AIAA Guidance, Navigation, and Control Conference
2 - 5 August 2010, Toronto, Ontario Canada

AIAA 2010-8162

Accuracy Comparison of an Operational and Experimental Strategic Conflict Probe

Confesor Santiago[1], Jeffrey Lehman[2], and Andrew Crowell[1]
Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City, NJ 08405

The primary function of administering the United States' National Airspace System (NAS) is the air traffic controller task of actively monitoring assigned aircraft and resolving the conflicts (i.e. losses of minimum separations between aircraft) anticipated some time in the future. The transformation of the current NAS is a planned evolution and will rely on communication digitally between air and ground and between aircraft with the use of enhanced trajectory-based operations and automation tools. To mitigate the safety risks of increased traffic growth and effectively designing automation to aid in the separation management task, research and development is required to improve the accuracy and usability of decision support tools. The paper presents the performance of an experimental strategic conflict probe model used in the development and validation of enhanced separation management concepts. Furthermore, this papers reports on a comprehensive design of experiment optimizing its performance and a comparison of its capability to an advanced operational system.

Nomenclature

Add_J_Cnt	=	number of radar track reports that define window used in processing the acceptance of a conflict prediction notification
Add_%_I_of_J	=	percentage of *Add_J_Cnt* used in deciding to present conflict notification
Del_J_Cnt	=	number of radar track reports that define window used in processing the deletion of a conflict prediction notification
Del_%_I_of_J	=	percentage of *Del_J_Cnt* used in deciding to delete conflict notification
DLT	=	amount of time for CP to remove the notification of a conflict that has actually ended
LH	=	look ahead time, in seconds, or time in the future that conflict predictions are probed
Min_HSep	=	minimum horizontal separation, in nmi, used in the probing for potential conflicts
R(FA)	=	likelihood that a predicted alert is indeed a false or nuisance alert
R(MA)	=	likelihood that the conflict probe does not predict the conflict when it occurs
WT	=	amount of lead time between the ATC being notified of the predicted conflict and the conflict actually occurring

I. Introduction

THE primary function of administering the United States' National Airspace System (NAS) is the air traffic controller task of actively monitoring assigned aircraft and resolving the conflicts (i.e. losses of minimum separations between aircraft) anticipated some time in the future. In the current air transportation system, Decision Support Tools (DSTs) aid air traffic controllers (ATC) in monitoring air traffic in order to maintain minimum separation standards between aircraft. These automated systems provide this service by predicting aircraft flight paths (trajectories), in order to foretell potential conflicts. The strategic conflict probe (CP) is one component of a DST that alerts ATC of potential conflicts normally up to twenty minutes into the future. The User Request Evaluation Tool (URET) is an operational DST deployed in all twenty en route Air Route Traffic Control Centers (ARTCC). Over the years the conflict prediction accuracy performance of URET has been researched and documented. Recently, a study investigating the properties of conflicts in the NAS was performed[1]. In this study an experimental strategic conflict probe model was employed. In order to validate the CP model, it was compared to

[1] Computer Scientist, FAA Simulation and Analysis Team, AJP-661, AIAA Member.
[2] Operations Research Analyst, FAA Simulation and Analysis Team, AJP-661.

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

the performance of URET using a six-hour Washington, DC Center air traffic sample. During this comparison, the experimental CP model was found to outperform certain aspects of URET. In order to properly design and develop enhanced automation to aid in the separation management task, further knowledge of these findings could offer important clues. This paper will present areas where the model bested URET, provide data to support these findings, and expose reasons for the results.

II. Previous Work and Background

There has been much research in studying the performance of operational strategic conflict probes[2, 3]. Currently, a replacement to the Host Computer System and URET is being developed. The replacement is called En Route Automation Modernization (ERAM). Furthermore, an ongoing project called Trajectory-Based Operations Separation Management: Modern Procedures is investigating ways to mitigate the safety risks of increased traffic growth. One prime task of this project is to identify requirements for ERAM to handle the new concepts of future separation management. Furthermore, the project involves computer human interface improvements to ERAM Post-Release 3 and algorithmic enhancements to the Conflict Probe to increase the usability of the tools for the near-term timeframe of NextGen. This paper supports this task. Through experimenting with our CP model the goal is to identify areas for improvement within the separation management and DST environment.

III. Overview of Strategic Conflict Probe

A conflict is a loss of the minimum separation standard when an aircraft enters the minimum separation area of another aircraft or a restricted airspace. Currently the minimum separation parameters in En Route airspace are generally five nautical miles horizontally and 1,000 feet vertically. There are two types of conflict probes utilized by ATCs in the NAS: tactical and strategic. A tactical conflict probe is used for more imminent predictions of conflicts (i.e. within 75 to 135 seconds from occurring). Tactical CPs use radar track data for the actual position, heading, and velocity of aircraft and ATC vertical clearances for probing of potential conflicts. A tactical CP alerts the ATC of an imminent conflict and action is taken immediately to resolve the issue. In a strategic CP, the prediction look ahead time is greater (i.e. twenty minutes). Also, strategic CPs use aircraft trajectories, future weather forecasts, and adaptation data to predict potential conflicts. The alerts from this type of CP are used more for strategic planning of airspaces and metering of the traffic. This paper focuses on comparing the performance of an operational CP and a newly developed CP model.

The operational CP used in our study is URET. The performance of URET has been studied for years[2,4]. Prior to this study, URET was executed by the ERAM developers during the ERAM Run for Record test scenario. The performance of URET was recorded and delivered to the authors as part of an ongoing task in regression testing of ERAM compared to URET.

The experimental CP model called Trajectory Conflict Probe (TrajCP) was developed in-house by the FAA's Simulation and Analysis (S&A) Team. As mentioned previously, TrajCP and the laboratory platform it resides in were originally created for a study on properties of conflicts within the NAS. The study investigated conflict properties such as distribution of conflict duration, encounter angle, vertical phase of flight, and frequencies per ARTCC across the NAS. At the core of TrajCP is a trajectory predictor (TP) developed to generate future flight paths of aircraft. TrajCP uses these predicted trajectories to probe for potential conflicts. TrajCP was designed in modular fashion where the TP is decoupled from the CP functionality. This creates the capability to use different methods of generating aircraft trajectory predictions for different purposes and studies. TrajCP can be configured to generate trajectories using standard trajectory modeling inputs or it can be set up to load pre-recorded trajectories from a database or file. For instance, TrajCP can predict potential conflict events using its own configurable CP methodology while using trajectories from an external TP such as URET. TrajCP is described, in detail, in Ref. 1 with flowcharts and diagrams depicting how conflict predictions are produced.

IV. Accuracy Metrics

When determining the accuracy of a CP there is a set of metrics used to quantify its performance[4-5]. Before metrics are calculated, the conflict predictions of a CP need to be evaluated. In order to evaluate the performance of a CP two datasets are required. The first is the collection of predicted conflicts the CP generated. Conflict predictions are normally generated during CP execution as ATC monitor and separate live traffic, however this data is archived for accuracy analyses described in this section. The second is the collection of actual conflicts determined through processing the air traffic scenario's ground truth data (flight paths as flown). This task is normally part of post processing testing functions where the radar track reports of all flights are compared with each in order to calculate where separation violations occurred. The evaluation of the predicted conflicts versus the actual

conflicts creates the performance results dataset. The techniques of how the performance of a CP is evaluated are presented in Appendix A of Ref. 3.

Table 1 details the different attributes (metrics) within evaluated CP results. There are three types of situational attributes and two attributes regarding timeliness. The three situational attributes are valid alerts (VAs), missed alerts (MAs), and false (nuisance) alerts (FAs). A VA occurs when a CP correctly predicts an actual conflict, a MA occurs when a CP does not predict an actual conflict, and a FA occurs when a CP predicts a conflict, but one does not occur. Consequently, the more VAs produced by a CP the better, and alternatively the less MAs and FAs produced by a CP the better. The optimal CP possesses a VA for every actual conflict represented in the air traffic data, and has no missed or falsely predicted conflicts.

Conflict Prediction Attributes	Description
Valid Alert (VA)	Actual conflict occurs, and CP correctly predicts event in timely fashion.
Missed Alert (MA)	Actual conflict occurs, and CP either does not predict or predicts in untimely fashion.
False Alert (FA)	No conflict occurs, and CP predicts conflict to have occurred.
Warning Time	The amount of lead time a VA provides prior to the start of the conflict.
Deletion Lag Time	The amount of time between the conflict ending and the prediction notification being deleted or removed.

Table 1: Metrics listing of evaluated conflict predictions.

Two statistical metrics are used that are related to the three situational attributes. The first is the rate of missed alerts (*R(MA)*), which is the number of missed alerts divided by the sum of the valid alerts and missed alerts. This gives the rate at which, given an actual conflict, it will be missed by the CP. The second metric is the rate of false alerts (*R(FA)*), which is the number of false alerts divided by the sum of the valid alerts and false alerts. This gives the rate at which, given an alert, an actual conflict does not occur during it.

Once the number of Valid Alerts (VAs) has been counted, another metric of conflict predictions is warning time (*WT*). This is the first of two timeliness attributes of a conflict prediction (Fig. 1). *WT* is the amount of lead time between the ATC being notified of the predicted conflict and the conflict actually occurring (amount of warning the ATC has to resolve the conflict). As illustrated in Figure 1, at t_1 a conflict notification would be present, and t_2 would be the start time of the actual conflict event. A large amount of *WT* is the most desirable for conflict predictions. The second timeliness attribute is deletion lag time. Deletion lag time (*DLT*) is the amount of time it takes a CP to remove the notification of a conflict after the conflict has actually ended. It is not desirable for a CP to have much delay in its awareness of a conflict resolution and ending. In Figure 1, *DLT* would be represented by the post-analysis calculation of the end of an actual conflict occurring at t_3 and at t_4 the system removing the conflict prediction notification. Normally, an ATC is aware of a conflict's end because they are the ones monitoring the situation and who have resolved it, hence this metric has not been used in previous operational testing of CPs in the past, as the other four metrics have. *DLT* is a response metric that arose specifically after the initial comparison of TrajCP with URET discussed in the previous study.

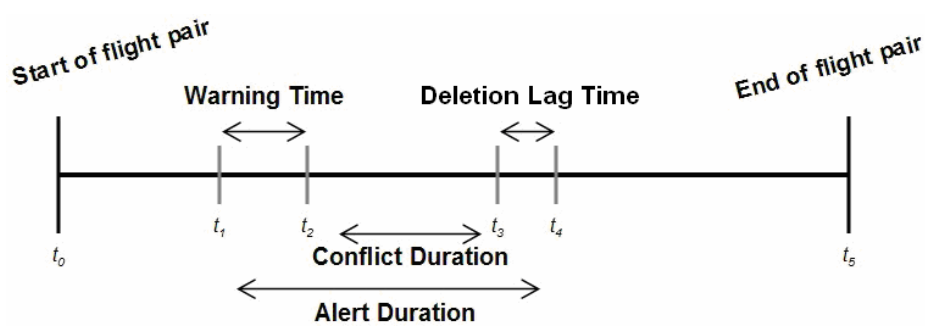

Figure 1. Diagram of the different time values in a conflict alert.

V. Study Settings

This paper focuses on detailing differences in accuracy between an operational CP (URET) and an experimental CP model (TrajCP). The differences are quantified by calculating the performance metrics described in Section IV then comparing the results. The experimental CP was studied under two settings where the TP varied; hence two comparisons are detailed below. These two settings were easily supported by TrajCP because of its unique design.

As mentioned previously, the design allows for different TPs to be plugged into TrajCP without modifications to the CP, which creates a controlled environment where the CP can be evaluated under various cases and ranges in TP accuracy.

A. Hybrid-Merge Trajectory Prediction Setting

The first setting used trajectory predictions from an experimental trajectory predictor developed in association with TrajCP. This trajectory prediction algorithm is called Hybrid-Merge (HM). The algorithm gets the name "*Hybrid*" because it chooses based on a set of thresholds to start with either linear extrapolation or flight plan of a flight for trajectory predictions, and "*Merge*" because if linear extrapolation is used, then the trajectory path is merged into the flight plan, if it intersects, and follows the flight plan for the remainder of the prediction. The complete detail of the HM algorithm is presented in Ref. 1. This setting is referred to as TrajCP-HM. Furthermore, the results of a NAS-wide conflict properties characteristic study[1] are based on using the HM algorithm. The comparison of these results with original URET is the main subject of this paper.

B. Recorded Trajectory Prediction Setting

The second setting utilized the trajectory predictions from URET itself. The URET trajectories were a result of the Formal Acceptance Test of ERAM where eight accuracy requirements were tested to ensure ERAM performed at least as well as URET in trajectory prediction and conflict prediction accuracy. These trajectories were exported from URET's System Analysis Recording and loaded into an Oracle database table. This trajectory predictor was developed where instead of modeling a trajectory prediction for a flight at a given time, an archived trajectory is simply retrieved. We refer to this algorithm as the Recorded Trajectory Engine. Since an external TP with its own algorithms generates trajectories with associated 4D flight path predictions and identifying build times, the active trajectories at a certain time for a flight pair are used by TrajCP in the probing for potentials conflicts. This setting is referred to as TrajCP-Recorded. The comparison of these results with original URET is presented only as a further investigation to substantiate the CP within TrajCP. In theory, the differences between TrajCP-Recorded and original URET lie strictly in the differences of the CP algorithms, since the trajectories used are the same. This offers a more focused investigation of CP functions, in general, because if TrajCP-Recorded outperforms URET in certain cases, these cases can be researched and possibly lead to identifying enhancements in predicting conflicts for NextGen separation management initiatives.

VI. Optimization of CP Model Parameters

The conflict probe modeling software is designed in a manner where the input parameters can be adjusted to affect the procedure in predicting conflicts, updating alerts, and removing alerts from the system. The objective of the design of experiment (DOE) is to determine the best levels to set the six factors (input parameters) to minimize the missed *(R(MA))* and false alert *(R(FA))* rates, and deletion lag time while maximizing the warning time. The model assumes a linear relationship between the variables and a normally distributed error.

A. Experimental Design

In the current study, the high and low settings for each of the experimental factors were determined so as to cover the operability region with respect to prior studies. Table 2 shows these operability region limits. *LH* refers to the modeled look ahead time, in seconds, or time in the future that conflict predictions are presented. *Add_J_Cnt* is the number of consecutive 10-second processed radar track reports that TrajCP uses to present a conflict notification, and ultimately defines the window of track reports that TrajCP uses to evaluate whether to present a notification of a predicted conflict. *Add_%_I_of_J* is the percentage of the *Add_J_Cnt* in which predicted conflicts must appear in order for the CP model to accept the predicted conflict as an alert to the ATC. For instance, an *Add_J_Cnt* of five and a *Add_%_I_of_J* of 100 equates to the requirement of consecutive predicted conflicts occurring for five track reports. *Del_J_Cnt* and *Del_%_I_of_J* are similar to *Add_J_Cnt* and *Add_%_I_of_J* except they apply to how the CP model deletes notifications of predicted conflicts by counting the number and percentage of consecutive track reports that do not contain predicted conflicts. Lastly, *Min_HSep* refers to the minimum horizontal separation, in nautical miles, over the duration of the notification.

Attempts were made to conduct screening experiments as a preliminary step to apply response surface methodology. Several conditions made such an approach difficult. With four responses, screening did not eliminate any factors. Also, the response metric with the highest importance rating, R(MA), had only eight distinct (out of 32) experimental values in the region of proposed experimentation. An additional constraint on the design was that Add_%_I_of_J and Del_%_I_of_J were integral multiples of Add_J_Cnt and Del_J_Ct, respectively. Although

Experimental Settings		LH	Add_J_Ct	Add_%_I_of_J	Del_J_Ct	Del_%_I_of_J	Min_HSep
Level	High	1500	25	100	30	100	9
	Low	600	5	60	10	60	5
	Scale	30 sec	integer	% / Integer	Integer	% / Integer	0.5 mile

Table 2: Experiment Operability Limits

rounding was a possibility, this condition somewhat limited the choice of design points in a sequential experiment. Finally, preliminary screening indicated that the response surface was fairly complex. Due to these conditions, a 2V6-1 design was applied as in the prior study, but to a modified operability region. As there are no replications for each treatment combination, there is no experimental error to estimate in the full model. The full model should be run in 2F-p experiments without replication 7, 8, and 9. Normal Probability Plots and Lenth Plots were used to determine the active factors on each response metric separately, and then a second generation model was run using the active factors to determine the optimal settings and corresponding response vector.

B. Experimental Results

The experiment described above was instituted in order to determine the optimal input parameters for the TrajCP model. Results from this experiment were used in the direct comparison of the CP model against URET, which is the main subject of this paper. Furthermore, to drive the experiment and measure the quantity of responses of the TrajCP software tool, a truth data set of conflicts and input air traffic scenario must be provided to TrajCP. The methods used involve creating a set of pseudo test conflicts by time shifting an actual air traffic recording. This is described in detail in Ref. 6. The original data source used in this study is an extraction of a ZDC recording from March 2005. It was originally developed for the formal accuracy testing of ERAM's Conflict Probe Tool[3]. It includes a sample of 418 flights that encounter a separation violation (conflict) during a 6-hour period of high traffic levels.

1. DOE Responses

The experimental design matrix and response metric values resulting from the experimental set of Hybrid-Merge Trajectory Engine (HMTE) runs can be found in Table 7 in the Appendix. The experimental design matrix and response metric values resulting from the experimental set of Recorded Trajectory Engine (RCTE) runs can be found in Table 8 in the Appendix. Figure 2 shows the *Run_ID* Level comparison of *R(FA)* and *R(MA)* by Trajectory Engine Type. Except for a handful of instances, HMTE *R(FA)* and *R(MA)* values are less than RCTE corresponding values.

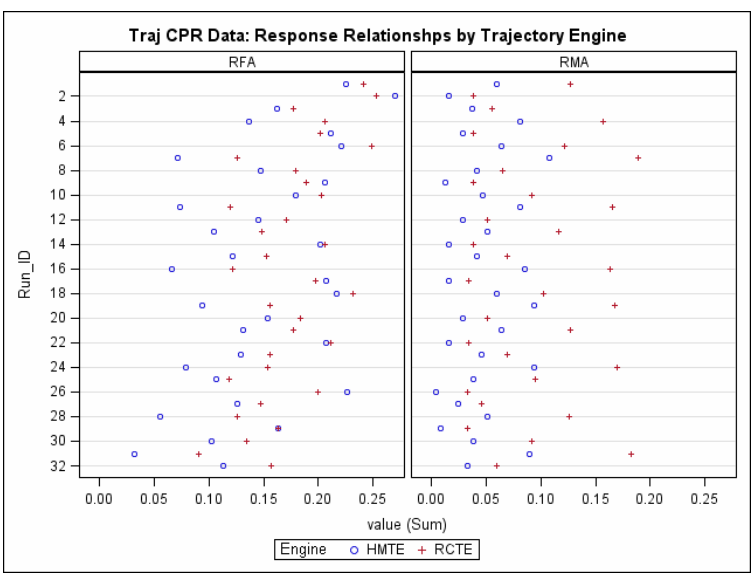

Figure 2: R(FA), R(MA) at each of 32 Experimental Settings and Trajectory Engine Type

2. DOE Modeling

A full regression analysis model was applied to response results of HMTE to fit the responses as a function of the experimental factors. Lenth and Normal Probability plots shown in Figures 3 and 4 in the Appendix showed that all main effects are significant on at least one of *R(FA)*, *R(MA)*, and *WT*. Therefore, all main effect factors were included in the multivariate model to identify optimal settings. In addition, the interaction *Min_HSep*Add_J_Ct* was active on *R(MA)*. Therefore, the model used to find the optimal settings included all main effect factors and the *Min_HSep*Add_J_Ct* interaction factor.

As was the case with the HMTE, all main effect factors were included in the multivariate model to identify optimal settings. In addition, the *Min_HSep*Add_J_Ct* was the strongest active interaction factor on *R(MA)*. Although a few other interactions such as *Min_HSep*Del_J_Ct* and *Min_HSep*LH_Time* were also marginally active, in consideration of overfitting, consistency with the HMTE model, and the fact that there was little marginal benefit of including the other two interactions, the same single interaction model was applied to the RCTE data to determine optimal settings and response values.

3. DOE Interpretations

Table 3 provides a summary of the relative influence of the main effect factors on each response for both engines. The cell values represent the ratio of the effect parameter estimate to the Margin of Error (ME) and the Simultaneous Margin of Error (SME), which are thresholds for determining whether a factor is active on a response. For example, *Add_J_Ct* and *Min_HSep* have the strongest influence on *R(MA)* for both trajectory engine types, whereas *Del_J_CT*, *Del_%_I_of_J*, and *Min_HSep* have the strongest influence on *DLT*.

Parameter Estimate Ratios: βi/ ME & βi / SME								
Factor	R(MA)		R(FA)		WT		DLT	
	HMTE	RCTE	HMTE	RCTE	HMTE	RCTE	HMTE	RCTE
A: LH	-0.71	-1.00	0.97	1.33	2.41	**7.97**	-0.25	-0.99
B: Add_J_Ct	**3.17**	**3.95**	**-3.48**	**-2.19**	**-6.22**	**-6.36**	0.31	-0.63
C: Add_%_I_of_J	1.09	1.18	-1.32	-0.91	-1.98	-2.37	0.15	0.21
D: Del_J_Ct	-1.53	-0.92	-1.75	**-2.04**	3.60	3.23	**49.79**	**30.35**
E: Del_%_I_of_J	-0.70	-0.66	-1.10	-1.06	2.61	1.37	**35.50**	**20.18**
F: Min_HSep	**-5.27**	**-9.02**	**2.45**	1.08	**11.45**	**10.05**	**25.48**	**18.40**
ME	0.0044	0.0052	0.0120	0.0105	4.37	7.01	1.74	2.65
SME	0.0084	0.0100	0.0228	0.0200	8.31	13.33	3.31	5.05

Table 3: Parameter Estimate Ratios

The models for both engines were used to find the optimal settings and response variables using desirability functions (see Figures 9 and 10 in Appendix). Subject matter experts were consulted, and shared their view that the appropriate importance ratings were (*R(MA) = 4, R(FA) = 4, WT= 3, DLT = 2*). The settings selected are shown in Table 4 with corresponding optimal response values, based upon the weights. For both trajectory engine settings used in the DOE optimization of TrajCP the desirable *Min_HSep* parameter was 9 nautical miles. Interestingly, both resulted in a *Add_%_I_of_J* equal to 100, however the HMTE (25) resulted in a much higher *Add_J_Cnt* than RCTE (5), which means TrajCP under the HMTE setting only presents notification of potential conflicts if a predicted conflict is realized for 25 consecutive radar track reports.

C. DOE Model Summary

A more compact study might admit a sequential experimental strategy such as a response surface methodology. For example, *R(FA)* and *WT* showed such strong negative correlation ($0.88 < r < 0.96$) that only one of these could be included in the model and the other inferred. As *Add_%_I_of_J* and *Del_%_I_of_J* were not strongly significant, these factors could be tested for removal. However, the issue with modeling data with so few missed alerts would have to be accounted for. Both the relationships between the factors and the response region itself are fairly complex.

Settings	RCTE	HMTE
Min_HSep	9	9
Add_J_Ct	5	25
Add_%_I_of_J	100	100
Del_J_Ct	10	25
Del_%_I_of_J	100	80
LH	1500	1500

Responses	RCTE	HMTE
R(MA)	0.035	0.035
R(FA)	0.207	0.134
WT	441	186
DLT	307	368

Table 4: Optimal Settings and Response Values

In multivariate response models, generally each response variable is optimized separately and the results are overlain. Alternatives to this strategy include multivariate techniques such as the construction and optimization of a desirability function or the formulation and execution of a mathematical program whereby constraints on the allowable ranges of the response variables with lower importance ratings are applied.

VII. Accuracy Analysis and Comparison

As previously mentioned, a CP modeling algorithm was used with an experimental TP (HMTE) as well as the operational tested URET TP (RCTE). Once the parameters for TrajCP were optimized using DOE modeling techniques, two different levels of accuracy analyses and comparisons were performed in this study.

The first analysis performed was a comparison between URET's CP and TrajCP-HM. In this comparison, both the CPs and the TPs are different, so the differences in the results require some analysis to determine whether they are caused by the CP or the TP. The second analysis was comparing TrajCP to the CP within URET that is currently deployed in the NAS. This analysis is performed by using the Recorded Trajectory Engine with the URET trajectories within the experimental CP algorithm, then comparing the results to the results of URET's CP. Performing this comparison will help to find areas where the CP algorithms differ, and give a basis for the comparison of the different TP algorithms.

Conflict Prediction Metric	URET	TrajCP-HM	TrajCP-Rec
Valid Alert (VA)	237	233	228
Missed Alert (MA)	5	9	8
False Alert (FA)	98	34	61
Verified Conflicts	242	242	236
Verified Alerts	335	267	289
R(MA)	0.0207	0.0372	0.0339
R(FA)	0.2925	0.1273	0.2111
Avg. Warning Time	00:04:44	00:03:59	00:08:32
Med. Warning Time	00:02:23	00:01:10	00:04:20
Avg. Deletion Lag Time Time	00:02:56	00:07:48	00:05:09

Table 5: Conflict prediction performance for each Conflict Probe analyzed.

To perform both comparisons we use several performance metric statistics of each CP setting as well as a pairing algorithm that matches each conflict prediction in one CP to its respective conflict prediction, if one exists, in the other CP[11]. The statistics used for comparison are listed in Table 5 and the results of the two algorithmic matching comparisons are shown in Table 6.

URET and TrajCP-HM each have the same number of verified conflicts. A verified conflict is an actual encroachment of one aircraft on another that has been accepted for analysis. Since the ground truth used for each CP setting is the same, the only way the number of verified conflicts can differ among the three is if there are actual conflicts that are discarded from the analysis because the flight was not adhering to the filed flight plan at the time the conflict occurred and the CP missed the alert. TrajCP-Recorded contains six less verified conflicts all of which were valid alerts in URET but discarded in TrajCP-Recorded because they were missed alerts but the flight was not adhering to the route at the time of the conflict.

Verified Alerts is the number of alerts that were presented to the ATC. This is the number of valid alerts plus the number of false alerts. *R(MA)* and *R(FA)* are the rates of missed alerts and false alerts, respectively, as described in

Section IV. The last three statistics are average warning time, median warning time, and average deletion lag time. These statistics were also described in detail in Section IV.

A. Accuracy Result Overview

The conflict prediction performance and accuracy results of URET, TrajCP-HM, and TrajCP-Recorded are detailed in Table 5. There are no major differences in the overall counts of VA. TrajCP has double the amount of MAs than URET in both settings, and TrajCP has significantly less FAs than URET. Interestingly, the average WT for TrajCP-HM is less than URET, however TrajCP-Rec outperformed URET in regard to WT. In regard to average deletion lag time, URET removes alerts of predicted conflicts much faster than the TrajCP model.

Comparison Result	URET vs. TrajCP-HM	URET vs. TrajCP-Rec
VA_MA	(a) 9	(h) 4
MA_VA	(b) 5	(i) 1
FA_NC	(c) 78	(j) 66
NC_FA	(d) 8	(k) 17
SAME_VA	(e) 225	(l) 223
SAME_MA	(f) 0	(m) 4
SAME_FA	(g) 19	(n) 27

Table 6: Comparison differences counts for conflict prediction performance: URET vs. TrajCP.

B. URET vs. TrajCP-HM

FA_NC cases are those in which the URET had a false alert, but TrajCP-HM had no comparable alert. Out of the 78 FA_NC (see Table 6(c)) that were in URET vs. TrajCP-HM, five were cases that were within five minutes before a SAME_VA. Of these five cases, the warning time was improved in TrajCP-HM over URET for two with an average improvement for all five of 00:03:07. Most other FA_NC cases were caused by TrajCP-HM using the I of J logic to delay the alert notification of a potential conflict. TrajCP has this capability as it generates a new trajectory every ten seconds; alternatively URET only generates a trajectory based on certain events such as out of conformance or new ATC clearance. More frequent trajectory generation helps to considerably reduce the number of false alerts.

The NC_FA cases are just the opposite of FA_NC. In these cases, there was a false alert in TrajCP-Recorded that was not comparable with any alert in URET. Five of the NC_FA cases (see Table 6(d)) in URET vs. TrajCP-HM were caused by the flight being off the filed flight plan (route). TrajCP-HM predicted the flight to fly straight until it intersected the route, where URET predicted it would immediately turn back toward the route. In each case, URET was more accurate and did not produce a false alert. The last three NC_FA cases were caused by TrajCP-HM's trajectory predictor being more accurate. The conflict prediction was run with a horizontal separation standard of 9 nmi. These three aircraft pairs did indeed come within 9 nmi, however, they did not come within the 5 nmi separation that is required for an actual conflict to occur.

The SAME_VA cases are those that have comparable valid alerts in both systems. There were two more SAME_VA cases in URET vs TrajCP-HM than there were in URET vs. TrajCP-Recorded. Although the TrajCP-HM algorithm was able to predict more valid alerts than TrajCP-Recorded, it sacrificed a lot of warning time to do so. The average decline of warning time in TrajCP-HM versus that of URET for the 225 SAME_VA cases (see Table 6(e)) was 00:00:32.

The MA_VA cases are those in which there is a missed alert in the URET that is the same conflict that received a valid alert in TrajCP-HM. The reason for five MA_VA cases (see Table 6(b)) in the URET vs. TrajCP-HM comparison is from URET having four MA because it was unable to build trajectories for at least one of the aircraft in each pair. The remaining one based on TrajCP-HM was caused by a large gap of about 4.5 minutes between two trajectory points in the URET trajectory surrounding the time at which the conflict occurred. TrajCP-HM performed a linear interpolation between the two trajectory points and was able to use the interpolated data to correctly predict the conflict. URET's CP did not perform this interpolation, and therefore missed the conflict due to lack of trajectory data. As resolution to the issue of URET failing to build an aircraft trajectory; a fail safe method could be designed to build a less accurate more tactical dead reckoning trajectory using active ATC cleared altitude.

Continuing the details of differences between URET and TrajCP-HM, VA_MA cases are just the opposite of MA_VA. In these cases, there is a valid alert in URET that was for a conflict that was missed in TrajCP-HM. The nine VA_MA cases (see Table 6(a)) were caused by a variety of reasons. Three were caused by one aircraft in the flight pair being given an interim clearance altitude close to the conflict. URET trajectories effectively ignore interim clearances, where HM trajectories predict the aircraft will level off at the interim clearance. This caused URET to be more accurate in its prediction and correctly predict the conflict would occur. Four VA_MA cases were caused by TrajCP-HM having poor along track error, resulting in a prediction that one aircraft would arrive to a location later than it actually did. One of the VA_MA cases was caused when a flight was slightly off the route, and

was headed away from the route. TrajCP-HM predicted the aircraft would fly in a straight line, where URET predicted the flight would turn back toward the route, which was correct.

Similar to the SAME_VA cases, the SAME_MA and SAME_FA cases are those in which both systems had similar results in not predicting an actual conflict, or predicting a conflict that did not occur, respectively. These cases are not particularly interesting, since this paper is most interested in determining the differences between the systems.

C. URET vs. TrajCP-Recorded

Out of the 66 FA_NC (see Table 6(j)) that were in URET vs. TrajCP-Recorded, twenty were cases that were within five minutes before a SAME_VA, which means the alerts were most likely absorbed within single notifications. Of these twenty cases, the warning time of the VA in the TrajCP-Recorded was improved over URET in fourteen. The average warning time improvement for all twenty cases was 00:06:14.

Since the same trajectories were used in both systems, and a higher separation standard was used in TrajCP-Recorded, it is valid to state that most FA_NC cases were directly related to the I of J logic included in TrajCP. The I of J regarding to the adding of a conflict prediction required 25 consecutive track reports having a potential conflict detected by the model. Thus, this protects against many nuisance alerts, but has negative effect on the timeliness of alert notifications, because of the natural delay in the I of J logic. By delaying the notification of the predicted conflict, TrajCP correctly removes false alerts from the system where URET's CP does not in many cases.

All of the seventeen NC_FA (see Table 6(k)) that were in the URET vs. TrajCP-Recorded were caused by the 9 nmi separation standard used in the TrajCP-Recorded CP as opposed to the 5 nmi separation standard used by the URET CP. Although the trajectories used in each system were generally the same, the more strict CP settings of TrajCP-Recorded caused several false alerts on aircraft pairs that were separated by greater than 5 nmi but less than 9 nmi. These extra FAs were the price to pay for the optimal CP results which ultimately resulted in less overall FAs.

Again, SAME_VA cases are those that have comparable valid alerts in both systems. This is where the URET CP and TrajCP each performed similarly, though the warning times were significantly different. The average improvement of warning time in TrajCP-Recorded over that in URET for the 223 SAME_VA cases (see Table 6(l)) was 00:03:39.

The single MA_VA (see Table 6(i)) in the URET vs. TrajCP-Recorded is the same as the last discussed MA_VA in the URET vs. TrajCP-HM and was caused by the same issue of interpolation.

Once more, VA_MA cases are just the opposite of MA_VA. In these cases, there is a valid alert in URET that was for a conflict that was missed in TrajCP-Recorded. It is not completely clear what caused the four VA_MA cases (see Table 6(h)) in the URET vs. TrajCP-Recorded comparison. The trajectories used to predict the conflict in URET do not appear to indicate that the aircraft pair will come into conflict. However, the trajectories do predict the aircraft will become very close to a conflict being separated only by 1000 feet vertically. Because only one foot less separation would have caused a conflict prediction in TrajCP-Recorded, the difference between the two CPs could have simply been round-off error in the altitudes.

There are several cases in the URET vs. TrajCP-Recorded comparison that are not shown in Table 6, but are worth mentioning. As discussed earlier there are six less verified conflicts in TrajCP-Recorded, and these are caused by the flight not adhering to the route during a missed alert. The routes and actual flight paths of the aircraft are the same in each CP system, therefore as TrajCP-Recorded was excused for missing the alert, it is interesting to note that URET was able to predict the alert, although it was excused for not meeting the minimum warning time requirement.

It was speculated that all differences found between URET and TrajCP-Recorded would be strictly due to differences in the Conflict Probe algorithm. The results of the analysis support this hypothesis, and identify the I of J logic, the higher separation standards, and the more discrete calculations of the CP model as the major influences on the URET vs. TrajCP-Recorded comparison.

VIII. Conclusion

As described previously, the performance of DSTs is an important requirement as an ATC monitors air traffic in order to maintain minimum separation standards among aircraft and airspace. The CP is one major component of a DST, and much research focuses on investigating ways to mitigate the safety risks of increased traffic growth through the enhancement of CP capabilities. This paper presented a new CP model named Trajectory Conflict Probe and its performance compared to an operational CP called URET. Design of experiment modeling techniques was employed to optimize the input parameters of the CP model. Notably it was found that there are benefits in raising

the look ahead time parameter to as high as 25 minutes and increasing the minimum horizontal parameter as high as 9 nautical miles. However, alone this indeed would increase the number of FAs the CP generates. This investigation also found that higher conflict probing frequency with I of J acceptance monitoring logic of conflict predictions helps protect against increased FAs. In general, the I of J logic provided a sense of conflict detection reaching some equilibrium point in which the conflict prediction is stable enough to present to the ATC as an alert to solve a separation problem. Case by case discrepancies were reviewed and reasons for these differences were presented.

Overall, the TrajCP's performance was verified and was found to be a valid model of CP functionality for research and development of new air traffic concepts. Further studies could investigate the performance of TrajCP given a higher fidelity trajectory predictor. Also, future work could involve the improvements of the CP model to include other specifications such as the use of a conflict probabilistic model and other advanced separation management prototyped functionality.

Appendix

The following appendix includes additional figures and tables from the design of experiment modeling performed in the optimization of the TrajCP model.

LH	Add J Ct	Add I % of J	Del J Ct	Del I% of J	Min Hsep	R(MA)	R(FA)	Warn Time	Del Lag Time
600	5	60	10	60	5	0.0598	0.2254	112.9091	89.0756
1500	5	60	10	60	9	0.0165	0.2699	223.1513	217.8776
600	25	60	10	60	9	0.0373	0.1625	155.6466	217.8776
1500	25	60	10	60	5	0.0812	0.1365	82.5581	88.7764
600	5	100	10	60	9	0.0289	0.2114	188.1702	217.8776
1500	5	100	10	60	5	0.0641	0.2206	110.5023	92.0588
600	25	100	10	60	5	0.1073	0.0714	59.1827	85.9494
1500	25	100	10	60	9	0.0413	0.1471	157.8879	219.5122
600	5	60	30	60	9	0.0124	0.2060	224.2259	384.8780
1500	5	60	30	60	5	0.0466	0.1788	148.4444	314.2324
600	25	60	30	60	5	0.0812	0.0733	79.9535	324.8333
1500	25	60	30	60	9	0.0288	0.1449	200.2542	378.8664
600	5	100	30	60	5	0.0511	0.1044	117.1749	318.7552
1500	5	100	30	60	9	0.0165	0.2013	238.0252	377.8862
600	25	100	30	60	9	0.0415	0.1217	158.0087	388.3333
1500	25	100	30	60	5	0.0858	0.0658	76.8545	322.5523
600	5	60	10	100	9	0.0165	0.2067	213.6975	345.8537
1500	5	60	10	100	5	0.0596	0.2163	141.5837	262.8870
600	25	60	10	100	5	0.0940	0.0940	78.5377	260.0840
1500	25	60	10	100	9	0.0288	0.1541	185.9322	346.9636
600	5	100	10	100	5	0.0641	0.1310	112.5114	262.6360
1500	5	100	10	100	9	0.0165	0.2067	219.1176	345.8537
600	25	100	10	100	9	0.0456	0.1288	150.0000	345.8537
1500	25	100	10	100	5	0.0944	0.0786	76.2559	261.4346
600	5	60	30	100	5	0.0381	0.1063	162.1145	404.0909
1500	5	60	30	100	9	0.0041	0.2261	287.5720	467.3279
600	25	60	30	100	9	0.0248	0.1259	208.9831	466.7073
1500	25	60	30	100	5	0.0511	0.0551	132.5112	405.9336
600	5	100	30	100	9	0.0082	0.1632	245.8921	466.7073
1500	5	100	30	100	5	0.0381	0.1028	174.6256	404.2562
600	25	100	30	100	5	0.0901	0.0320	79.8585	407.4583
1500	25	100	30	100	9	0.0329	0.1132	211.2766	467.3279

Table 7: TrajCP with HMTE Setting Optimize Experiment Results.

LH	Add J Ct	Add I % of J	Del J Ct	Del I% of J	Min Hsep	R(MA)	R(FA)	Warn Time	Del Lag Time
600	5	3	10	6	5	0.1268	0.2408	207.9570	72.5381
1500	5	3	10	6	9	0.0385	0.2525	447.8667	204.2982
600	25	15	10	6	9	0.0560	0.1767	262.7397	200.9692
1500	25	15	10	6	5	0.1571	0.2063	225.8757	78.3938
600	5	5	10	6	9	0.0385	0.2021	316.8444	206.1842
1500	5	5	10	6	5	0.1221	0.2490	266.4706	72.0707
600	25	25	10	6	5	0.1893	0.1257	116.8862	77.2193
1500	25	25	10	6	9	0.0649	0.1787	328.5185	199.0708
600	5	3	30	18	9	0.0380	0.1886	371.9737	358.9130
1500	5	3	30	18	5	0.0922	0.2024	342.9442	265.4634
600	25	15	30	18	5	0.1659	0.1200	183.0114	278.8144
1500	25	15	30	18	9	0.0511	0.1710	425.3363	349.2576
600	5	5	30	18	5	0.1163	0.1480	232.2632	283.6816
1500	5	5	30	18	9	0.0380	0.2056	484.8246	354.3043
600	25	25	30	18	9	0.0690	0.1529	246.4352	356.1674
1500	25	25	30	18	5	0.1635	0.1212	237.5862	269.1146
600	5	3	10	10	9	0.0339	0.1972	350.0877	316.6667
1500	5	3	10	10	5	0.1023	0.2311	304.6632	216.0199
600	25	15	10	10	5	0.1675	0.1553	165.5747	224.9474
1500	25	15	10	10	9	0.0513	0.1838	397.2973	297.9039
600	5	5	10	10	5	0.1268	0.1770	211.3978	221.0714
1500	5	5	10	10	9	0.0339	0.2111	455.6579	309.2208
600	25	25	10	10	9	0.0690	0.1563	229.6296	307.1806
1500	25	25	10	10	5	0.1699	0.1535	211.3450	212.3810
600	5	3	30	30	5	0.0946	0.1184	272.3383	356.0577
1500	5	3	30	30	9	0.0333	0.2000	531.5086	429.9571
600	25	15	30	30	9	0.0464	0.1472	312.0796	430.6466
1500	25	15	30	30	5	0.1256	0.1256	294.5213	351.9802
600	5	5	30	30	9	0.0335	0.1630	370.0433	435.7940
1500	5	5	30	30	5	0.0913	0.1348	354.6734	364.3478
600	25	25	30	30	5	0.1827	0.0909	144.1765	354.4737
1500	25	25	30	30	9	0.0598	0.1571	383.8182	427.3362

Table 8: TrajCP with RCTE Setting Optimize Experimental Results.

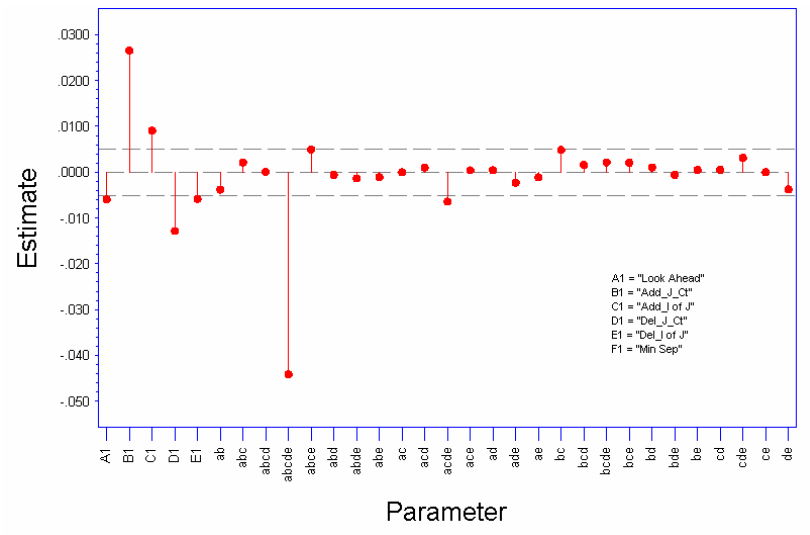

Figure 3: Lenth Plot for R(MA) for TrajCP-HM

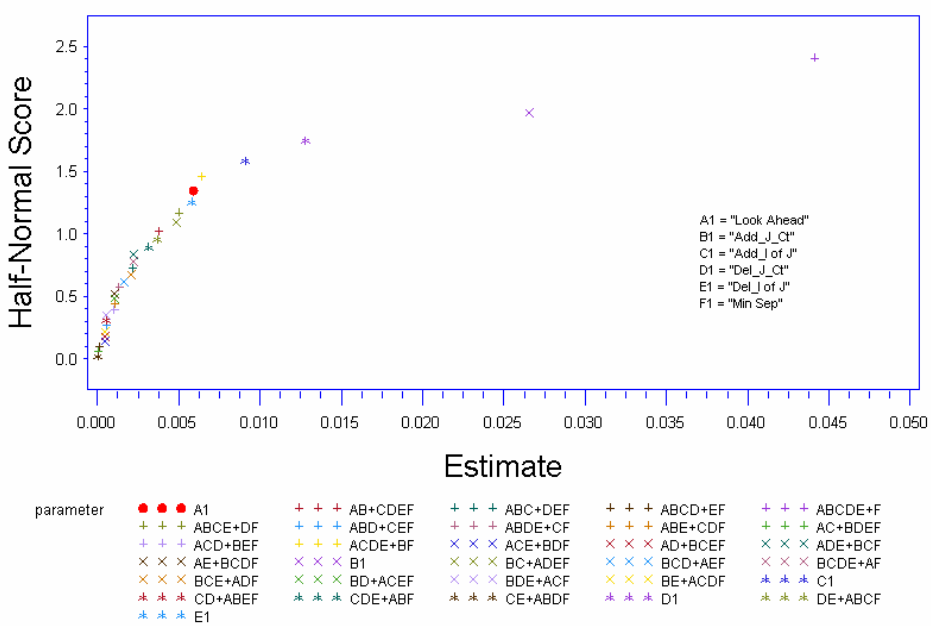

Figure 4: Half-Normal Probability Plot for R(MA) for TrjCP-Recorded

Importance Weights				Opt. Criteria	Optimal Settings							Optimal Response Values			
R(MA)	R(FA)	Warn Time	Del Lag Time	Desirability	LH	Add J Ct	Add I % of J	Del J Ct	Del I% of J	Min Hsep	Min Detect Time	R(MA)	R(FA)	Warn Time	Del Lag Time
4	***4***	4	3	0.538	1474	22.1	97	10.4	61.8	9	-	0.04	0.165	168	216
			2	0.547	1416	13.3	100	25.1	89.1	9	-	0.0212	0.1636	224	413
		3	3	0.546	1487	25	98.5	10	60	9	-	0.044	0.152	`60	206
		3	***2***	***0.5417***	***1475***	***21.3***	***100***	***18.8***	***88.3***	***9***	***-***	***0.0305***	***0.144***	***189***	***355***
	3	3	3	0.554	1494	24.9	66.3	10	60.3	9	-	0.0366	0.177	174	207
		3	2	0.548	1340	11.9	100	21.5	82	9	-	0.0226	0.18	214	360
			1	***0.586***	***1500***	***15***	***100***	***30***	***80***	***9***	***-***	***0.0221***	***0.157***	***226***	***435***

Table 9: TrajCP-HM Sensitivity Analysis of Desirability Function Optimization.

Importance Weights				Opt. Criteria	Optimal Settings							Optimal Response Values			
R(MA)	R(FA)	Warn Time	Del Lag Time	Desirability	LH	Add J Ct	Add I % of J	Del J Ct	Del I% of J	Min Hsep	Min Detect Time	R(MA)	R(FA)	Warn Time	Del Lag Time
4	***4***	4	3	0.571	1490	13.6	96.8	10	80.1	9	-	0.047	0.203	394	254
			2	0.583	1451	9.6	99.3	10.3	98.9	9	-	0.04	0.198	415	305
		3	3	0.569	1488	20.4	98.7	10.68	71.3	9	-	0.056	0.193	358	234
		3	***2***	***0.585***	***1500***	***17.5***	***97.9***	***10.61***	***95.5***	***9***	***-***	***0.049***	***0.186***	***384***	***296***
	3	3	3	0.588	1499	12	99.2	10.2	65.1	9	-	0.048	0.213	394	217
		3	2	0.59	1500	7.3	98.3	10.28	93.7	9	-	0.037	0.205	426	295
		3	**1**	**0.625**	**1500**	**5**	**100**	**15**	**100**	**9**		**0.034**	**0.201**	**452**	**341.5**

Table 10: TrajCP-Recorded Sensitivity Analysis of Desirability Function Optimization.

References

[1] Paglione, M. M., Santiago, C., Crowell, A., and Oaks, R. D. "Analysis of the Aircraft to Aircraft Conflict Properties in the National Airspace System," *Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference*, August 2008.

[2] Cale, M. L, Paglione, M. M., Ryan, H. F., Timoteo, D., and Oaks, R. D., "User Request Evaluation Tool (URET) Conflict Prediction Accuracy Report," *DOT/FAA/CT-TN98/8*, April 1998.

[3] Ryan, H. F., Chandler, G., Santiago, C., Paglione, M. M., Liu, S. "Evaluation of En Route Automation's Trajectory Generation and Strategic Alert Processing: Analysis of ERAM Performance", *DOT/FAA/TC-TN08/10*, November, 2008.

[4] Cale, M. L, Kazunas, S., M. M. Paglione, H. F. Ryan, "User Request Evalution Tool (URET) Algorithm Assessment Report," DOT/FAA/CT-TN97/4, February 1997.

[5] Paglione, M. M., Oaks, R. D., Ryan, H. "Methodology for Evaluating and Regression Testing a Conflict Probe," *AIAA-IEEE Digital Avionics Conference*, October 2004.

[6] Oaks, R. D., Paglione, M. M., "Generation of Realistic Air Traffic Scenarios Using a Genetic Algorithm," *AIAA-IEEE Digital Avionics Conference*, October 2002.

[7] Hicks, Charles R., *Fundamental Concepts in the Design of Experiments, Fifth Edition*, New York, NY, Oxford University Press, 1999.

[8] Daniel, C., "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments," *Technometrics*, Vol. 1, 1959, pp. 311-342.

[8] Lenth, Russel V., "Lenth's Method for the Analysis of Unreplicated Experiments," *Wiley Encyclopedia of Statistics in Quality and Reliability*. URL: http://www.wiley.com/legacy/wileychi/eqr/docs/sample_1.pdf, 2006.

[9] Lenth, Russel V., "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, Vol. 31, No. 4, November 1989, pp. 469-473.

[10] Myers, Raymond H., Douglas C. Montgomery, and Christine M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Hoboken, NJ, John Wiley & Sons, 2009.

[11] Crowell, Andrew, Confesor Santiago, "An Algorithmic Method for Regression Analysis of Conflict Probe Accuracy," *Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference*, August 2009.