. Autoroutes selected get displayed via highlighting along with the original flight plan route. Otherwise the flight plan route on the filed route is kept.

The state data is stored, which is associated with the flight plan converted route such as the fix index, along route distance, last significant point, and indicators of successful route conversion and the route is complete. If the along route distance does not pass final checks, then the route could be rejected.

Finally, the list of restrictions for the flight plan route are determined and stored in the flight plan converted route without regard to the activity status of the restriction. Trajectory modeling will later consider the activation status of these restrictions when building the flight plan trajectory. See Figure 2 for a flowchart of this process[3].

[3] See Lockheed Martin Corporation (2004) as source information.

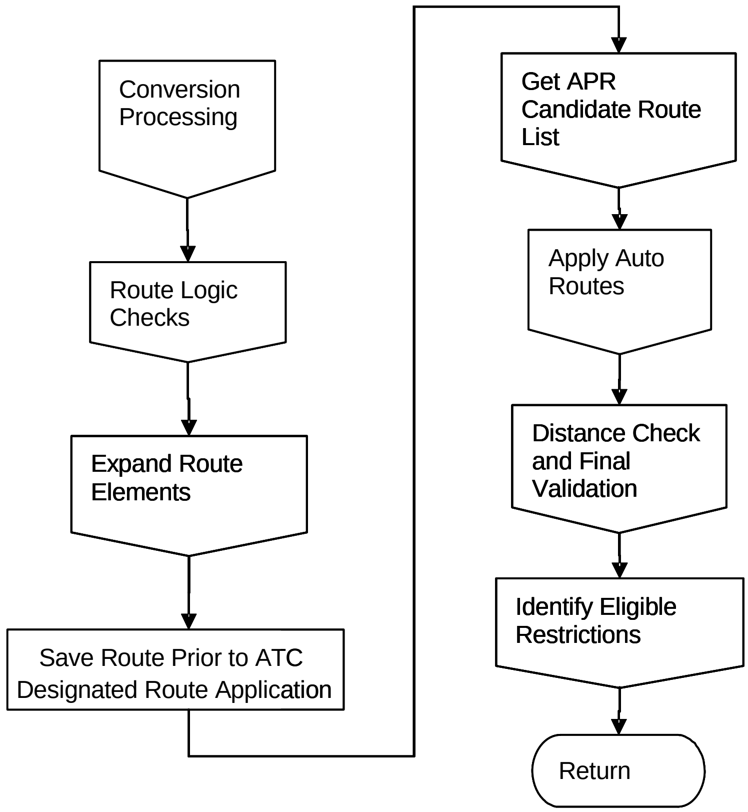


Figure 2: ERAM Route Conversion Process.

2.3 Comparison of Systems

From the flowcharts of the HCS (Figure 1) and ERAM systems (Figure 2), it is apparent that there are some minor differences in the order of the processes. Most of these differences should have no impact on the results other than possibly a timing effect. The only process that will cause a measurable change in the two systems is the process of boundary crossing insertion for the HCS.

A major design change for the ERAM system is the conversion of end-to-end processing instead of each Center. Therefore ERAM has no need for boundary crossings since it does not cross boundaries by design. In the case where a flight crosses into U.S. airspace, the point of entry is not considered a boundary crossing into U.S. airspace but rather the departure point of the flight.

Based on interviews with Tech Center personnel, the HCS uses an objective means to determine the origination point and termination point for routes spanning Centers. Hand-off begins to occur before the accepting Center takes a flight. Each Center has predefined fixed lines where a flight plan originates and terminates. This fixed line is usually 20 nautical miles before the Center boundary, but may differ based on agreements between adjacent Centers.

Nonetheless the software algorithms to perform the processing are being rewritten for ERAM. Therefore there may be significant differences in the results of ERAM versus the HCS regardless that they follow similar processes.

3. Data Collection

Prior to the implementation of any testing strategy, sample data was assembled for analysis. Actual flight data was collected for approximately six hours from the Washington Center at Leesburg, Virginia (ZDC) starting on March 17, 2005 around 18:00. Within that time frame, 2,610 flights were recorded for a total of 577,760 track reports. The minimum number of track reports for a flight was 2 while the maximum was 990.

4. Data Reduction

Since the HCS is the current automation that processes the converted routes and is currently available, the data collected from the ZDC Center was run through the HCS. Legacy software tools developed by the Simulation and Analysis Group for evaluating URET have been applied by the AMTWG to check for reasonableness of the collected data, interpolate it, and compute adherence to the flight plan. This check requires design thresholds including definitions on what constitutes a reasonable position report and how much of a gap in time and space can be repaired with interpolation[4]. From the processed data, a lateral adherence metric is chosen as the closest straight-line distance of the aircraft's track position report projected onto the aircraft's current route segment in nautical miles[5]. This metric is computed from the aircraft's track position and two point positions of the cleared route of flight using vector multiplication. The vector multiplication produces a lateral deviation with a negative or positive value indicating deviation left or right of the current route segment, respectively. This signed lateral deviation value contributes information about the dispersion of the data but does not contribute much else towards the goal of the test. Therefore in this study the lateral distance is the primary lateral adherence metric, which is computed by taking the absolute value of the signed lateral deviation[6]. The resulting absolute value is the lateral distance in nautical miles between the cleared route of flight to the current track position (see Figure 3). The resulting data is stored in a relational database.

Again with the objective of this study to compare a pair of system's converted route processing, the lateral adherence is chosen as an indirect metric for this comparison. For this metric to be effective, both systems are input the same surveillance radar data. As a consequence, the difference between same aircraft's lateral adherences is the metric that allows direct comparison of the two systems. For this lateral adherence difference, the common position becomes subtracted out, resulting in a difference between the two converted routes. The comparison will be explained in more detail in Section 5.2, which presents statistical methods that quantify a level of confidence when contrasting the lateral adherences.

[4] See Paglione et al., (2000) section 3.5 for a more complete description of computing the lateral deviation metric.
[5] A nautical mile is 1,852 meters, or 1.852 kilometers. In the English measurement system, a nautical mile is 1.1508 miles, or 6,076 feet.
[6] See Oaks & Paglione, (2005) for more information on determining lateral adherence.

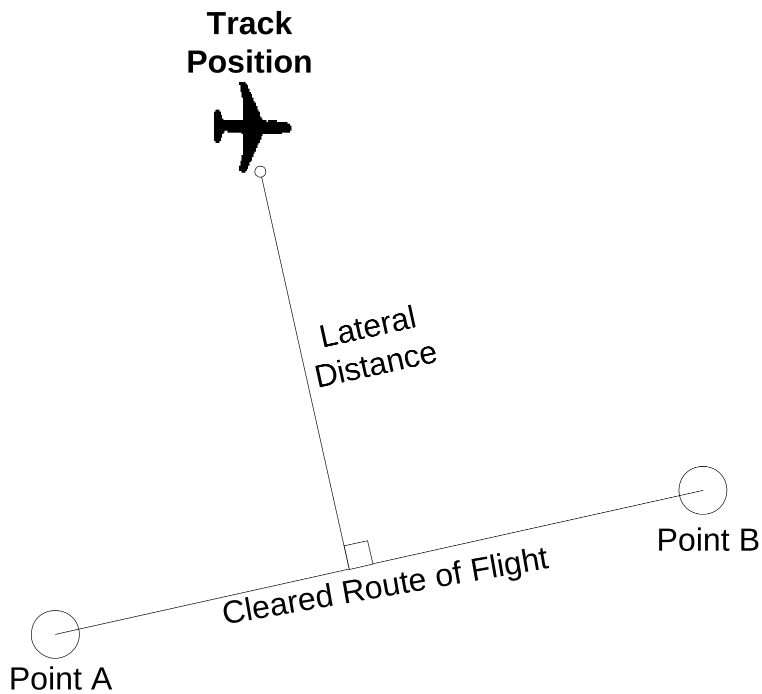


Figure 3: Lateral Distance

Although the measure of suitability addresses the ERAM system, initially the ZDC data will be run through URET as a substitute for ERAM[7]. Even though the results from a comparison against URET will not have a direct impact on the final testing of ERAM, this approach will provide an opportunity to determine the statistics necessary to make a good comparison of the two systems. Once ERAM is available to the William J Hughes Technical Center (WJHTC), ERAM will process the converted routes using the same raw data as the HCS, and the toolsets and statistical analyses will be ready to perform the actual tests. The same AMTWG toolsets that were applied to the HCS have been applied to URET to check for reasonableness, interpolate the data, and compute the lateral distances. Again the results were stored in a relational database.

The converted route process sub-team of the AMTWG developed further tools to extract the sample datasets from the relational database tables. For any flight with at least 30 track reports, the arithmetic mean of lateral distances for each flight are computed for the HCS and then URET. Of all the flights, 46 were eliminated having less than 30 track reports. The remaining 2,564 flights for both the HCS and URET were formatted into comma-delimited text files and transferred from a Unix server to a local Windows 2000 personal computer. Using the commercial statistical software package SPSS for Windows (version 10.0.7), the comma-delimited files were imported and merged by flight.

The resulting variables are the flight number and both the lateral distance and the signed lateral deviation for each flight. Since the flight metrics are averaged, there is only one observation per flight number. The signed lateral deviations are useful to determine the standard deviation of the

[7] Due to time constraints, AMTWG used the MITRE Prototype of URET (Version D34-R6-P4) for this study. To confirm the findings, AMTWG recently checked the results against the fielded production version of URET. The results produced the same conclusions. For more details, see the Appendix of this technical note.

adherence metric, but the mean is of minimal value as the values should be approximately symmetric around zero. Hence the lateral distance variables are used for all other statistics regarding the performance of each system.

5. Data Strategy

An analysis strategy can be applied to process the results. The proposed strategy uses both descriptive and inferential statistics. Although the strategy discusses the ERAM system, the URET is the system used in section 6. URET is substituted for ERAM in order to test the analysis strategy before ERAM is available for such testing.

The data from both systems will be analyzed individually using descriptive statistics, and then the comparison of the two systems will be performed using inferential statistics. The inferential statistics strategy addresses whether the ERAM flight data processing of converted routes meets or exceeds the current automation of the HCS.

5.1 Descriptive Statistics

Prior to addressing a full analysis, a random raw sample of 50 flights will be chosen. The sample data will be observed for outliers, including the distribution of outliers, and the mean and standard deviation of the lateral distances. The collected information may be useful as the study progresses.

In its raw form, the data is by flight by track report with a potentially different set of points in time for each flight. In order to make the data more applicable to the inferential statistical techniques, the mean lateral distance of each flight is computed resulting in one record per flight. By taking the mean average of each flight, the Central Limit Theorem (CLT) can be applied to obtain an approximately normal dataset[8]. The CLT states that the shape of the sampling distribution of the mean will approximate a normal distribution if the sample size is sufficiently large, where sufficiently large is usually greater than 30.

Furthermore the CLT states that the mean of the sampling distribution approximates the mean of the population. Therefore the mean of the mean lateral distances will reveal the arithmetic average deviation for each system. Although a random "raw" sample of flights should provide similar descriptive statistics, the unknown distribution of the data limits the number of applicable statistical tests. By applying the CLT, the distribution of means should approximate a normal distribution, which greatly increases the power of the statistical tests. If the resulting distribution of means is symmetric as in a normal distribution, the skewness should be close to 0.

Another result of the CLT is that the standard deviation of the population is approximated by the estimated standard error of the distribution of means (see Equation 3). Due to the very large sample size for this study, the standard error may not be an accurate estimate. Nonetheless the standard deviation of the sample means of the signed lateral deviations will provide some information about the dispersion of the means.

In order to reduce the impact of outliers, any outlying lateral distance will be excluded from the calculations. Eliminating extreme cases from the computation of the mean generates a better estimate of central tendency. The threshold for the outliers will be based on the data

[8] See Naiman et al., (1995) for an overview of the Central Limit Theorem

observations. When it appears that the observations of lateral distances are no longer increasing smoothly and become sporadic, the threshold will be determined. Although it would be pleasing to eliminate any outlier greater than 5 nm or so, the threshold must be supported by the actual data.

Another method to reduce the impact of outliers is to examine the median of the sample. By definition, 50% of the values will be less than the median value. A high mean with a low median will indicate that most lateral distances are low but a relatively few outliers increase the mean.

Since the measure of suitability does not require ERAM to be perfect, there is no need to test ERAM against the ideal outcome of mean deviation zero. This fact is another reason why only the most extreme outliers were eliminated. If both the HCS and ERAM result in a large lateral distance, ERAM may still be an improvement. Nonetheless the most important descriptive statistics should be the mean of the lateral distances for each system and the median. In theory, the system with better accuracy should be represented by the mean and median.

5.2 Inferential Statistics

The second part of the strategy is the real test where the samples from the two systems are compared. In order to verify the results, two different hypothesis tests will be conducted using a nonparametric test and a parametric test. The decision to use both tests resulted from peculiarities in the data. Although the CLT should prepare the data for the parametric test, the nonparametric test will be used as a verification of the results, assuming the results are equivalent.

5.2.1 Nonparametric Test

The Wilcoxon signed-rank test is a nonparametric alternative to a paired samples t test when the data distribution is not sufficiently normalized to support a t test[9]. The absolute differences between the variables of two paired systems are ranked and the ranks are split into three groups. For System A – System B, the negative ranks category contains those cases for which the value of System B surpasses the value of System A, while the positive ranks category contains the opposite. The third group contains any ties. Within this study, the desired outcome is that the lateral distances of the HCS surpass those for ERAM since a smaller lateral distance indicates better performance. The Wilcoxon signed-rank test detects differences in the distributions of the two related variables[10]. Set up a hypothesis test with System A representing the data distribution from the HCS and similarly System B for ERAM as follows:

H_0: System A = System B

H_1: System A ≠ System B

To test the hypothesis, first compute the sum of the ranks for the less frequent sign and reference this value as the Wilcoxon test statistic, W. Next standardize W and compare the standardized W, Z_W, against the normal distribution.

To standardize W, the expected value of W, E(W), is computed using Equation 1.

---

[9] See Witte (1993) for an overview of the Wilcoxon Signed-Rank Test

[10] See Mendenhall et al., (1990) for a detailed description of the Wilcoxon Signed-Rank Test formulas and computations

$$E(W) = \frac{n(n+1)}{4}$$ **Equation 1: Expected value of the Wilcoxon test statistic**

where n is the number of data pairs.

The variance of W is needed also for the standardization process and is computed using Equation 2.

$$Var(W) = \frac{n(n+1)(2n+1)}{24}$$ **Equation 2: Variance of Wilcoxon test statistic**

Finally, the standardized Wilcoxon test statistic, Z_W, is computed using Equation 3.

$$Z_W = \frac{W - E(W)}{\sqrt{Var(W)}}$$ **Equation 3: Standardized Wilcoxon test statistic**

The hypothesis test is to reject H_0 in favor of H_1 when $Z_W < z_{\alpha/2}$ or $Z_W > z_{1-\alpha/2}$, where z is the standard normal probability for significance α. Small significance values (less than 0.05) indicate that the two variables differ in distribution. If the two variables are significantly different, then the categories will indicate which system demonstrates better performance.

5.2.2 Parametric Test

Although informative, the Wilcoxon signed-rank test is not a very strong test because it is nonparametric. Since both systems will process the converted routes using the same original data, each mean observation in one sample is paired on a one-to-one basis with a single mean observation in the other sample. This situation is known as two dependent samples, which is preferred in this case to independent samples, as the influence of other variables should cancel out. The paired-samples t test procedure compares the means of two variables that represent the same group at different times[11]. The paired-samples t test is a strong test, but it does require data that approximates a normal distribution. In theory, the data should approximate a normal distribution as long as the raw data samples are independent and identically distributed, which they appear to be. It should be pointed out that the dependency of the two samples is different than the independent samples within each population. The dependency is between the HCS sample and the ERAM (or URET as a stand in) sample, and not within each system's samples.

As discussed in the descriptive statistics section of the strategy, the mean of each flight will be computed in order to approximate normal data. For each flight, the lateral distances arithmetic mean will be computed. The means from each system will be matched up and the difference will be taken resulting in one set of difference scores, as in Equation 4.

[11] See Witte (1993) for an overview of the Paired-Samples t Test

$$D_i = X_i - Y_i$$ **Equation 4: Difference of Paired Samples**

where D_i is the difference score for each observation, X_i is the arithmetic mean lateral distance for each flight from the HCS, and Y_i is the arithmetic mean lateral distance for each flight from the ERAM system.

Next the arithmetic mean of the difference scores (D_i) will be computed, resulting in a single statistic $\overline{D}$. This statistic will be useful for hypothesis testing.

Using the mean of the difference scores, there are two outcomes that will be acceptable. The original goal for ERAM is that it meets or exceeds the HCS. For ERAM to exceed the HCS, the standardized mean of the difference scores, μ_D, must be greater than zero. To understand this reasoning, the original lateral distance observations from HCS would be greater than the observations for ERAM if ERAM surpasses the HCS. Therefore the difference of each observation would be a positive number. The standardized mean of the difference scores would be representative of the population indicating ERAM surpasses the HCS when the statistic is greater than zero.

The hypothesis test for this result is set up using the following equivalent symbolic statements:

H_0: $\mu_D \leq 0$

H_1: $\mu_D > 0$

H_0 and H_1 represent their common meanings of null hypothesis and alternative hypothesis, respectively, and as previously stated, μ_D is the standardized mean of the difference scores.

In order to test the null hypothesis, the t ratio for two population means must be computed to standardize the mean of the difference scores. The formula for the t ratio, in this case, is defined in Equation 5.

$$t_{ratio} = \mu_D = \frac{\overline{D} - \mu_{D_{hyp}}}{S_{\overline{D}}}$$ **Equation 5: t ratio for Paired T Test**

where $\overline{D}$ represents the sample mean of the difference scores, $\mu_{D_{hyp}}$ represents the hypothesized mean (of zero) for all difference scores in the population, and $S_{\overline{D}}$ represents the estimated standard error of the mean of the difference scores. The estimated standard error of the mean of the difference scores is defined as in Equation 6.

$$S_{\overline{D}} = \frac{S_D}{\sqrt{n}}$$ **Equation 6: Estimated standard error of the mean of the difference scores**

where S_D represents the sample standard deviation of the observed difference scores and n equals the number of difference scores.

Comparing the t sampling distribution with n-1 degrees of freedom versus the t ratio computed in Equation 5 allows one to reject or fail to reject the null hypothesis. A low significance value for the t test (typically less than 0.05) indicates that there is a significant difference between the two variables. When the confidence interval for the mean difference does not contain zero, the result provides additional support that the difference is significant.

If the null hypothesis can be rejected, the outcome will be a strong decision. Unfortunately failing to reject the null hypothesis does not lead to any conclusion. Therefore more statistical testing would be needed if that case occurs. If the difference scores are statistically the same, in which case μ_D is close to zero, then ERAM does meet the HCS performance. The case of the HCS exceeding ERAM needs to be ruled out. Therefore a hypothesis test would be performed under the following conditions:

H_0: $\mu_D = 0$

H_1: $\mu_D < 0$

where the symbols are defined as before. Although positive results of this hypothesis test would be a weak decision, rejecting the hypothesis would provide a definitive yet undesirable conclusion that the HCS performance probably exceeds ERAM.

6. Data Analysis

6.1 Descriptive Statistics

A sample of all track reports for 50 flights was randomly chosen resulting in 11,104 track reports. This sample will be referenced as the “raw sample” from this point forward. No observations were dropped, including any values that may be considered outliers. By chance, none of the sample flights had less than 30 track reports. This raw sample set was analyzed in respect to the HCS and the URET.

In order to strengthen the statistical tests, the Central Limit Theorem was applied to normalize the data. The mean lateral distance for each flight was computed resulting in one record per flight. This sample dataset of the means will be referenced as the “mean sample” from this point forward. The CLT states that the shape of the sampling distribution of the mean will approximate a normal curve if the sample size is sufficiently large, where sufficiently large is usually greater than 30. In the ZDC mean sample dataset alone, there were 2,610 flights of which 2,564 had more than 30 track positions. Since there were over 2500 usable flights, it is safe to assume that the sample size was sufficiently large.

As stated in the data strategy section, eliminating any flights with unusually extreme mean values refined the data further. The absolute value of 19 nautical miles was chosen for the signed lateral deviations based on observations in the data. It appears that the observations of signed lateral deviations smoothly approach the 19 nm threshold and then become sporadic. This threshold eliminated 28 flights from the mean sample dataset. Although it would be desirable for no data points greater than 5 nm or so, the data does not support this inclination. It may be beneficial for a future study to determine why the lateral distances approach these values greater than 5 nm, but it is out of scope for this study. If either the HCS or the URET exceeded the threshold, the entire flight was eliminated. Only 28 flights, or approximately 1%, met this criterion, and the resulting mean sample dataset had 2,536 observations. Eliminating the 1% extreme values is well within common statistical practices.

6.1.1 Host Computer System

A perfect prediction for either system would have a lateral distance of zero. The arithmetic mean of the lateral distances from the raw sample of 50 flights is 2.97 nm for the HCS.

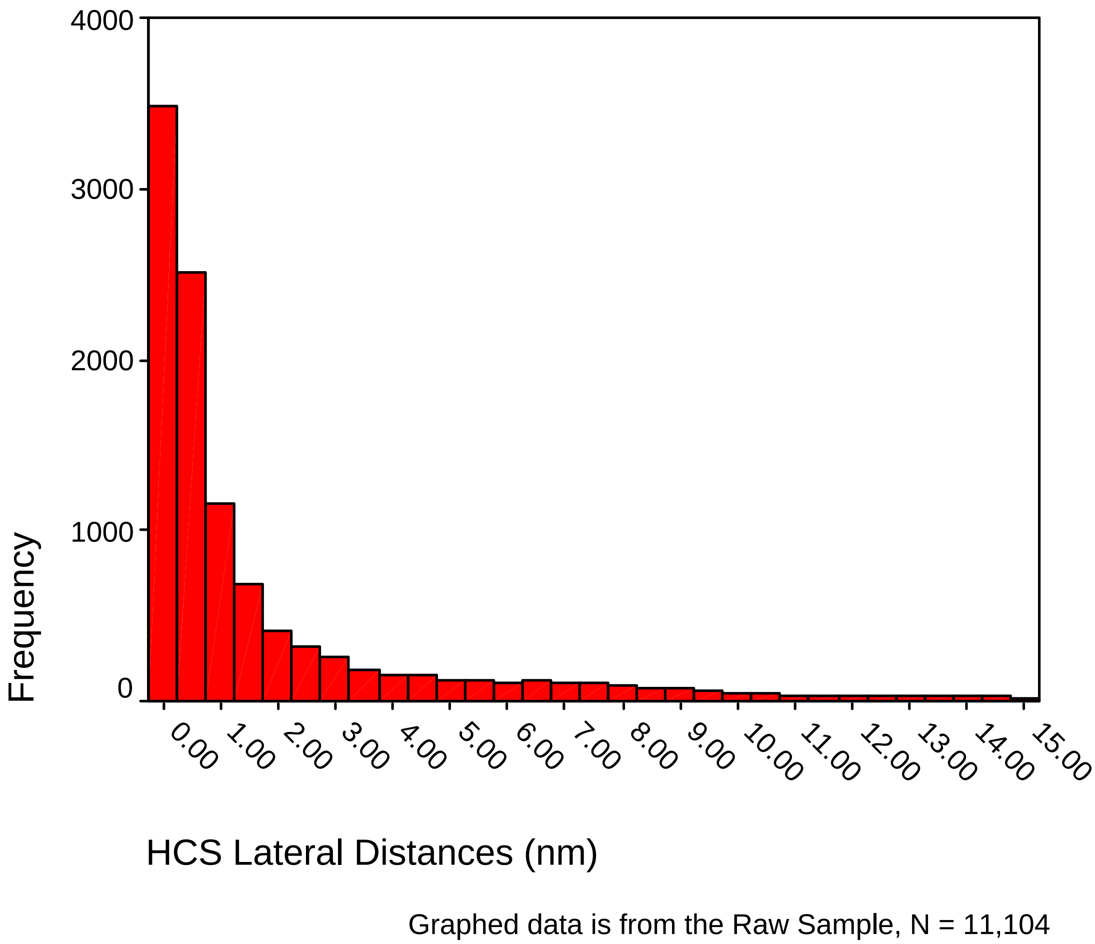


Figure 4: Histogram of a Raw Sample of HCS Lateral Distances

For readability, the axis of the histogram in Figure 4 goes only to 15 nm. The histogram shows the lateral distances from the HCS of the raw sample across the horizontal axis and the flight count (or frequency) on the vertical axis. Somewhat apparent by this graph, most of the observations are close to zero. The mean is almost 3 nm due to the extreme values, most of which are rare above 10 nm. The median supports this conclusion with a value of 0.62 nm.

By taking the arithmetic mean of the lateral distance for each flight and then taking the arithmetic mean of each flight's arithmetic mean, the average deviation of a flight's track position to the HCS processed flight route is 3.4 nm. According to the CLT, this mean should represent the parent population. Furthermore the mean and median should equal, but the median is quite smaller than the mean. The median is 2.5 nm, which indicates that 50% of the means are 2.5 nm or less but does not represent the median of the raw data. Although the CLT will normalize the signed lateral deviation data (see Figure 6), the "unsigned" lateral distances are not totally bell shaped probably due to the loss of the negative numbers. When the "negative numbers" are added to the "positive number" count, the bell curve becomes skewed with a skewness of 1.8 for the HCS. The histograms in Figure 5 and Figure 8 illustrate this skewing of the distributions for both systems.

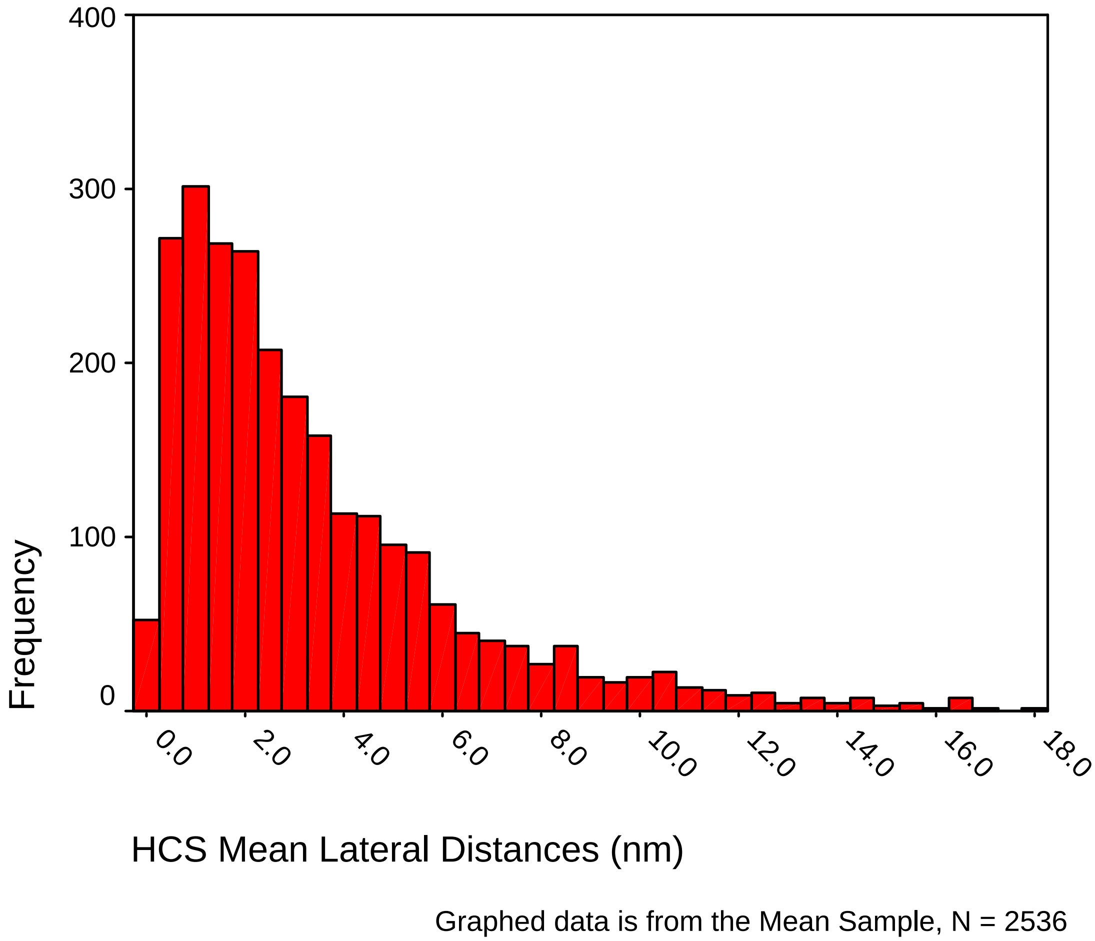


Figure 5: Histogram of HCS Mean Lateral Distances

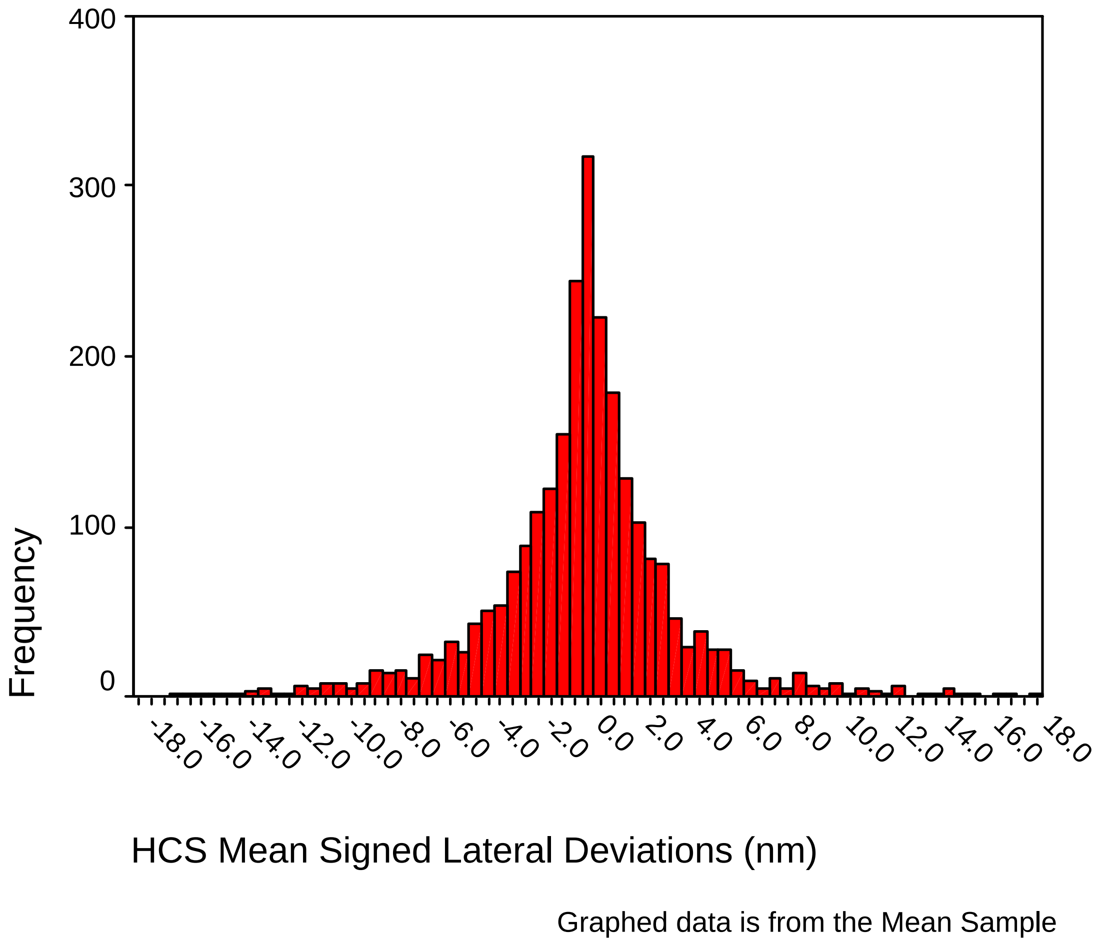


Figure 6: Histogram of HCS Mean Signed Lateral Deviations

As illustrated in Figure 6, the mean sample of signed lateral deviations approximates a bell curve with a skewness of 0.04. Since the mean and median are close to 0 nm, the averages are not very indicative of the data. The standard error, which should approximate the standard deviation of the parent population, is 0.07, but this statistic appears to be a poor estimation. The standard deviation of the mean sample is 3.7 nm, which at least indicates the dispersion of the sample means.

The different analysis techniques can be collected to try to get a picture of the descriptive statistics. The signed lateral deviation provides the standard deviation of the means of 3.7. The lateral distance from the mean sample provides the mean of 3.4 nm. The raw sample of 50 flights has a median of 0.62 nm, which indicates the mean is affected by extreme values.

6.1.2 User Request Evaluation Tool

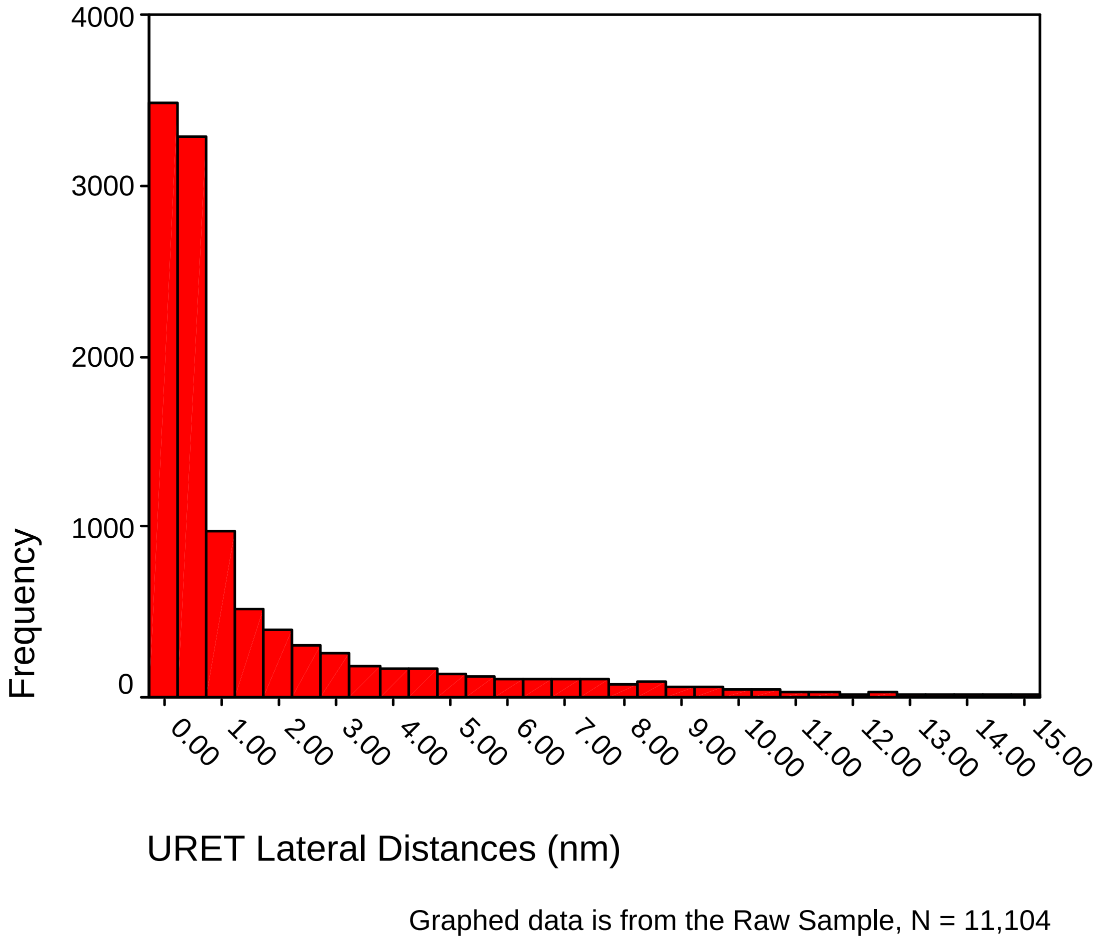
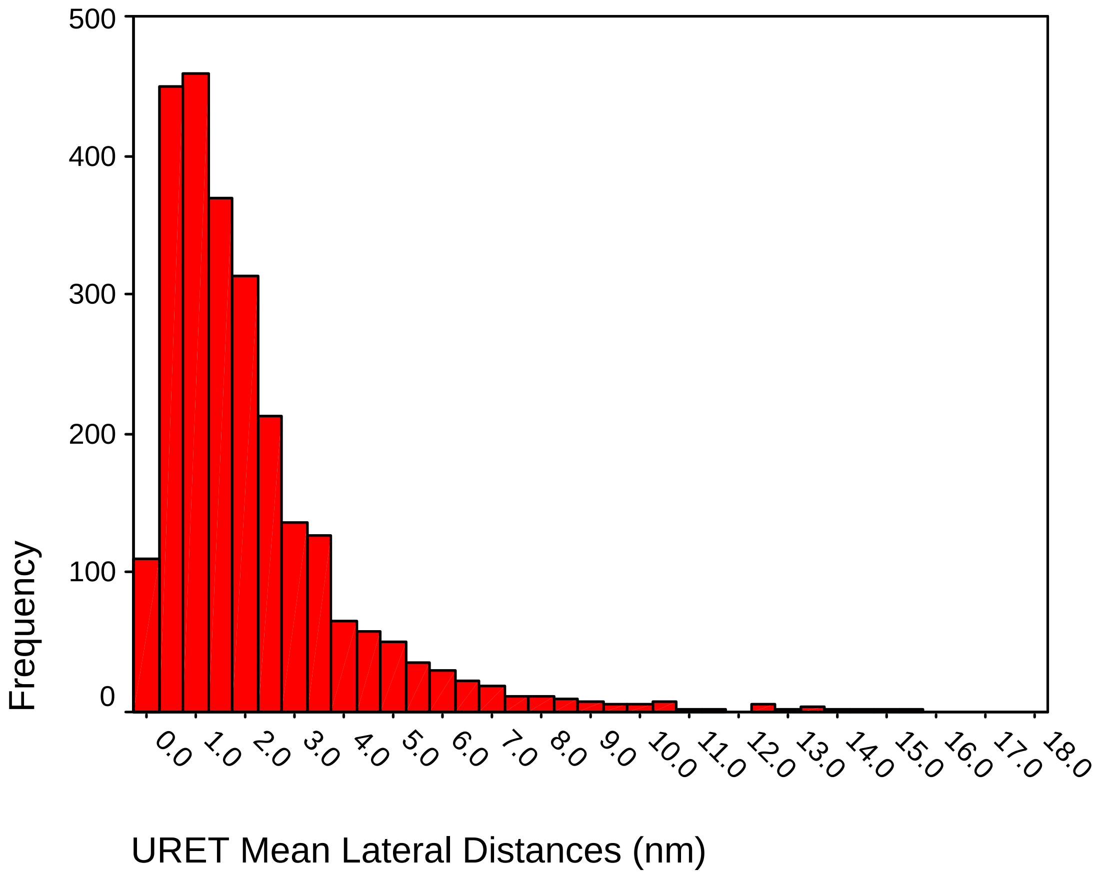


Figure 7: Histogram of a Raw Sample of URET Lateral Distances

As shown by the histogram in Figure 7, URET has a similar situation to the HCS, although all the statistics do improve. The raw sample's arithmetic mean of the lateral distances is 1.89 nm for the URET. The median lateral distance for URET is 0.50 nm, indicating that 50% of the lateral distances from this raw sample are less than a half-mile. Unfortunately as also shown in Figure 7, values out at least to 15 nm occur often enough to be displayed in the histogram.

By taking the arithmetic mean of the lateral distance for each flight and then taking the arithmetic mean of each flight's arithmetic mean, the average deviation of a flight's track position to the URET processed flight route is 2.2 nm.



Graphed data from the Mean Sample, N = 2536

Figure 8: Histogram of URET Mean Lateral Distances

As illustrated in the histogram in Figure 8, there are not as many obvious data points past 10 nm. An observation of the mean sample data does show some outliers but not as many as the HCS results.

Again similar to the HCS case, the mean sample distribution remains skewed for the lateral distances with a skewness of 2.7. Comparable to the HCS results, the absolute values of the lateral distances appear to upset the curve of the distribution somewhat by folding the “negative numbers” onto the positive side of the curve.

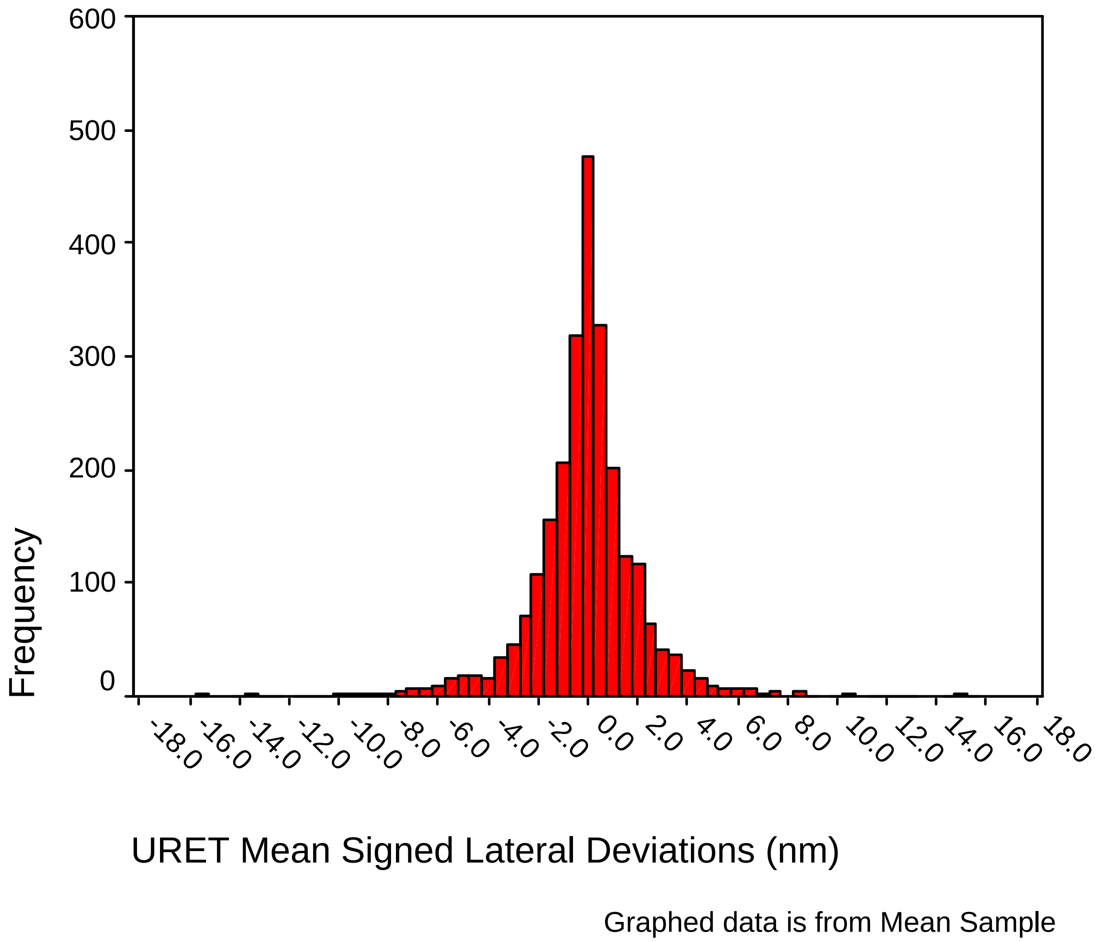


Figure 9: Histogram of URET Mean Signed Lateral Deviations

The mean sample of signed lateral deviations, as shown in Figure 9, approximates a bell curve with a skewness of –0.16. Since the mean and median are close to 0 nm, the averages are not very indicative of the data. The standard error, which should approximate the standard deviation of the parent population, is 0.05, but this statistic appears to be a poor estimation. The standard deviation of the mean sample is 2.4 nm, which at least indicates the dispersion of the sample means.

Compiling the statistics gives a median of 0.50 nm from the raw sample, a mean lateral distance of 2.2 nm from the mean sample, and a standard deviation of the mean signed lateral deviations of 2.4.

6.1.3 Descriptive Statistics Analysis

The raw sample's median of the lateral distances, which is probably the better estimate of the average, is 0.62 nm for the HCS and 0.50 nm for URET. These findings translate into a 20% improvement for the URET. When the CLT is applied, the mean of the mean sample is 3.4 nm for the HCS and 2.2 nm for URET. The standard deviations of the mean of signed lateral deviations contribute to the finding that URET performs better than the HCS.

6.2 Inferential Statistics of HCS versus URET

6.2.1 Nonparametric Test

Although the CLT will approximate the normal distribution for large samples, we have seen that the absolute values of the lateral distances skew the curve somewhat. Therefore, a nonparametric test was performed in order to eliminate any uncertainty based on the assumption of normality. The mean sample was used for the Wilcoxon signed-rank test, and the ranks are displayed in Figure 10.

HCS – URET	N	Sum of Ranks
Negative Ranks	758[a]	670,173
Positive Ranks	1778[b]	2,546,743
Ties	0[c]	
Total	2536	
	a. HCS < URET b. HCS > URET c. HCS = URET	

Figure 10: Wilcoxon Signed-Rank Test Ranks

By standardizing the sum of the ranks for the less frequent sign, in this case the negative ranks, differences in the distributions of the two variables can be tested. The standard z statistic at the 0.001 significance level of a 2-tailed test indicates that the null hypothesis should be rejected if $Z_W < -3.30$ or $Z_W > 3.30$. The significance of the test indicates the probability of the results.

	HCS - URET
Z_W	-25.44[a]
Significance, α (2-tailed)	0.00[b]
	a. Based on negative ranks, where HCS < URET b. Approximately zero

Figure 11: Wilcoxon Signed-Rank Test Results

Since $Z_W = -25.44 < -3.30$, the null hypothesis that HCS = URET is rejected. Actually the significance value for the results, as shown in Figure 11, is almost nonexistent. This result indicates that the two variable differ in value with an almost certain probability. Since URET outperforms the HCS 1,778 times, and the HCS outperforms URET only 758 times, it is clear from this test that URET performance exceeds the performance of the HCS by a factor of approximately 2.3.

In order to see the result visually, the scatter plot in Figure 12 was graphed using the mean sample data. The HCS lateral distances from the mean sample are on the y-axis or vertical axis, and the URET lateral distances are plotted on the x-axis or horizontal axis.

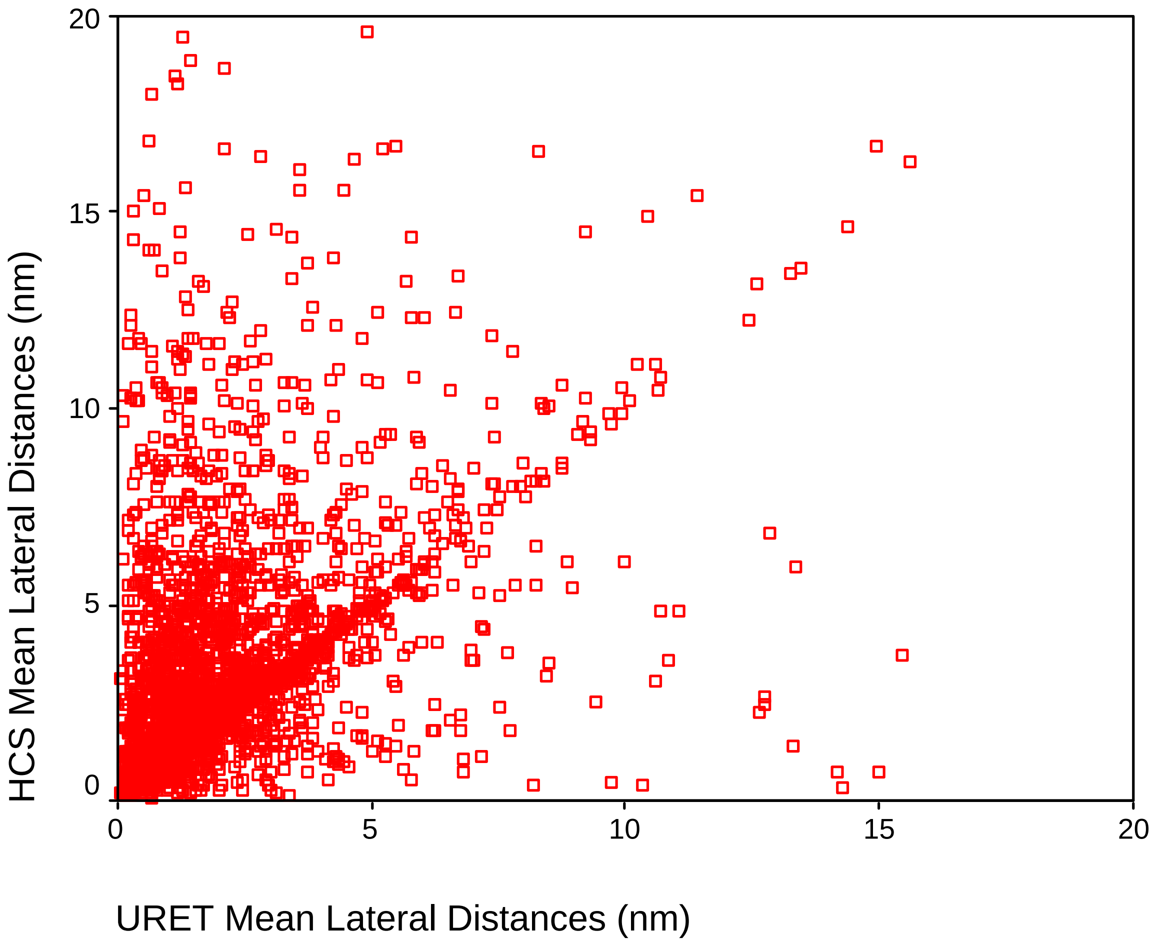


Figure 12: Scatter Plot of Lateral Distances

If the mean lateral distances were equal, the resulting points on the diagram would lie on the diagonal. It is quite evident from the scatter plot that there are more points above the diagonal than below the diagonal. This result indicates that the HCS appears to have a larger lateral distance on average than the URET, which further supports the previously discussed test.

6.2.2 Parametric Test

In order to perform any parametric tests, the underlying data should be normal or at least approximate a normal distribution. Applying the CLT should normalize the distribution, but the distribution of mean lateral distances is skewed somewhat. Taking the differences of the means of lateral distances results in a new distribution, called $\overline{D}$. This difference distribution does approximate a bell curve, although it is weighted to one side as shown in Figure 13. The weighted side indicates that the HCS has larger lateral distances on average than URET.

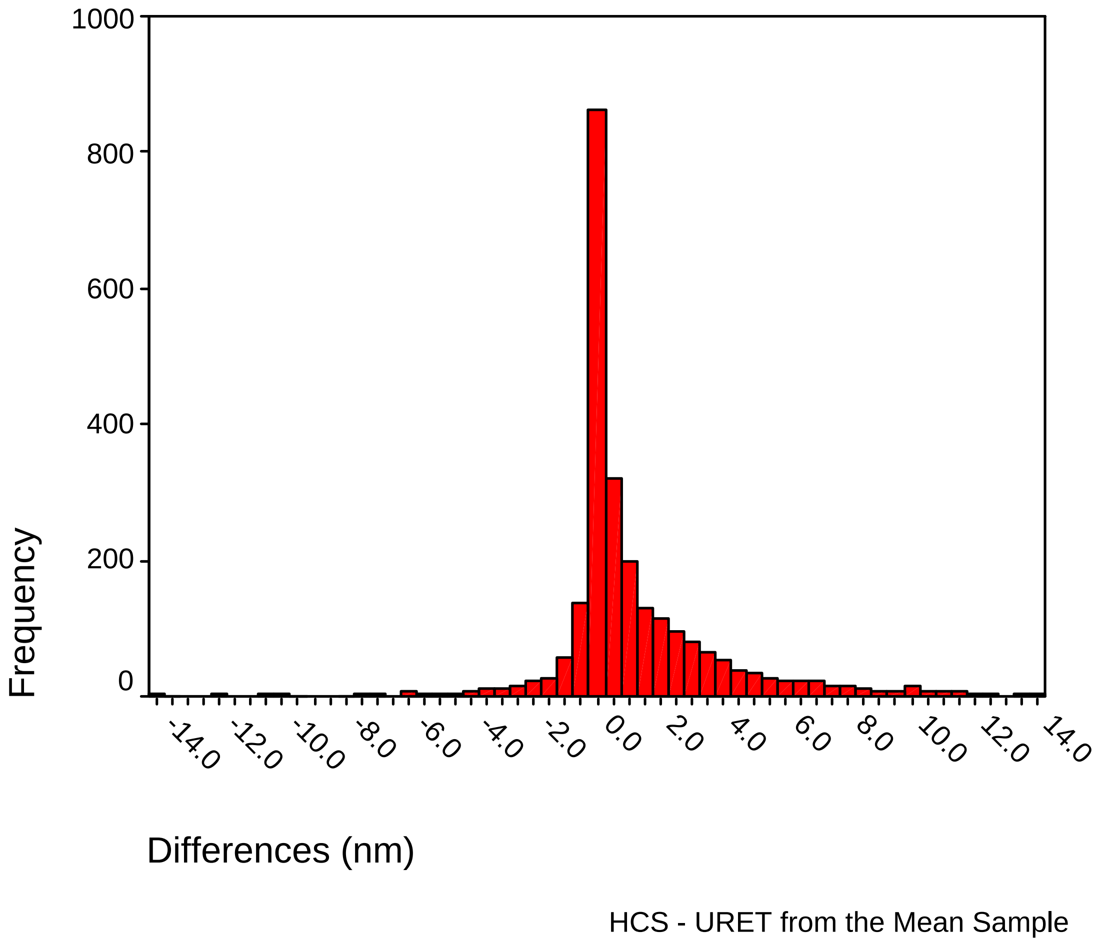


Figure 13: Histogram Representing Differences of Mean Lateral Distances

The paired t test assumes the differences $\overline{D}$ is a normalized distribution, but this assumption can be relaxed when the sample size is large, which it is. The alternative test for non-normalized data is the Wilcoxon signed-rank test[12], which is less sensitive than the paired t test. Since the dataset here is very large and close to normal, it is almost certainly safe to proceed with the paired t test.

The average difference (HCS – URET) of the mean lateral distances is 1.3 nm. Since URET lateral distances were subtracted from the HCS lateral distances, a positive number indicates the HCS had a larger lateral distance than the URET. Therefore an average difference of 1.3 nm indicates that on average the mean of the HCS mean lateral distances is about 1.3 nm greater than the mean of the URET mean lateral distances.

In order to test the hypothesis that the URET lateral distances exceed the HCS lateral distances significantly, a hypothesis test is performed. Assume the HCS performs the same as URET, and attempt to show that URET performance has a 99.9% probability of exceeding the HCS.

[12] See section 6.2.1

H_0: $\mu_D = 0$

H_1: $\mu_D > 0$

The standard t statistic with 2,535 degrees of freedom at the 0.001 significance level indicates that the null hypothesis should be rejected if $\mu_D > 3.09$. Computing the t ratio for this data sample results in a standardized mean of the difference scores of $\mu_D = 22.09$.

Since 22.09 > 3.09, reject the null hypothesis. Therefore it can be shown with greater than 99.9% certainty that URET performance in regards to the converted route probably exceeds the HCS. Actually the significance of this test (or α) is much better than 0.001. The statistical software package used for the test claimed α is 4×10^{-23}, but the significant digits of such a small number are questionable.

Adding to the outcome, the 95% confidence interval of the difference is 1.17 nm $< \overline{D} <$ 1.40 nm. Since the confidence interval does not contain zero, this result is further evidence that the difference is significant.

7. Conclusion

Using the ZDC sample data, the HCS and the URET processed the converted routes of 2610 flights. By design the resulting two data samples should vary only by their processing method. In order to perform strong statistical tests, the Central Limit Theorem was applied. The average value of the lateral distances was computed for each flight resulting in one statistic, the arithmetic mean, for the lateral distances per flight per system.

The arithmetic mean of the lateral distances show that the HCS has an average 3.4 nautical mile lateral distance after the outliers are eliminated. When URET is observed, the average lateral distance after removing outliers is 2.2 nautical miles. The median values of a random raw sample of 50 flights show an even better statistic with HCS deviating by 0.62 nm and URET deviating by 0.50 nm. Since the ideal case would be a lateral distance of 0 nautical miles, URET shows a substantial improvement from the HCS.

The comparison of the two systems reveals a strong result also. The data indicates that there is at least 99.9% probability that URET exceeds the performance of the HCS when the lateral distance is used as the metric. A nonparametric statistical test, the Wilcoxon signed-rank test, supports this conclusion. Furthermore the data indicates there is a 95% probability that the HCS will have between 1.17 nm and 1.40 nm greater lateral distances than URET.

The results of this study are quite conclusive. In every test attempted, URET outperformed the HCS as measured by the lateral distance between the cleared route of flight to the current track position.

Finally, the strategy and statistical tests described in this technical note should apply equally well to the converted route processing performance of any two systems. Although the results may differ, the same strategy and tests should produce definitive results for ERAM and the HCS.

8. Acronym List

ADIZ – Air Defense Identification Zone

AMTWG – Automation Metrics Test Working Group

APR – ATC Preferred Routes

ARTCC – Air Route Traffic Control Center

ATC – Air Traffic Control

CLT – Central Limit Theorem

COI – Critical Operational Issue

CPT – Conflict Probe Tool

DS – Display System

ERAM – En Route Automation Modernization

FAA – Federal Aviation Administration

FDP – Flight Data Processing

HCS – Host Computer System

NAS – National Airspace System

nm – Nautical Miles

PAR – Preferential Arrival Routes

PDAR – Preferential Departure Arrival Route

SDP – Surveillance Data Processing

SID – Standard Instrument Departure

STAR – Standard Terminal Arrival Route

TPM – Technical Performance Measurements

URET – User Request Evaluation Tool

WJHTC – William J Hughes Technical Center

ZDC – Washington Center at Leesburg, Virginia

9. References

1. Computer Sciences Corporation, (2001) "Functional Audit of Existing En Route Capabilities for the En Route Software Development and Support III (ERSDS III) Contract," Contract No. DTFA01-00-C-00047, July 31.

2. FAA, (2002) "Blueprint for NAS Modernization 2002 Update," October.

3. Lockheed Martin Corporation, (2004) "En Route Automation Modernization (ERAM), Software Requirements Specification (SRS) Algorithm Document, Volume II, Flight Data Processing (FDP) and Weather Data Processing (WDP) Algorithms," Document Number FAA-ERAM-2004-0026, May 21.

4. Lockheed Martin Corporation (2004), "En Route Automation Modernization (ERAM), Software Requirements Specification (SRS), Flight Services (FLTS)," Document Number FAA-ERAM-2004-0018, May 21.

5. Mendenhall, William, Dennis D. Wackerly & Richard L. Scheaffer, (1990) Mathematic Statistics with Applications, 4th ed., PWS-KENT Publishing Company.

6. Naiman, Arnold, Gene Zirkel & Robert Rosenfeld, (1995) Understanding Statistics, 4th ed., McGraw-Hill Science/Engineering/Math.

7. Oaks, Robert & Mike Paglione, (2005) "Determination of Lateral Flight Adherence in Recorded Air Traffic Data," Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference.

8. Paglione, Mike M., Robert D. Oaks, Dr. Hollis F. Ryan & J. Scott Summerill, Description of Accuracy Scenarios for the Acceptance Testing of the User Request Evaluation Tool (URET) / Core Capability Limited Deployment (CCLD), January 27, 2000.

9. Witte, Robert S., (1993) Statistics, 4th ed., Harcourt Brace Jovanovich College Publishers.

10. WJHTC/ACB-330, (2005) "En Route Automation Modernization Automation Metrics and Preliminary Test Implementation Plan Automation Metrics Test Working Group," Version 2.7, Atlantic City: FAA, June.

Appendix – Verification of Results Using URET Production Version

Due to time constraints, AMTWG used the MITRE Prototype of URET (Version D34-R6-P4) for this study. To confirm the findings, AMTWG recently checked the results against the fielded production version of URET. A different scenario using the same March 17, 2005 ZDC data was run through the HCS and the production version of URET. Other than the version of URET, the data reduction was the same as that discussed in section 4.

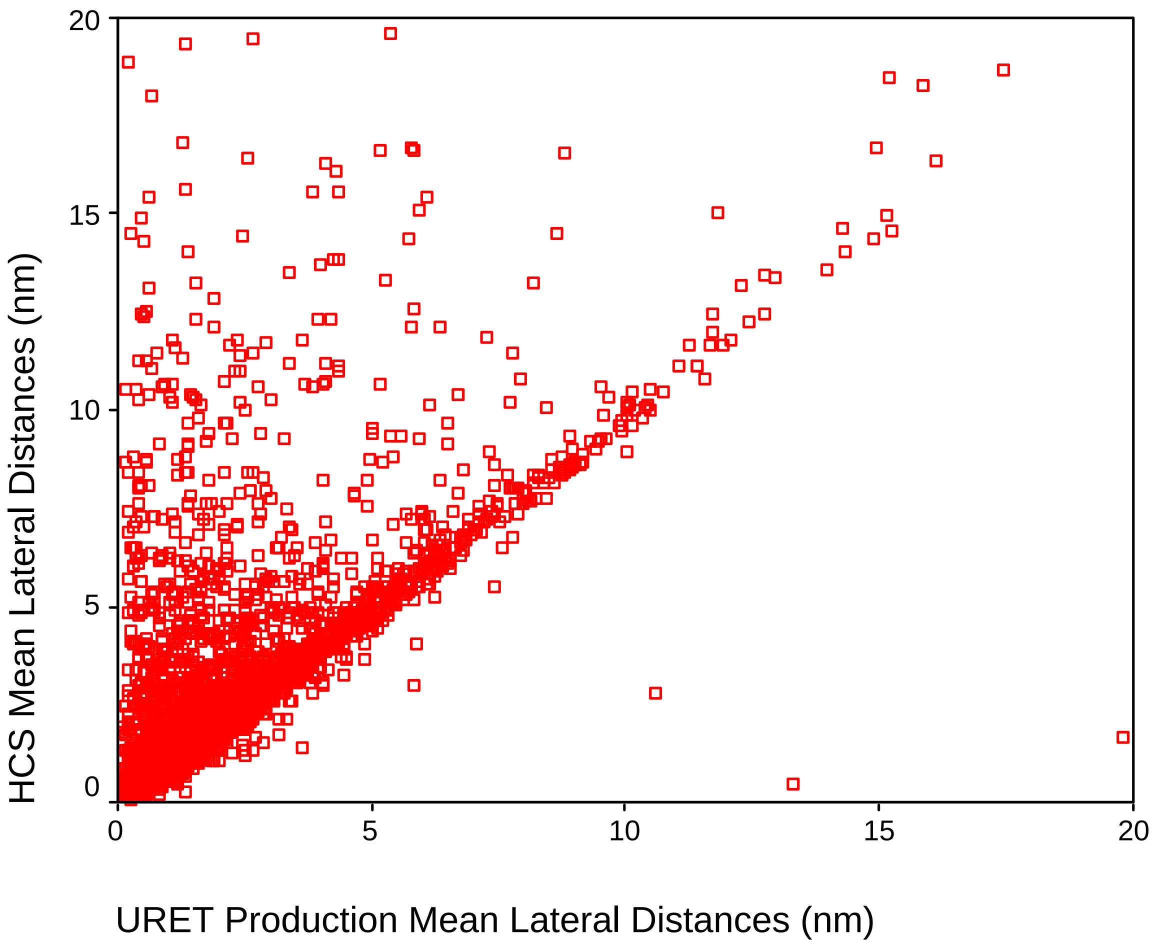


Figure 14: Scatter Plot of Lateral Distances (Production Version)

The graphical results of the verification are shown in Figure 14. The HCS lateral distances from the mean sample are on the y-axis or vertical axis, and the production URET lateral distances are plotted on the x-axis or horizontal axis. If the mean lateral distances were equal, the resulting points on the diagram would lie on the diagonal. It is quite evident from the scatter plot that most of the points are on or above the diagonal. This result indicates that for the most part the HCS appears to have a larger lateral distance on average than the production version of URET.

The results of the descriptive statistics are shown in Figure 15.

Sample Size (n)	2,538 flights
HCS Mean Lateral Distance (MLD)	3.50 nm
95% Confidence Interval for HCS MLD	3.36 nm < μ_{HCS} < 3.62 nm
Production URET Mean Lateral Distance	2.42 nm
95% Confidence Interval for URET MLD	2.33 nm < μ_{URET} < 2.52 nm
Mean Difference	1.07 nm
95% Confidence Interval for the Difference	0.97 nm < $\overline{D}$ < 1.17 nm

Figure 15: Descriptive Statistics Using Production URET

In order to analyze the comparison, the nonparametric Wilcoxon Signed-Rank test obtained the results shown in Figure 16.

HCS – URET	N	Sum of Ranks
Negative Ranks	834[a]	721,252.5
Positive Ranks	1704[b]	2,500,738.5
Ties	0[c]	
Total	2538	
	a. HCS < URET b. HCS > URET c. HCS = URET	

Figure 16: Wilcoxon Signed-Rank Test Ranks

By standardizing the sum of the ranks for the less frequent sign, in this case the negative ranks, differences in the distributions of the two variables can be tested. The standard z statistic at the 0.001 significance level of a 2-tailed test indicates that the null hypothesis should be rejected if $Z_W < -3.30$ or $Z_W > 3.30$. The significance of the test indicates the probability of the results, as shown in Figure 17.

	HCS - URET
Z_W	-24.10[a]
Significance, α (2-tailed)	0.00[b]
	a. Based on negative ranks, where HCS < URET b. Approximately zero

Figure 17: Wilcoxon Signed-Rank Test Results

Since $Z_W = -24.10 < -3.30$, the null hypothesis that HCS = URET is rejected.

Finally, the parametric paired-samples t test was used to analyze the results. The standard t statistic with 2,537 degrees of freedom at the 0.001 significance level indicates that the null hypothesis should be rejected if $\mu_D > 3.09$. Computing the t ratio for this data sample results in a standardized mean of the difference scores of $\mu_D = 21.50$. Since 21.50 > 3.09, reject the null hypothesis. Therefore it can be shown with greater than 99.9% certainty that the production version of URET performance in regards to the converted route probably exceeds the HCS.

Although the values are slightly different when using the two different scenarios and two versions of URET, the results are definitely the same. In every test attempted, the production version of URET outperformed the HCS as measured by the lateral distance between the cleared route of flight to the current track position.