7-2013

Validation of New Technology using Legacy Metrics: Examination of Surf-IA Alerting for Runway Incursion Incidents

Robert Edward Joslin
Embry-Riddle Aeronautical University - Worldwide

Scholarly Commons Citation
Joslin, Robert Edward, "Validation of New Technology using Legacy Metrics: Examination of Surf-IA Alerting for Runway Incursion Incidents" (2013). *Dissertations and Theses*. 56.
https://commons.erau.edu/edt/56

VALIDATION OF NEW TECHNOLOGY USING LEGACY METRICS: EXAMINATION OF SURF-IA ALERTING FOR RUNWAY INCURSION INCIDENTS

by

Robert Edward Joslin

A Dissertation Submitted to the College of Aviation
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University
Daytona Beach, Florida
July 2013

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

VALIDATION OF NEW TECHNOLOGY USING LEGACY METRICS: EXAMINATION OF SURF-IA ALERTING FOR RUNWAY INCURSION INCIDENTS

by

Robert Edward Joslin

This Dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Alan J. Stolzer, Professor, Daytona Beach Campus; and Dissertation Committee Members Dr. Dothang Truong, Associate Professor, Daytona Beach Campus; Dr. Haydee M. Cuevas, Assistant Professor, Daytona Beach Campus; and Dr. Antonio I. Cortes, External Member, and has been approved by the Dissertation Committee. It was submitted to the College of Aviation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Aviation

Dissertation Committee:

Alan J. Stolzer, Ph.D.
Committee Chair

Dothang Truong, Ph.D.
Committee Member

Haydee M. Cuevas
Haydee M. Cuevas, Ph.D.
Committee Member

Antonio I. Cortes, Ph.D.
Committee Member

Alan J. Stolzer, Ph.D.
Department Chair, Doctoral Studies

Tim Brady, Ph.D.
Dean, College of Aviation

Robert Oxley, Ph.D.
Associate Vice President of Academics

8-8-2019
Date

ABSTRACT

Researcher: Robert Edward Joslin

Title: VALIDATION OF NEW TECHNOLOGY USING LEGACY METRICS: EXAMINATION OF SURF-IA ALERTING FOR RUNWAY INCURSION INCIDENTS

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2013

New flight deck technology designed to mitigate runway incursions may not be effective in triggering a flight deck *alert* to avoid high speed surface collisions for runway incursions classified as *serious* by legacy metrics. This study demonstrated an innovative method of utilizing expert raters and actual high-risk incidents to identify shortcomings of using legacy metrics to measure the effectiveness of new technology designed to mitigate hazardous incidents. Expert raters were used to validate the Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA) model for providing alerts to pilots to reduce the occurrence of pilot deviation type runway incursion incidents categorized as *serious* (Category A or B) by the FAA/ICAO Runway Incursion Severity Classification (RISC) model.

This study used archival data from Aviation Safety Information Analysis and Sharing (ASIAS) incident reports and video reenactments developed by the FAA Office of Runway Safety. Two expert raters reviewed nine pilot deviation type *serious* runway incursion incidents. The raters applied the baseline minimally compliant implementation of the RTCA/DO 323 SURF-IA model to determine which incidents would have an alerting SURF-IA outcome. Inter-rater reliability was determined by percentage agreement and Cohen's kappa and indicated perfect agreement between the raters who assessed six of the incidents with a SURF-IA *alerting* outcome and three as *non-alerting*. Specific *aircraft states* were identified in the baseline SURF-IA model that precluded an outcome of a Warning or Caution *alert* for all pilot deviation type runway incursion incidents classified as *serious* by the FAA/ICAO RISC model: (a) wrong runway departures, (b) no *alert* if traffic entered runway after ownship lift-off from same runway, and (c) helicopter operations.

The study concluded that the SURF-IA model did not yield an outcome of a Warning or Caution *alert* for all pilot deviation type runway incursion incidents classified as *serious* by the FAA/ICAO RISC model. Even if the SURF-IA model had performed to design, the best it could have achieved would have been a 70% alerting outcome for incidents classified as *serious* by the legacy RISC model metric. In the qualitative analysis both raters indicated that neither the legacy RISC definition of *on-runway* nor the SURF-IA definition was appropriate. Hence, the raters' recommendation was not to adopt either model's definition, but rather develop an entirely new definition through further study. The raters were explicit about the criticality of appropriate and harmonized definitions used in the models.

The different outcomes between the RISC and SURF-IA models may result in misleading information when using the reduction in *serious* runway incursion incidents as a metric for the benefit of SURF-IA technology. It is recommended that prior to using the ASIAS runway incursion data as a metric for the benefit of SURF-IA, the FAA develop a process for identifying and tracking ASIAS reported PD type *serious* runway

incursion incidents which will not trigger a SURF-IA *alert*. Consideration should be made to improving the SURF-IA model technical capabilities to accommodate all possible *aircraft states* that the RISC model would classify as *serious* runway incursion incidents.

ACKNOWLEDGEMENTS

The participation of expert raters and data support from the National Aeronautics and Space Administration-Langley Research Center, MITRE Corporation-Center for Advanced Aviation System Development, and the Federal Aviation Administration Office of Runway Safety were invaluable for this study. The critical review, insights, and sage guidance from the dissertation committee were not only essential to the completion of this dissertation, but also provided the tools to continue to grow as a researcher, scholar, and educator. A special thanks to Embry-Riddle Aeronautical University for their generous support of Veterans education through their full contribution to the Yellow-Ribbon program. Finally, this extended educational endeavor would not have been possible without the support and understanding of my family.

TABLE OF CONTENTS

Appendices

LIST OF TABLES

Page

Table

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Federal Aviation Administration (FAA) predicts that domestic passenger capacity will grow from 731 million in Fiscal Year (FY) 2011 to 1.2 billion in FY 2032 (FAA, 2011a). To accommodate the demand for additional capacity in a safe and efficient manner, the FAA has implemented the Next Generation Air Transportation System (NextGen). NextGen is a comprehensive overhaul of the National Airspace System (NAS) that integrates new and existing communication, navigation, and surveillance technologies. The cornerstone enabling technology to accommodate the growth in the number and frequency of aircraft operations, both in the air and on the ground, will be a satellite-based navigation and surveillance system that implements various applications of Automatic Dependent Surveillance-Broadcast (ADS-B) (FAA, 2011b, 2012a, 2012b).

The FAA forecasts that airport tower operations associated with the increased domestic capacity will increase by 23% between FY 2012 and FY 2032, which corresponds to an increased number of runway operations (FAA, 2011a). As a proactive measure for mitigating runway incursions, in 2011 the FAA issued Advisory Circular (AC) 20-172 -Airworthiness Approval for ADS-B(In) Systems and Applications, which included Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA) as one of the NextGen ADS-B(In) applications (FAA, 2011b).

The deadliest accident in worldwide aviation history occurred on March 27, 1977, and involved a runway incursion and collision of two B-747 aircraft at Tenerife in the Canary Islands resulting in 583 fatalities (http://www.ntsb.gov/safety/mwl-4.html). Since

1990, there have been seven fatal runway incursion accidents in the United States resulting in 112 fatalities, the deadliest of which occurred in 2006 with 49 fatalities (Table 1) (NTSB, 2007).

Table 1

Fatal U.S. runway incursion accidents 1990-2012

Year	Location	Airline/Flight	Aircraft	Fatalities
2006	Lexington, KY	Comair 5191	CL-600	49
2000	Sarasota, FL	N89827/N79960	C-152, C-172	4
1996	Quincy, IL	United Express 5925/N1127D	Beech 1900, King Air A90	14
1994	St. Louis, MO	TWA 427/N441KM	MD 82, C441	2
1991	Los Angeles, CA	USAir 1493/ SkyWest 5569	B-737, Metroliner	34
1990	Detroit, MI	NWA 1482/NWA 299	DC-9, B-727	8
1990	Atlanta, GA	EAL 111/N44UE	B-727, King Air A100	1

Note: Adapted from the NTSB database website
http://www.ntsb.gov/aviationquery/index.aspx and ALPA website
http://www.alpa.org/portals/alpa/runwaysafety/NTSBRunwaySafetyfact sheet.pdf

On June 30, 2010, the FAA chartered an Aviation Rulemaking Committee (ARC) for ADS-B(In) that recommended the implementation of SURF-IA to mitigate runway incursions as one of the top ten ADS-B(In) priorities to support NextGen (ADS-B, 2011). SURF-IA is a new avionics system that alerts pilots of potentially dangerous runway incursions. Runway incursions are defined by the FAA and International Civil Aviation Organization (ICAO) as the incorrect presence of an aircraft on a surface designated for takeoff and landing, and are grouped into three types: (a) operational error/ deviation/ incident (OE/D/I), (b) vehicle-pedestrian deviation (V/PD), and (c) pilot deviation (PD)

(FAA, 2009; ICAO, 2007). For over two decades the FAA has used the rate and number of runway incursions as a metric for measuring runway safety; however, there has not been a significant reduction in runway incursions. The ADS-B(In) ARC proposed using the change in rate of PD type runway incursions as the FAA metric for assessing the benefit and effectiveness of SURF-IA flight deck technology as a mitigation for runway incursions (ADS-B, 2011). Pilot deviation type runway incursions comprised over 60% of all runway incursions, which are the type of runway incursion that SURF-IA was designed to mitigate. The FAA Office of Runway Safety tracks and classifies runway incursions using a Runway Incursion Severity Classification (RISC) model (FAA, 2006, August). However, the RISC model used by the FAA for classification of a runway incursion differs from the Radio Technical Commission for Aeronautics (RTCA) developed model used to activate SURF-IA flight deck *alerts* of a runway incursion that will be displayed to pilots (Cardosi, Hannon, Sheridan, & Davis, 2005; FAA, 2007a; RTCA, 2010). When introducing new technology, the FAA often utilizes international consensus organizations, such as RTCA and the Society of Automotive Engineers (SAE), to develop specifications, requirements, and standards, which the FAA then adopts in part or in toto as regulatory guidance.

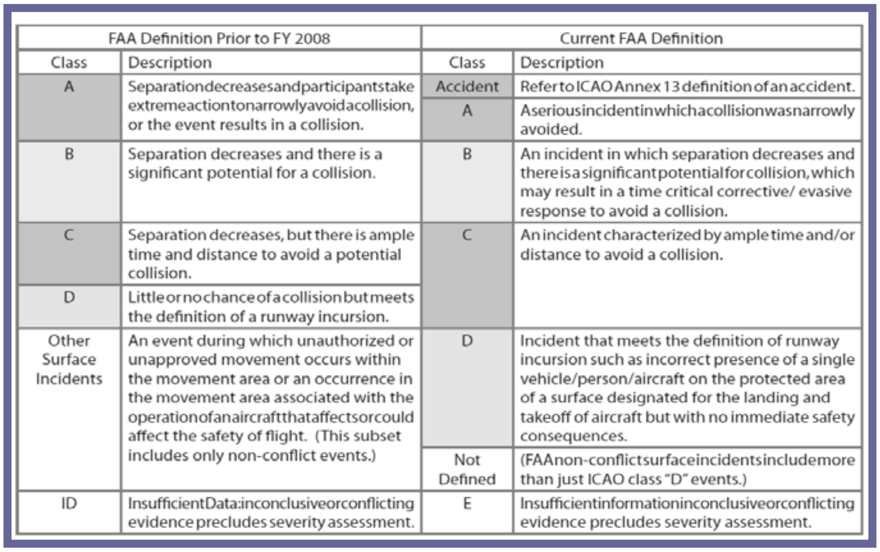
This study evaluated the effectiveness of the SURF-IA model for providing *alerts* to pilots to reduce the occurrence of PD type *serious* (Category A or B) runway incursion incidents as defined by the RISC model. Quantitative data were used to identify the state of ownship (vernacular for pilot's own aircraft) and traffic aircraft for incidents when the outcome severity of the SURF-IA model was not validated by matching the outcome severity of the RISC model. Qualitative data were used to describe specific *aircraft* *states* and factors in the models. The *aircraft states* were defined by the true position of ownship and traffic aircraft in the runway environment. The practical findings of this study may be used to:

(a) specify recommendations for modifications that may enhance the SURF-IA and RISC models, and (b) utilize the modifications to enhance validation of metrics for measuring the benefits of SURF-IA technology. Furthermore, the implications of this study's approach, which assessed the validity of using legacy metrics to measure the effectiveness of new flight deck technology, can be applied to other emerging NextGen flight deck technology that have measurable operational outcomes. Traditional or legacy metrics that have been historically used to assess existing technology may have little or no meaning for the models and algorithms utilized in new technology. For example, Traffic Situation Awareness with Alerts (TSAA) is another NextGen technology that will enhance safety in general aviation (GA) aircraft by providing flight deck alerts to mitigate aircraft mid-air collisions (FAA, 2012a). The FAA (2012c) defines a Near Mid-Air Collision (NMAC) as an incident associated with the operation of an aircraft in which the possibility of collision has been reported by one of the involved flight crew, and results in a recorded proximity of less than 500 feet vertical and 0.5 nautical miles lateral to another aircraft. The effectiveness of TSAA technology will likely be measured through NMAC reports, which are tracked through the Aviation Safety Information and Analysis System (ASIAS); however, the algorithm that triggers flight deck alerts for the proposed TSAA system may have a model with different aircraft proximity thresholds. Hence, an aircraft with TSAA on a flight deck display may be presented with information that safe separation was maintained from another aircraft, but may still precipitate a

NMAC report from the other traffic. This study provided a methodological framework for evaluating legacy metrics to ensure they present valid indications of the safety performance of new technology. The methodology can be extended to other transportation modes as well as medicine, law enforcement, nuclear power-plants, and other safety critical fields.

Runway Incursion Defined

In October 2007, the FAA adopted the ICAO definition of runway incursion as any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designated for the landing and take-off of aircraft. The revised definition expanded the legacy FAA definition to include surface incidents and wrong runway departures. In addition, the revised definition changed some of the descriptors for the categories and types of runway incursions (Figure 1) (FAA, 2009).



FAA Definition Prior to FY 2008		Current FAA Definition	
Class	Description	Class	Description
A	Separation decreases and participants take extreme action to narrowly avoid a collision, or the event results in a collision.	Accident	Refer to ICAO Annex 13 definition of an accident.
		A	A serious incident in which a collision was narrowly avoided.
B	Separation decreases and there is a significant potential for a collision.	B	An incident in which separation decreases and there is a significant potential for collision, which may result in a time critical corrective/ evasive response to avoid a collision.
C	Separation decreases, but there is ample time and distance to avoid a potential collision.	C	An incident characterized by ample time and/or distance to avoid a collision.
D	Little or no chance of a collision but meets the definition of a runway incursion.		
Other Surface Incidents	An event during which unauthorized or unapproved movement occurs within the movement area or an occurrence in the movement area associated with the operation of an aircraft that affects or could affect the safety of flight. (This subset includes only non-conflict events.)	D	Incident that meets the definition of runway incursion such as incorrect presence of a single vehicle/person/aircraft on the protected area of a surface designated for the landing and takeoff of aircraft but with no immediate safety consequences.
		Not Defined	(FAA non-conflict surface incidents include more than just ICAO class "D" events.)
ID	Insufficient Data: inconclusive or conflicting evidence precludes severity assessment.	E	Insufficient information inconclusive or conflicting evidence precludes severity assessment.

Figure 1. FAA definition of runway incursion severity classifications. Adapted from the "National Runway Safety Plan 2009-2011," by the FAA, 2009.

Runway incursions are only reported and tracked by the FAA Office of Runway Safety at airports that have an operating air traffic control tower (FAA, 2009). The FAA defines Category A and Category B runway incursions as *serious* incidents.

Runway Incursion Statistics

FAA runway incursion data for FY 2008 through FY 2012 indicate that there has been an increase in the rate and the number of runway incursions (Figures 2 and 3). Although there have been other types of aviation related accidents and incidents on the runway surface, such as runway excursions and loss of control (LOC) on the ground, both the FAA and the National Transportation Safety Board (NTSB) consider runway incursions as the top surface hazard (FAA, 2012b; NTSB, 2012).

The most prevalent type of runway incursion was caused by a pilot deviation, defined by the Flight Standards Information Management System (FSMIS) as actions of a pilot that resulted in a failure to comply with air traffic control (ATC) clearance and/or instructions (Figure 2). The overall rate of runway incursions per million operations from FY 2008 to FY 2012 increased from 17.16 to 21.02, of which over 60% were PD type in every year (Figure 3). Pilot deviation type runway incursions are the type of runway incursions intended to be mitigated by SURF-IA.

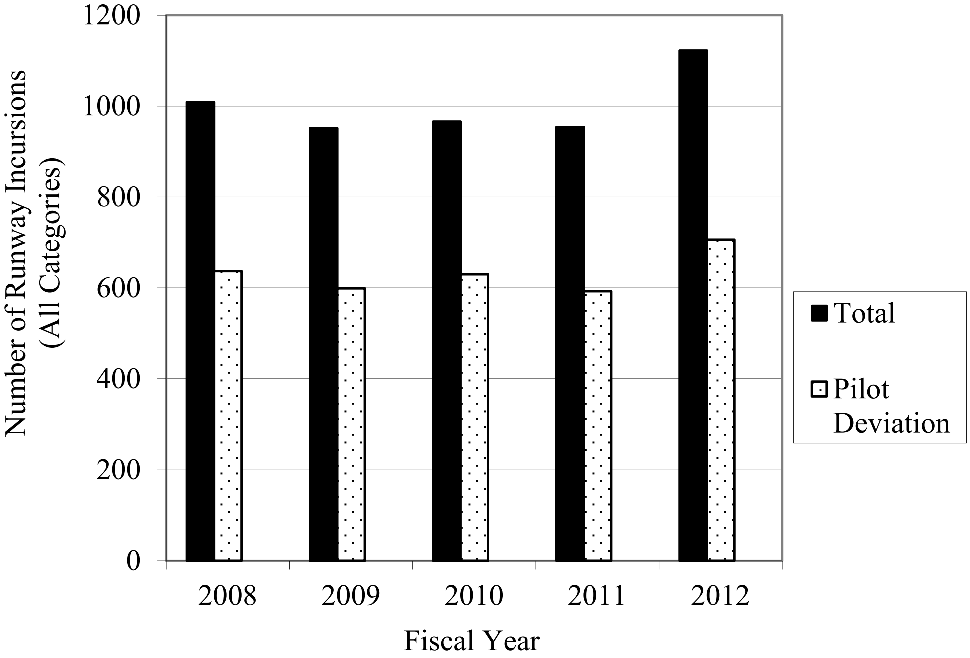


Figure 2. Runway incursions FY 2008-FY 2012. Adapted from http://www.faa.gov/ airports/runway_safety/ statistics

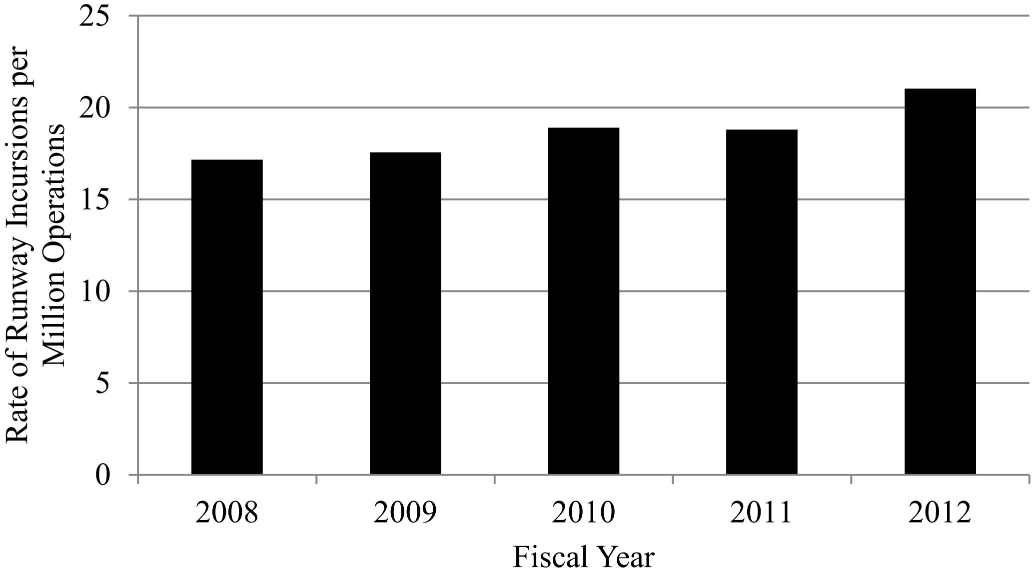


Figure 3. Rate of runway incursions FY 2008-FY 2012. Adapted from http://www/faa.gov/airports/runway_safety/

Benefits of Runway Incursion Mitigation

Aviation rulemaking committees, which are chartered by the FAA, conduct benefit-cost analyses as part of their evaluations of investments in new technology that are subject to FAA decision-making, such as SURF-IA (FAA, 2007b). All costs incurred (or costs avoided) that may result from proposed FAA investments and regulations are considered in the analyses. Avoided accidents are one of the principal benefits and are valued using the avoided injury and property damage costs recommended in the FAA Guide for Economic Values for FAA Investment and Regulatory Decisions (FAA, 2007b). Accurately quantifying benefits of SURF-IA in terms of avoidance of hull loss and fatalities is difficult because, fortunately, there have been few actual runway incursion accidents. However, the ADS-B(In) ARC report estimated the benefits from avoidance of runway incursions from FY 2011-FY 2025 to be $55M for the U.S. domestic air transport community alone (ADS-B, 2011).

> Risk management in aviation illustrates how organizations cooperate, by capturing near miss information to augment the sparse history of crashes and injuries. Data from incident reporting systems on near misses have been effectively used to redesign aircraft, air traffic control systems, airports, and pilot training, and to reduce human error (Barach & Small, 2000, p. 762).

Considering Category A and Category B runway incursion incidents, and not just accidents in the benefit analysis for SURF-IA, would follow historical precedents such as the mandate for Collision Avoidance System (CAS), which considered mid-air collision (MAC) accidents along with near mid-air collision incidents.

The language in the economic analysis in the rulemaking for CAS demonstrated the precedent for using incident data to supplement accident data for the benefit analysis of new flight deck technology (Collision Avoidance Systems, 2003).

> It is assumed that the risk of a near mid-air collision (NMAC) is proportional to the pair probabilities. The risk of a NMAC is used rather than the risk of a MAC, because most of the statistical models used in studying the safety of TCAS II were derived from encounter data and not from MAC data. (CAS, 2003, p.15896)

Based in part on the Collision Avoidance System (2003) precedent, the ADS-B(In) ARC report recommended measuring reduced frequency of pilot deviation type runway incursion incidents associated with the use of SURF-IA flight deck technology, to assess the effectiveness and benefit of the new technology (ADS-B, 2011).

> The ARC recommends the FAA analyze the rate of pilot deviation type runway incursions at the 44 airports where the SURF-IA ADS-B(In) application is initially implemented to assess the application’s benefits. (ADS-B, 2011, p. 45)

SURF-IA Model

The SURF-IA Safety and Performance Interoperability Requirements (SPR) were developed through the Radio Technical Commission for Aeronautics (RTCA) Special

Committee 186 (SC-186) based on studies by Jones and Prinzel (2006), and Prinzel and Jones (2006). Using ADS-B technology, SURF-IA mitigates runway incursions by enhancing pilot situation awareness of other aircraft through a Cockpit Display of Traffic Information (CDTI) (Jones & Prinzel, 2006; Jones, Prinzel, Otero, & Barker, 2009; Moertl & Nickum, 2008; Prinzel & Jones, 2006). SURF-IA technology, enabled by ADS-B(In), was designed for use by pilots operating within the airport surface movement area.

Automatic Dependent Surveillance-Broadcast is the FAA's satellite-based successor to radar. ADS-B makes use of Global Positioning System (GPS) technology to determine and share precise aircraft location information, and streams additional flight information to the flight deck of properly equipped aircraft. It is *automatic* because no interrogation from an external source is required for operation, and it is *dependent* because it relies on on-board equipment to provide surveillance information (i.e., position, altitude, speed, and heading) obtained from a GPS receiver (RTCA, 2010). Any user within line-of-sight broadcast range can receive and process ADS-B messages using an appropriate receiver. The CDTI is enhanced with SURF-IA visual alerts, aural alerts, and indications that highlight traffic and runway status through alphanumeric information and symbology (Figure 4) (Jones & Prinzel, 2006; Jones, Prinzel, Otero, & Barker, 2009; Jones et al., 2010; RTCA, 2010).

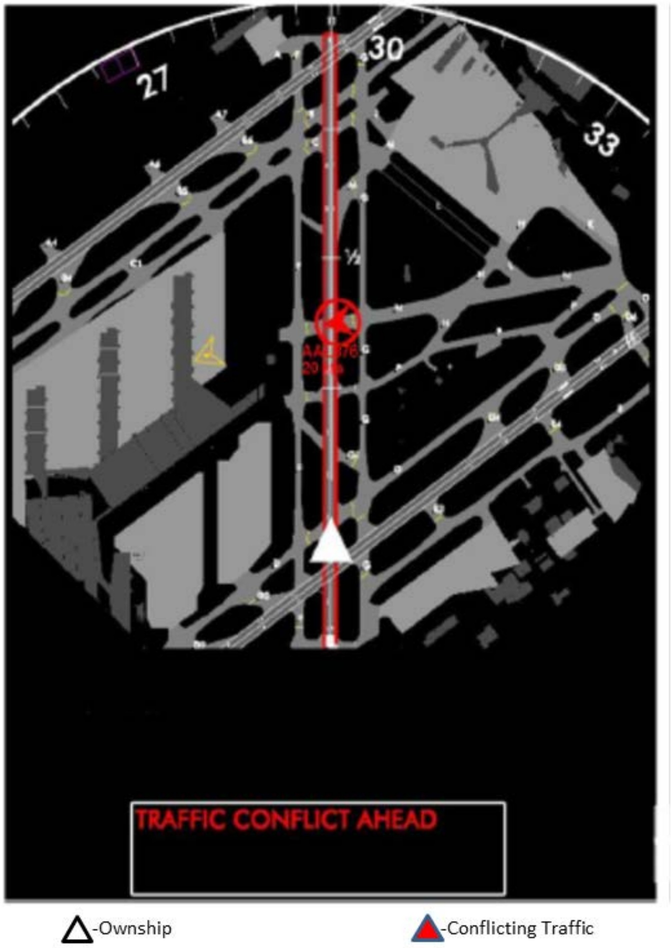


Figure 4. Conceptual cockpit display of a Warning alert from SURF-IA. Adapted from "Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA)," by Radio Technical Commission for Aeronautics (RTCA/DO-323).

The starting point for the Minimum Operational Performance Standards (MOPS) of the model for SURF-IA alerting are set forth in the Safety and Performance Interoperability Requirements (SPR) Document developed by RTCA (2010). The SPR identified the baseline SURF-IA model as the version with limitations to the technical capabilities for

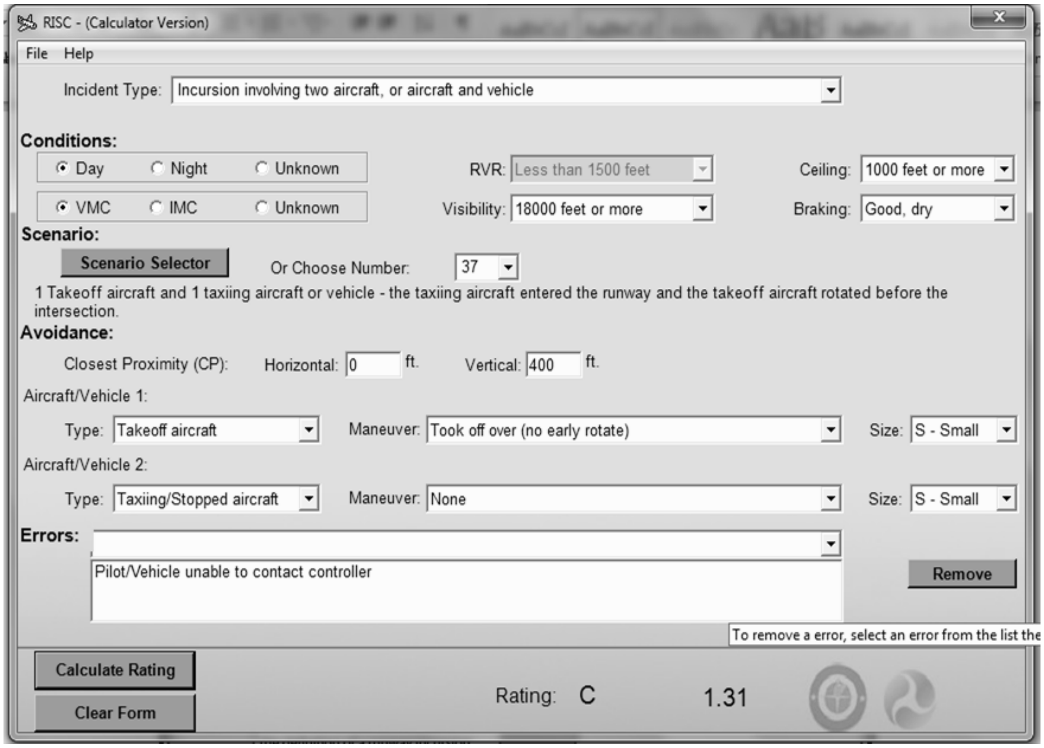
- alerting and indications about potential collisions in airport ramp areas;

- alerting and indications about potential collisions on airport taxiways;
- technological integration between ground-based alerting logic and flight deck based alerting logic;
- SURF-IA for helicopters;
- SURF-IA for ground vehicles;
- explicit consideration of aircraft movement intent information. This version inferred movement intent based on current traffic position and movement (e.g., takeoff, landing, crossing);
- detection of wrong runway usage (including closed runways), unless there is conflicting traffic;
- alerting after lift-off;
- use of CDTI and indications for surface movement efficiency, such as supporting taxi-operations during low visibility conditions;
- directive alerting;
- predictive alerting.

Runway Incursion Severity Classification (RISC) Model

The RISC model was based on the research conducted by Sheridan (2004), Sheridan, Cardosi, and Hannon (2004), and Cardosi, Hannon, Sheridan, and Davis (2005). It has been used since 2006 by the FAA Office of Runway Safety as part of its quality management system (QMS) to validate runway incursion severity classifications (FAA, 2006, August; FAA, 2011a, March; FAA, 2011b, March). A computer program that automated the RISC model was developed by the FAA and VOLPE National

Transportation System Center with the aim of standardizing assessments of runway incursion events among the FAA and ICAO member states (ICAO, 2007). The primary factors considered in the RISC model were horizontal/vertical proximity of the aircraft and/or vehicle/pedestrian, geometry of the encounter, evasive or corrective action, available reaction time, environmental conditions, and factors that affected system performance such as communication failures/errors (Figure 5).



RISC - (Calculator Version)

Incident Type: Incursion involving two aircraft, or aircraft and vehicle

Conditions:

● Day ○ Night ○ Unknown

● VMC ○ IMC ○ Unknown

RVR: Less than 1500 feet

Visibility: 18000 feet or more

Ceiling: 1000 feet or more

Braking: Good, dry

Scenario:

Or Choose Number: 37

1 Takeoff aircraft and 1 taxiing aircraft or vehicle - the taxiing aircraft entered the runway and the takeoff aircraft rotated before the intersection.

Avoidance:

Closest Proximity (CP): Horizontal: 0 ft. Vertical: 400 ft.

Aircraft/Vehicle 1:

Type: Takeoff aircraft Maneuver: Took off over (no early rotate) Size: S - Small

Aircraft/Vehicle 2:

Type: Taxiing/Stopped aircraft Maneuver: None Size: S - Small

Errors:

Pilot/Vehicle unable to contact controller

To remove a error, select an error from the list the

Rating: C 1.31

Figure 5. Runway Incursion Severity Classification calculator. Adapted from the "Manual on the Prevention of Runway Incursions," by ICAO, 2007.

The RISC model was initially developed from research conducted by Sheridan, Cardosi, and Hannon (2004) for rating the severity of close-call events in transportation, medicine, police, and security, and then refined by Sheridan (2004) and Cardosi, Hannon, Sheridan, and Davis (2005) to specifically focus on runway incursions.

The ICAO (2007) Manual on the Prevention of Runway Incursions cited Sheridan (2004) as the governing document that provided the mathematics behind the RISC model for objectively categorizing runway incursion incidents as Category A, B, or C. Sheridan's (2004) so-called interpolation method for rating severity of runway incursions established *a priori* criteria and rules for the RISC model based on the objective factual data and quantitative estimates extracted from ATC runway incursion reports used by the FAA Office of Runway Safety to classify incidents. The model identified a set of independent runway incursion scenarios (e.g., one landing aircraft, one taxiing aircraft) that broadly subsumed all incursions. The baseline severity was determined by closest horizontal or vertical proximity for the given scenario. A smaller set of common scalable factors (e.g., visibility, ceiling, Runway Visual Range (RVR), day/night) rated on a 10-point scale were used to further characterize the severity of the scenario beyond aircraft proximity. For cases where the scalable factors were all zero, closest proximity alone was adequate to characterize the severity of the incident. Each factor's weighted score was used to calculate a final score that determined the severity category. The Runway Incursion Severity Classification (RISC) calculator is a computer program that classifies the outcome of runway incursions into one of three severity categories: "A", "B", or "C". Category D runway incursions are considered non-conflicting (Figure 1). The calculator was populated with data extracted from pilot deviation reports (FAA Form 8020-17) of runway incursions submitted by the ATC personnel who observed the incident (FAA, 2007a, 2007c, 2010a).

In the initial RISC validation, the model ratings matched the ratings from the FAA Office of Runway Safety in only 67% of the cases. Hence, Cardosi, Hannon, Sheridan, and Davis (2005) recommended the incorporation of the following improvements to the RISC model prior to formal implementation by the FAA and ICAO: (a) only consider closest proximity achieved unintentionally; (b) not consider an incident where ATC directed participating aircraft to intentionally and knowingly taxi toward each other on the same runway; and (c) classify aircraft fly-over or land-over scenarios as Category B instead of Category C.

Following the aforementioned model adjustments, the FAA (2006, August) assessed the RISC as valid and reliable and formalized its use by the FAA Office of Runway Safety Quality Management System (QMS) process for runway incursion severity classification (FAA, 2011a, March; FAA, 2011b, March). The International Civil Aviation Organization (2007) also formalized the use of the Runway Incursion Severity Classification (RISC) computer program for modeling and standardizing the classification of outcomes from runway incursions to provide consistent ratings by applying the same decision processes used by expert raters from the FAA and ICAO (FAA, 2006, August; FAA, 2011a, March; FAA, 2011b, March; ICAO, 2007). ICAO asserted that, "such consistency is deemed essential for being able to examine trends over time or see the effects of mitigation strategies" (ICAO, 2007, p. H-1).

RISC and SURF-IA Model Comparison

The RISC model was based primarily on *aircraft state,* environmental factors, and non-temporal quantitative factors for closest horizontal or vertical (*overflight*) proximity and could consider ATC intervention for intentional incursions. The SURF-IA model for *alerts* relied on GPS derived quantitative *aircraft state* factors that considered horizontal and vertical proximity as well as temporal closure considerations. However, the proximity information entered into the RISC model from ATC deviation reports at times was based on subjective observations of incidents with regard to how close two aircraft came to colliding, rather than the precise GPS derived instrument readings for position and time that were used in SURF-IA (GAO, 2007). An overview of the primary model factors is provided in Table 2.

Table 2

Comparison of primary model factors for RISC and SURF-IA

Factors for Model	FAA RISC Model	SURF-IA Alerts Model
Horizontal Separation	√	√
Vertical Separation	√	√
Aircraft Geometry	√	√
Runway Visual Range	√	×
Ceiling/Visibility	√	×
Braking Condition	√	×
Closure Rate	×	√
Day/Night	√	×
VMC/IMC	√	×
Aircraft Size	√	×
Aircraft Maneuver	√	√
Human Errors	√	×
ATC Intervention	√	×
On-Runway Criteria	Hold Short Line	Runway Shoulder
Distance from Runway	< 1 mile from runway threshold	≤ 35 secs to runway threshold

√ Considered × Not considered

Note: Adapted from “Safety, performance and interoperability requirements document for enhanced traffic situational awareness on the airport surface with indications and alerts (SURF-IA),” by Radio Technical Commission for Aeronautics (RTCA), 2010, and “A method for rating the severity of runway incursions,” by Cardosi, et al., *Proceedings of the USA/Europe 6th Air Traffic Management Research and Development Seminar*, 2005.

Other comparisons between SURF-IA *alerting* model and the RISC model for a *serious* (Category A or Category B) classification that could affect the outcomes were as follows:

1) SURF-IA model will *alert* when conflicting aircraft is within 35 seconds of the runway threshold (RTHRE), while the RISC model considers traffic at less than one nautical mile (nm) from the runway threshold as a *serious* runway incursion.
2) Even when a take-off clearance was cancelled by ATC, SURF-IA will *alert* when conflicting traffic was on the same runway with either aircraft moving at greater than 40 knots with closure. The RISC model accounted for air traffic controller instructions and interventions.
3) SURF-IA will *alert* for any conflicting traffic operating below 1000 feet above the airfield elevation (AFE) with horizontal and/or vertical closure, while the RISC model for a *serious* incident typically requires less than 4000 feet horizontal separation or less than 100 feet vertical *overflight* separation.
4) The SURF-IA model defined an on-runway condition for an aircraft not lined up with the runway as any part of the aircraft inside the runway shoulder. The SURF-IA model also considered an aircraft to have met the on-runway condition when it was approximately lined-up with the runway and was within one runway width of the runway centerline (Figure 6). However, the FAA (2010a) Runway Safety Program and the RISC model defined an on-runway condition to be

when any part of the aircraft was inside the runway hold position markings (i.e., hold line), which was encompassed by the runway safety area (RSA). A comparative depiction of the on-runway condition for the SURF-IA model and RISC model is provided in Figure 7 (FAA, 1989, 2010a; RTCA, 2010).

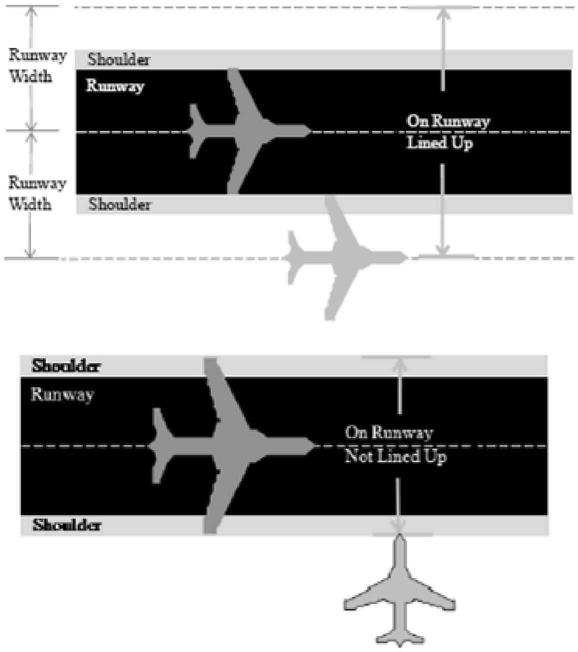


Figure 6. SURF-IA on-runway conditions. Adapted from “Safety, performance and interoperability requirements document for enhanced traffic situational awareness on the airport surface with indications and alerts (SURF-IA),” by RTCA, 2010.

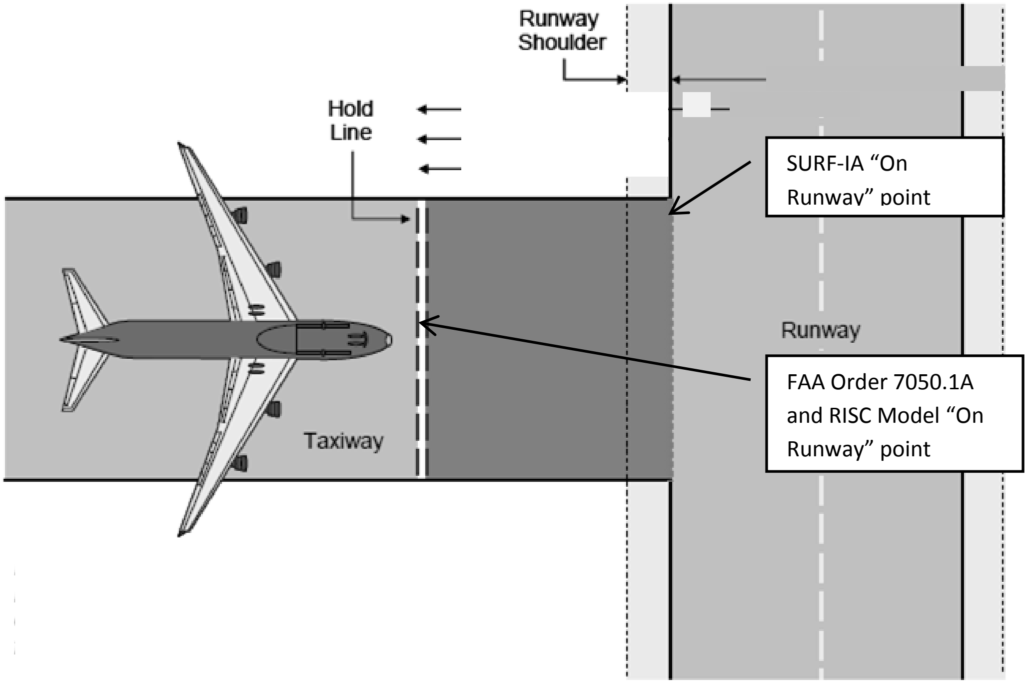


Figure 7. U.S. airport surface geometry on-runway points. Adapted from "Safety, performance and interoperability requirements document for enhanced traffic situational awareness on the airport surface with indications and alerts (SURF-IA)," by RTCA, 2010, "Runway Safety Program," by FAA, 2010."Airport Design," by FAA, 1989.

Statement of the Problem

The SURF-IA model may not trigger a flight deck Warning or Caution *alert* to prompt pilot action for pilot deviation type runway incursions classified by the FAA/ICAO RISC model as s*erious* (Category A or Category B) to avoid a potential high speed surface collision between two aircraft.

Purpose Statement

The purpose of this study was to use actual runway incursion incidents to validate the SURF-IA model for providing *alerts* to pilots to reduce the occurrence of pilot deviation type *serious* (Category A or B) runway incursion incidents as defined by the

RISC model. The study used two expert raters to determine if the SURF-IA model would have provided an *alert* for the scenarios from archived historical ASIAS reports of pilot deviation type runway incursion incidents classified as *serious* (Category A or B) by the RISC model.

Research Questions

Does the SURF-IA model yield an outcome of a Warning or Caution *alert* for runway incursion incidents classified as *serious* (Category A or B) by the FAA/ICAO RISC model? What are the *aircraft states* in the SURF-IA model that preclude an outcome of a Warning or Caution *alert* for runway incursion incidents classified as *serious* (Category A or Category B) by the FAA/ICAO RISC model?

Significance of the Study

For over a decade, the reduction of runway incursions has been a top strategic objective for the FAA and also identified as one of aviation's most critical continuing challenges by the National Transportation Safety Board, European Aviation Safety Agency (EASA), and the International Civil Aviation Organization (EASA, 2010; FAA, 2012b; ICAO, 2007; NTSB, 2012). From FY 2005-FY 2010 over 60% of runway incursions were from pilot deviations (FAA, 2009, 2011c). The FAA opens an enforcement investigation upon receipt of a pilot deviation report (FAA Form 8020-17) for a runway incursion that involves possible regulatory violations (FAA, 2007a, 2007c, 2010a). If the investigation reveals a violation of an FAA regulation, the pilot may be subject to legal enforcement action such as pilot certificate action or civil penalty (FAA, 2007c, 2010a, 2010c).

One of the NextGen technological mitigations for runway incursions will be flight deck *alerts* using SURF-IA. The metric to measure the benefit of SURF-IA as a runway incursion mitigation strategy will be assessed by the FAA Office of Runway Safety through analysis of statistics for change in the rate of pilot deviation type runway incursions using the legacy RISC criteria (ADS-B, 2011). Different outcome severities from the RISC and SURF-IA models, when applied to the same runway incursion incident, may result in misleading information when using the reduction in runway incursion incidents classified as *serious* by the RISC model as the metric for assessing the effectiveness and benefit of SURF-IA technology. The outcome differences may also result in a pilot not receiving a Warning or Caution *alert* from the SURF-IA flight deck technology for an event that would be categorized and reportable to the FAA by the RISC model as a *serious* runway incursion, and may result in legal disputes over FAA enforcement actions for runway incursions from pilot deviations (OSC, 2008).

Delimitations

Only pilot deviation (PD) type runway incursion incidents of Category A and Category B that occurred in the United States between FY 2007-FY 2012 for which the FAA Office of Runway Safety created video-reenactments were considered in this study. The SURF-IA logic for this study only considered *alerts* and did not include or consider SURF-IA *indications*. Hence, the raters only rated whether or not a SURF-IA *alert* would have been triggered, without considering whether or not a runway incursion would have been avoided. The study did not evaluate the physical location of the SURF-IA *alert* annunciation on the CDTI in the pilot's field-of-view, navigational positional accuracy, or aircraft deceleration/braking performance.

For the SURF-IA model, any incident that would have triggered an *alert* (Warning or Caution) was coded as an *alerting* incident and considered equivalent to a Category A or Category B *serious* runway incursion from the RISC model. Any other outcome was coded as a *non-alerting* incident, hence not a *serious* incident. OE/D/I incidents were used in the pilot study as surrogates for PD type incidents in order to maximize the number of PD type incidents available for the main study.

Limitations and Assumptions

This analysis assumed that the flight deck SURF-IA *alerting* system assessed in this study complied with the baseline minimum performance requirements stipulated in the Safety, Performance, and Interoperability Requirements (SPR) Document RTCA/DO 323 developed by RTCA (2010) with the known limitations as published in the SPR (Table 3). The FAA conducted a study on the risk mitigation related to ADS-B surface detection performance issues in support of future ADS-B surface applications such as SURF-IA, and identified necessary improvements to the airport ground infrastructure to resolve operational SURF-IA issues (FAA, 2012d). As of 2012, the FAA had identified 44 airports already outfitted or in the process of being equipped with the necessary ground infrastructure to resolve the operational SURF-IA issues of multi-path, drop-outs, line-of-sight, and position accuracy on the ground (FAA, 2012a, 2012d). This study assumed that the airports where the incidents occurred had the overall infrastructure necessary to support aircraft operating with minimally compliant SURF-IA systems that would *alert* per RTCA/DO-323.

Table 3

Capabilities not in baseline SURF-IA

Alerting and indications about potential collisions in airport ramp areas
Alerting and indications about potential collisions on airport taxiways
Technological integration between ground-based alerting logic and flight deck based alerting logic
SURF IA for helicopters
SURF IA for ground vehicle
Explicit consideration of aircraft movement intent information. This version inferred movement intent based on current traffic position and movement (e.g., take-off, landing, crossing)
Detection of wrong runway usage (including closed runways) in absence of a conflict traffic
Alerting after lift-off
Use of CDTI and indications for surface movement efficiency such as supporting taxi-operations during low visibility conditions
Directive alerting
Predictive alerting

Note: Adapted from “Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA),” by Radio Technical Commission for Aeronautics (RTCA/DO-323), 2010.

By definition, a runway incursion incident classified as *serious* (Category A or B) required an immediate or time critical response by the pilot in either aircraft, which is analogous to SURF-IA *alerts* (Warning or Caution) that also require a similar pilot response as shown in Table 4. In the event a runway incursion is in progress, or about to occur, incursion detection and aural/graphical alerting on the flight deck by the SURF-IA model allows evasive or corrective action to be taken immediately (Green, 2006). Hence, it was assumed that a *serious* (Category A or Category B) runway incursion outcome from the RISC model was equivalent, in terms of potential outcome severity, to a SURF-IA *alert* outcome. The model for SURF-IA used the 14CFR §25.1322 definition of an *alert*, modified to only consider Warnings and Cautions and not Advisories (FAA, 2010b, RTCA, 2010). During the development of the SURF-IA model, Jones and Prinzel (2006) mapped Category A and Category B *serious* runway incursion incidents to SURF-IA alerts, which was consistent with the U.S.C. 14CFR §25.1322 and associated FAA (2010b) Advisory Circular definition for a Warning or Caution *alert* (Table 4). Whereas the FAA (2010b) defined three possible levels of alerting (Warning, Caution, and Advisory), SURF-IA *alerts* only considered Warnings and Cautions, which are annunciated for traffic in non-normal operational conditions when a conflict is predicted on the airport surface (RTCA, 2010). SURF-IA also provided Runway Status Indications (RSI) and Traffic Indications (TI) for traffic in normal operational conditions, consisting only of runway and/or traffic highlighting on a CDTI with no aural annunciations. The RSI and TI *indications* are intended to remind the pilot to verify runway status prior to proceeding and to increase the flight crews' situational awareness about particular relevant traffic that could affect runway safety, but do not require any time critical or immediate action.

Table 4

Runway incursion category vs. SURF-IA Alerting

FAA Definition of *Serious* Runway Incursion	SURF-IA Definition of *Alerting* Runway Incursion
Category A: A serious incident in which a collision was narrowly avoided	Warning *alert*: Requires immediate flight crew awareness and immediate flight crew response
Category B: An incident in which separation decreases and there is significant potential for a collision, which may result in time critical corrective/evasive response to avoid a collision	Caution *alert*: Requires immediate flight crew awareness and subsequent flight crew response

Note: Adapted from "Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA)," by Radio Technical Commission for Aeronautics (RTCA/DO-323), 2010, and from "Manual on the prevention of runway incursions" by ICAO, 2007.

The following statistical assumptions for inter-rater reliability were made based on Cohen (1960): (a) coding of either *alerting* or *non-alerting* was mutually exclusive, independent, and collectively exhaustive; (b) runway incursion incidents were independent events; and (c) each rater generated a rating without knowledge and without influence of the other rater's rating.

Disclaimer

The research presented in this study was solely from the author and does not represent an official position of the Federal Aviation Administration or the Department of Transportation. The use of trademarks or names of manufacturers in this study was for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers.

Definition of Terms

Advisory	Requires flight crew awareness and may require subsequent flight crew response
Alert	A generic term to describe a flight deck annunciation, meant to attract the attention of, and identify to the flight-crew a non-normal operational or aircraft system condition.
Automatic Dependent Surveillance Broadcast	A function on an aircraft or surface vehicle operating within the surface movement area that periodically broadcasts its state vector (horizontal and vertical position, horizontal and vertical velocity) and other information. ADS-B is automatic because no external stimulus is required to elicit a transmission. It is dependent because it relies on on-board navigation sources and on-board broadcast transmission systems to provide surveillance information to other users.
Category A	A serious runway incursion incident in which a collision was narrowly avoided.
Category B	A runway incursion incident in which separation decreases and there is significant potential for a collision, which may result in time critical corrective/evasive response to avoid a collision.
Category C	A runway incursion incident characterized by ample time and/or distance to avoid a collision.
Category D	An incident that meets the definition of runway incursion such as the incorrect presence of a single vehicle/person/aircraft on the protected area of a surface designated for the landing and take-off of aircraft but with no immediate safety consequences.
Caution	Requires immediate flight crew awareness and a less urgent subsequent flight crew response than a Warning alert.
Convergence	Progressively decreasing distance between ownship and traffic determined either by position reports, or velocity and directionality
False Alert	An incorrect or spurious alert caused by failure of the alerting system including the sensor.
Fly-over	When an aircraft attempts to land on the same runway and aborts the landing and flies over the traffic.
Indication	Identify to the flight crew a normal operational condition that could become a runway safety hazard. Indications do not actively attract attention from flight crews but provide enhanced situation relevant information to facilitate flight crew perception of safety hazards (Indications are not Alerts).
Land-over	When an aircraft attempts to land on the same runway and lands over the traffic.
Missed Alert	Condition where an alert is needed but not provided.
Nuisance Alert	An alert generated by a system that is functioning as designed but is inappropriate or unnecessary for the particular condition.
Ownship	Vernacular for pilot's own aircraft.

Pilot Deviation	Actions of a pilot that resulted in a failure to comply with air traffic control (ATC) clearance and/or instructions.
Runway Excursion	When an aircraft on the runway surface departs the end (overrun) or the side (veer-off) of the runway surface during a take-off or landing.
Runway Incursion	Any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designated for the landing and take-off of aircraft.
Warning	Requires immediate flight crew awareness and immediate flight crew response.

List of Acronyms

AC	Advisory Circular
ACSS	Aviation Communication and Surveillance Systems
ADS-B	Automatic Dependent Surveillance-Broadcast
AFE	Airfield Elevation
ALPA	Air Line Pilots Association
ARC	Aviation Rulemaking Committee
ASDE-X	Airport Surface Detection Equipment-Model X
ASIAS	Aviation Safety Information Analysis and Sharing
ASRS	Aviation Safety Reporting System
ATC	Air Traffic Control
ATCAM	Airport Traffic Collision Avoidance Model
BI	Bias Index
CAS	Collision Avoidance System
CDTI	Cockpit Display of Traffic Information
CFR	Code of Federal Regulations
CHS	Charleston International Airport
CY	Calendar Year
DAB	Daytona Beach International Airport
DEN	Denver International Airport
DFW	Dallas-Fort Worth International Airport
DO	Document
DVT	Phoenix Deer Valley Airport

EASA	European Aviation Safety Agency
ERAU	Embry-Riddle Aeronautical University
FAA	Federal Aviation Administration
FY	Fiscal Year
FSIMS	Flight Standards Information Management Systems
GA	General Aviation
GBA	Ground-Based Alerting
GFK	Grand Forks International Airport
GPS	Global Positioning System
HPN	White Plains/Westchester County Airport
HQR	Handling Qualities Rating
HWO	Hollywood/North Perry Airport
ICAO	International Civil Aviation Organization
IMC	Instrument Meteorological Conditions
IRB	Institutional Review Board
JFK	New York/John F. Kennedy International Airport
LACM	Low Altitude Conflict Monitor
LAHSO	Land and Hold Short
LOC	Loss of Control
MAC	Mid-Air Collision
MKE	Milwaukee/General Mitchell International Airport
MOPS	Minimum Operational Performance Standards
NAS	National Airspace System

NASA	National Aeronautics and Space Administration
NextGen	Next Generation Air Transportation System
nm	Nautical Mile
NMAC	Near Mid-Air Collision
NPRM	Notice of Proposed Rulemaking
NTSB	National Transportation Safety Board
OE/D/I	Operational Error/Deviation/Incident
PABAK	Prevalence-Adjusted Bias-Adjusted Kappa
PD	Pilot Deviation
PDARS	Performance Data Analysis and Reporting System
PHL	Philadelphia International Airport
PI	Prevalence Index
QMS	Quality Management System
RIAAS	Runway Incursion Advisory and Alerting System
RIPS	Runway Incursion Prevention System
RISC	Runway Incursion Severity Classification
RSA	Runway Safety Area
RSI	Runway Status Indication
RSM	Runway Safety Monitor
RTCA	Radio Technical Commission for Aeronautics
RTHRE	Runway Threshold
RVR	Runway Visual Range
SAE	Society of Automotive Engineers

SC	Special Committee
SEA	Seattle-Tacoma International Airport
SPR	Safety and Performance Interoperability Requirements
SPSS	Statistical Package for the Social Sciences
SSAP	Surface Surveillance Analysis Platform
SURF-IA	Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts
TCM	Taxi Conflict Monitor
TCAS	Traffic Alert and Collision Avoidance System
TI	Traffic Indication
TSAA	Traffic Situation Awareness with Alerts
VMC	Visual Meteorological Conditions
V/PD	Vehicle/Pedestrian Deviation
WAL	Wallops Flight Facility

CHAPTER II

REVIEW OF THE RELEVANT LITERATURE

The literature review indicated a gap in the body of knowledge in three areas relevant to this study as related to SURF-IA model outcomes: assessment of inter-rater reliability, validity of scenarios, and validity of runway incursion type to determine SURF-IA outcomes relevant to a cockpit display of traffic information (CDTI). None of the other studies in the literature review focused exclusively on PD type runway incursion incidents or *serious* (Category A and Category B) incidents. This study was confined to pilot deviations, which were the type of incidents the FAA has proposed to use as a metric to assess the benefit of SURF-IA. Furthermore, previous SURF-IA studies that rated outcomes did not provide any measure of inter-rater reliability, and researchers conducted their analyses primarily based on generalized data from scenarios said to be representative of actual runway incursion incidents. This study improved on the reliability and validity of previous studies by measuring inter-rater reliability, and utilizing detailed video reenactments of actual *serious* runway incursion incidents classified and recorded by the FAA in ASIAS. ASIAS is an on-line data and information sharing repository consisting of 131 databases related to aviation safety and aviation standards (FAA, 2012e). One of the databases is populated and maintained by the FAA Office of Runway with data from runway incursion incident reports that have been categorized for severity in accordance with the RISC model.

SURF-IA Model Development

SURF-IA flight deck technology for runway incursion *alerting* was developed in response to NTSB safety recommendations starting in 1973, and reinforced by the NTSB in 1990 when the hazard of airport runway incursions was first placed on its list of *Most Wanted Transportation Safety Improvements* (NTSB, 2012). Runway incursions have remained on the list every year since, culminating in the specific NTSB recommendation to install flight deck technology for runway incursion *alerting* leading to the development of SURF-IA as one of the NextGen ADS-B(In) applications.

> Require that all *14 Code of Federal Regulations Part 91K, 121, and 135* operators install on their aircraft cockpit moving map displays or an automatic system that alerts pilots when a takeoff is attempted on a taxiway or a runway other than the one intended. (NTSB, 2007, p. 94)

SURF-IA for runway incursion *alerting* on the flight deck required the maturation of the Global Positioning System (GPS) as an enabling component to provide precise position and timing information; hence, research in this area is some of the most recent (Young & Jones, 2001). The literature review in this area focused on the development of the model for SURF-IA *alerting* and not for SURF-IA *indications*.

The concept for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA) was defined by RTCA Special Committee (SC-186) Automatic Dependent Surveillance-Broadcast (ADS-B) in support of the FAA implementation of ADS-B technologies to mitigate runway incursions (RTCA, 2010). The *alerting* model for SURF-IA was adapted by SC-186 through analysis conducted under government contract by the MITRE Corporation. The SURF-IA model was based on the Runway Safety Monitor (RSM) algorithm initially developed as part of the NASA Runway Incursion Prevention System (RIPS) research. The genesis of the SURF-IA model stemmed from NTSB recommendations for the FAA to implement flight deck technology that provided immediate warnings of probable runway incursions directly to the flight-crew (Moertl & Nickum, 2008; P. Moertl, personal communication, October 8, 2012; NTSB, 2000, 2007). A runway incursion *alert,* as defined by RSM, was not necessarily a warning of an impending collision but rather a means of notifying the pilot that a hazardous situation on the runway was detected so that evasive action could be taken to avoid an accident (Green, 2002).

NASA RIPS encompassed three different technologies under the Airport Traffic Collision Avoidance Monitor (ATCAM): (a) Taxi Conflict Monitor (TCM) for ground taxi conflicts anywhere in the airport movement and ramp areas; (b) Low Altitude Conflict Monitor (LACM) for air-to-air conflicts; and (c) Runway Safety Monitor (RSM) for runway incursion conflicts (Green, Otero, Barker, & Jones, 2009). The first RSM flight demonstrations were conducted at Dallas-Fort Worth International Airport (DFW) alongside two other candidate runway incursion alerting models: Runway Incursion Advisory and Alerting Systems (RIAAS), and Ground-Based Alerting (GBA) system. All three models considered the operational state of ownship and traffic determined by the location relative to the runway, speed, track angle, and acceleration. The primary difference between the RSM model and the other candidate models, which were subsequently dismissed, was that RSM detection of incursions considered other criteria such as aircraft position within a three-dimensional virtual protection zone around a runway that was being used by the ownship, along with separation, and closure rate between ownship and conflicting traffic. The other considerations incorporated into the algorithms for RIAAS and GBA were unique criteria associated with specific scenarios (Young & Jones, 2001). In the initial demonstration flights at Dallas-Fort Worth International Airport (DFW) the RSM algorithm *alert* generation rate was 91%, yielding four missed alerts and four false alerts, with all flight profiles following generic scenarios (Jones, Quach, & Young, 2001). Generic scenarios were a subsumed amalgamation of many different runway incursion incidents; hence, they could not be directly mapped to a specific actual runway incursion incident recorded in ASIAS.

The DFW test only involved single runway incursion scenarios. Consequently, the RSM model was enhanced to accommodate crossing runways and intersecting flight paths, and adjusted to reduce the number of false and missed alerts (Cassell, Evers, Esche, & Sleep, 2002). In March 2002, a full mission simulation at NASA Langley Research Center evaluated the enhanced and improved RSM incursion detection algorithm and associated alerting concepts, while once again only used flight profiles with generic scenarios (Jones, 2002). In 2004 the revised RSM model was flight tested at Wallops Flight Facility (WAL) alongside a runway incursion decision algorithm called PathProx, which like RIAS and GBA used a scenario based *alerting* scheme (Jones, 2004). The RSM algorithm *alert* generation rate was 100% with no missed or false alerts, while the scenario-based PathProx only alerted 41% of the time when an *alert* was expected (Jones, 2004). The research concluded that the RSM model would significantly enhance runway safety, but should be validated with further simulations and flight demonstrations.

Following the Wallops flight demonstration, another simulation study by Prinzel and Jones (2006) tested the RSM utilizing runway incursion scenarios developed by NASA, which were then categorized by the RISC model as Category A, B, C, or D through an independent analysis by three raters from the FAA Office of Runway Safety.

Flight demonstrations of the SURF-IA alerting model based on the improved RSM model were flown in 2009 by Aviation Communication and Surveillance Systems (ACSS) under contract with the FAA at Philadelphia International Airport (PHL), and evaluated for technical feasibility and safety effectiveness using six generic *alerting* scenarios (ACSS, 2010). This was followed by an additional demonstration flown with generic scenario profiles by Honeywell in 2010 at Seattle-Tacoma International Airport (SEA), also under FAA contract, to refine and mature the model prior to incorporation into the Safety, Performance, and Interoperability Requirements (SPR) document for SURF-IA (Honeywell, 2010; RTCA 2010).

The literature review indicated that previous research, simulations, and demonstrations of SURF-IA had all followed the accepted practice of benchmarking performance of conflict alerting algorithms using generic conflict scenario profiles, and not data from specific actual runway incursion incidents (Latimer, 2012). Latimer's (2012) research on creating a conceptual detection and avoidance model recognized the value of using actual incidents to examine outcomes of conflict alerting models and even presented a mix of generic scenarios and actual incidents; however, the study ultimately only utilized the generic scenarios. An analysis by Moertl, Lascara, Higgins, and Baker (2012, June) to estimate the safety benefits of SURF-IA based on the minimum RTCA (2010) SURF-IA requirement utilized data from a set of 24 historical Category A and

Category B runway incursions from FY 2007 of which 12 were PD type. The runway incursions were reconstructed from limited available information, and required detailed assumptions about aircraft movement and timing based on aircraft typical performance characteristics, such as aircraft velocity and distance travelled down the runway during take-off/landing, and aircraft speed and altitude during an approach to a landing. The study considered both *indications* and *alerts,* and used three raters who assessed the effect of SURF-IA on reducing the severity of runway incursion incidents (i.e., Category A to Category B or Category B to Category C); however, no measures of inter-rater reliability were presented and the study utilized the FAA definition of runway incursion prior to it being harmonized with ICAO in 2008. Moertl et al. (2012, June) concluded that only 33% of the pilot deviation type incidents would have provided either a SURF-IA *alert* or *indication*. Lascara and Moertl (2012) subsequently developed a software tool called the Surface Surveillance Analysis Platform (SSAP) to determine, verify, and validate SURF-IA outputs from historic runway incursions; however, SSAP used outputs different than required by minimally compliant RTCA (2010) SURF-IA technology. Only four runway incursion scenarios were analyzed of which only two were from actual incidents: one operational error (O/E) and one pilot deviation (PD). Both incidents were classified as Category C runway incursions and one triggered a SURF-IA *alert*. The most recent SURF-IA study by Jones et al. (2012) evaluated the SURF-IA algorithm at various levels of horizontal position accuracy for seven runway incursion scenarios and did not focus on *alerts* or PD type incidents. Although some of the analyses used in the development of SURF-IA utilized expert raters, none reported any measure of inter-rater reliability.

Expert Raters and Inter-rater Reliability

There was abundant literature addressing inter-rater reliability with seminal work by Cohen (1960) and Fleiss (1971), which had been used extensively in a variety of studies where models were rated by experts. The relevant statistical literature for this study revolved around the precedent for rating categorical data with two raters and measuring inter-rater reliability using percentage agreement and Cohen's kappa.

The use of expert raters and measures of inter-rater reliability to validate models has been used extensively in research, primarily in medicine and social science. However, there was no *de facto* requisite number of expert raters, which have ranged from a minimum of two up to dozens of raters. There was general agreement in the literature that the various indices of inter-rater reliability each have advantages and disadvantages; hence, at least two indices should be used to measure inter-rater consistency. The terms of inter-rater consistency, inter-rater agreement, inter-observer reliability, inter-judge reliability, and inter-rater reliability have been used interchangeably in the literature and were considered synonymous for this study.

Based on a review of the relevant literature, it was concluded that percentage agreement is the simplest measure of inter-rater reliability; however, it does not take into account the agreement that would be expected by chance alone. Consequently, percentage agreement is normally augmented by one or more complex measures that indicate the proportion of agreement beyond that expected by chance, such as Cohen's kappa, Cohen's weighted kappa, or Fleiss kappa, depending on the number of raters and whether the variables are nominal, ordinal, or continuous. Fleiss' kappa is a variant of Cohen's kappa that works for any constant number of raters assigning categorical ratings to a fixed number of items. The simplest use of Cohen's kappa is where two raters provided an independent single dichotomous nominal rating for each case, while a weighted Cohen's kappa is appropriate for ordinal and continuous ratings to assess the level of disagreement by attaching greater emphasis to large differences between ratings than small differences (Sim & Wright, 2005). Kappa (*k*) indicates the proportion of agreement beyond what is expected by chance and takes the form of the following equation:

$$K = \frac{(\,observed\ agreement - chance\ agreement)}{1 - chance\ agreement}$$

The data for a 2-category (dichotomous) nominal scale are usually displayed on a 2 x 2 contingency table, as presented in Table 5 (Gwet, 2001).

Table 5

Contingency table for runway incursion alerting

		Rater 1		
		Alert	No Alert	Total
Rater 2	Alert	a	b	$g_1 = a + b$
	No Alert	c	d	$g_2 = c + d$
	Total	$f_1 = a + c$	$f_2 = b + d$	$N = f_1 + f_2$

Note: Adapted from "The kappa statistic in reliability studies: Use, interpretation, and sample size requirements," by J. Sim, and C.C. Wright, 2005, *Journal of the American Physical Therapy Association*, *85*(3), 257-268.

Sim and Wright (2005) recommended the following method to determine kappa. The frequency of observed agreement is obtained by summing the frequencies of the main diagonal cells ($a + d$). The proportion of observed agreement (P_o) is obtained by dividing the frequency by the total number of ratings (n). The frequency of chance agreement for *alerting* and *non-alerting* ratings is calculated by multiplying the marginal totals corresponding to each cell on the main diagonal and dividing by n. The proportion of expected agreement (P_c) is obtained by summing across chance agreement in these cells and dividing by n. The values of P_c and P_o are used to determine k as shown in the following equations (Cohen, 1960; Sim & Wright, 2005):

$$P_o = (a + d)/n$$

$$P_c = \left[\frac{(f1\ x\ g1)}{n} + \frac{(f2\ x\ g2)}{n} \right] \Big/ n$$

$$k = \frac{(Po - Pc)}{1 - Pc}$$

The range of possible values of kappa is from -1.0 to +1.0, where the latter indicates perfect agreement by the raters in every case. Zero indicates agreement no better than by chance, and negative values indicates agreement worse than expected by chance. Green (1997) suggested that kappas greater than .75 were considered to have a high agreement beyond chance, values below .40 have a low agreement, and values between .40 and .75 represented a fair to good level of agreement beyond chance alone. Landis and Koch (1977) had a somewhat more refined scale for standards of strength of agreement for the kappa coefficient, which appeared to be the most widely accepted in the literature reviewed for this study (Table 6). The minimally acceptable value of kappa depends on the context (Laura & William, 1999). Medical studies have defined clinically important kappa values ranging from less than .50 to greater than .80 depending on the level of medical risk, while academic textbooks and statisticians generally recommend a kappa value of .60 or greater. Some studies have recognized any kappa value above zero, better than by chance, as minimally acceptable. Overall, the preponderance of authors have acknowledged that kappa tends to be lower than other measures of inter-rater agreement since it corrects for chance, and that the kappa divisions for strength of agreement are arbitrary (Landis & Koch, 1977; Leech, Barret, & Morgan, 2008). This study followed the guidance from Sim and Wright (2005) for a small sample size of nine cases in a 2-rater study, which required a kappa of .90 to be statistically significant ($p < .05$) for the dichotomous variable of *alerting* or *non-alerting*. The high kappa value of .90 was consistent with other research involving elevated risk and was considered appropriate within the context of this study (Laura & William, 1999).

Table 6

Generally accepted standards of agreement for kappa

Kappa Statistic	Strength of Agreement
< 0.0	Poor
0.0-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Note: Adapted from “The measurement of observer agreement for categorical data coefficient of agreement for nominal scale,” by J. Landis, and G. Koch, 1977, *Biometrics(33)*1, 159-174.

The kappa statistic alone is appropriate if the marginal totals for the 2 x 2 contingency table are relatively balanced. However, if a statistical computer program such as *SPSS* is used to calculate kappa when the prevalence of a given response is very high, the value of kappa may indicate a low level of reliability even when the observed proportion of agreement is quite high (Cunningham, 2009; Sim & Wright, 2005). For example, the proportion of observed agreement in Table 7 was 95%; however, the *SPSS* calculated kappa was -0.0163, indicating poor agreement according to the Landis and Koch (1977) generally accepted standards of agreement.

Table 7

Prevalence and bias paradox

Rater A * Rater B Cross-tabulation

			Rater B		
			No Alert	Alert	Total
Rater A	No Alert	Count	95	4	99
		% of Total	95%	4%	99%
	Alert	Count	1	0	1
		% of Total	1%	0%	1%
Total		Count	96	4	100
		% of Total	96%	4%	100.0%

Note: Adapted from "More than Just the Kappa Coefficient: A Program to Fully Characterize Inter-Rater Reliability between Two Raters," by M. Cunningham, 2009, *SAS Global Forum*,242.

Prevalence is the proportion of agreement on ratings of the attribute, sometimes called symmetry of agreement. A high prevalence index (PI) indicates a high chance agreement and leads to reduced kappa values, and vice versa. Bias index (BI) is the extent to which the raters disagree on the proportion of ratings. When there is a large bias, kappa is higher than when bias is low or absent. Prevalence index and bias index have been used to interpret and inform the magnitude of kappa using the following formulas (Sim & Wright, 2005):

$$\textit{Prevalence Index} = |a - d|/n$$

$$\textit{Bias Index} = |b - c|/n$$

In the example from Table 7, the calculated PI of .95 and the BI of .03 accounted for the misleading low kappa (-.0163), even though there was a 95% observed agreement (Cunningham, 2009).

Some authors have suggested computing a kappa adjusted for prevalence and bias (Byrt, Bishop, & Carlin, 1993; Cunningham, 2009; Sim & Wright, 2005). Prevalence is taken into account by computing the average value of cell a and cell d in Table 5 and substituting that value into the actual value of those cells. Bias is accommodated by substituting the average of cells b and c for those actual cell values. The resulting kappa coefficient is referred to as prevalence-adjusted bias-adjusted kappa (PABAK).

The use of two raters providing nominal ratings and measuring agreement with the kappa statistic has been previously used in a variety of peer reviewed research and doctoral dissertations. Kilpikoski et al. (2002) conducted a study where two clinicians rated low back pain for 39 patients using a dichotomous rating scale. Sim and Wright (2005) cited another related study in a book by McKenzie (1981) that provided data for the agreement of two raters for categorical classification of spinal pain. Wrisley (1998) utilized two raters and kappa statistics to assess the inter-rater reliability for rating the performance of 30 human subjects completing walking tasks on level surfaces and climbing stairs. An example of an aviation related study that relied on two raters was a NASA analysis of airspace violations that categorized the apparent factors and causes into eight dimensions. The study involved 22 records, which were coded independently by two raters with only percentage agreement used as a measure of inter-rater reliability

(Zuschlag, 2005). Another aviation related study utilized two radiologists who independently rated spine x-rays of F-16 pilots to classify spinal degeneration, with inter-rater reliability assessed by percentage agreement and Cohen's kappa (Hendriksen & Holewijn, 1999).

A multitude of doctoral dissertations have used two raters and inter-rater reliability measures of percentage agreement and kappa. A university research study by Mata (1993) assessing the value of chest x-rays as a screening tool for the diagnosis of skeletal disorders used two physicians as raters with inter-rater reliability measured by percentage agreement and kappa. D'Amato (2008) rated a sample of 218 dreams of adult adopted women using two raters, and more recently Arany (2012) used two raters in a dissertation involving the coding of family mediation agreements.

Sim and Wright (2005) suggested: (a) constructing a confidence interval around the kappa value obtained using the standard error (*SE*) of kappa (*k*) and the *z* score corresponding to the desired level of confidence to reflect sampling error; and (b) testing the significance of kappa against a value that represents a minimum acceptable level of agreement, rather than against zero, thereby testing whether its plausible values lie above an acceptable threshold. Sim and Wright (2005) also presented a table with the number of subjects (incidents) required in a 2-rater study to detect a statistically significant kappa for $p \leq .05$ on a dichotomous variable (Table C1).

Notably, Shoukri (2004) asserted that when seeking to detect a kappa ≥ 0.40 on a dichotomous variable, using more than three raters had little effect on the power of hypothesis testing or the width of the confidence intervals, and suggested that increasing the number of subjects (incidents) was the more effective strategy for maximizing statistical power.

Metric Validity

Decisions to fund and support new technology in any field (e.g., aviation, medical, information technology) have hinged on the demonstration of a benefit as measured by applicable valid metrics (Laupacis, Feeny, Detsky, & Tugwell, 1992). However, the rapidity in which new technology has developed often results in the technology outpacing the validity of the metrics used to measure its effectiveness (Bughin, Shenkan, & Singer, 2009).

The legacy/traditional metrics recorded in ASIAS, such as those for runway incursions and near mid-air collisions, can be considered *operational metrics* that can be used to determine the continued funding and support of new technology after initial operational introduction into the NAS. However, when new technology is in its infancy and still in the research and developmental phase, the validity is determined by engineering metrics, which can also be outdated, invalid, or otherwise not harmonized with the new technology. Numerous legacy/traditional rating scales have been used as engineering metrics during the certification of new aircraft technology. One of the most recognizable is the Handling Qualities Rating (HQR) Scale depicted in Figure 8, which was first developed in 1966 by Harper and Cooper (1966, 1986). Handling qualities are those qualities or characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform a specified task (Cooper & Harper, 1969). The HQR scale, commonly called the Cooper-Harper scale, is a decision tree that considers task performance and workload in determining the rating. Task performance is quantitative and readily measured as either adequate or desired, for example: (a) maintained runway centerline within ± 3 feet (adequate); or (b) maintained runway centerline within ± 1 feet (desired). Workload is subjective and qualitatively assessed on a continuum for pilot compensation ranging from minimal to intense.

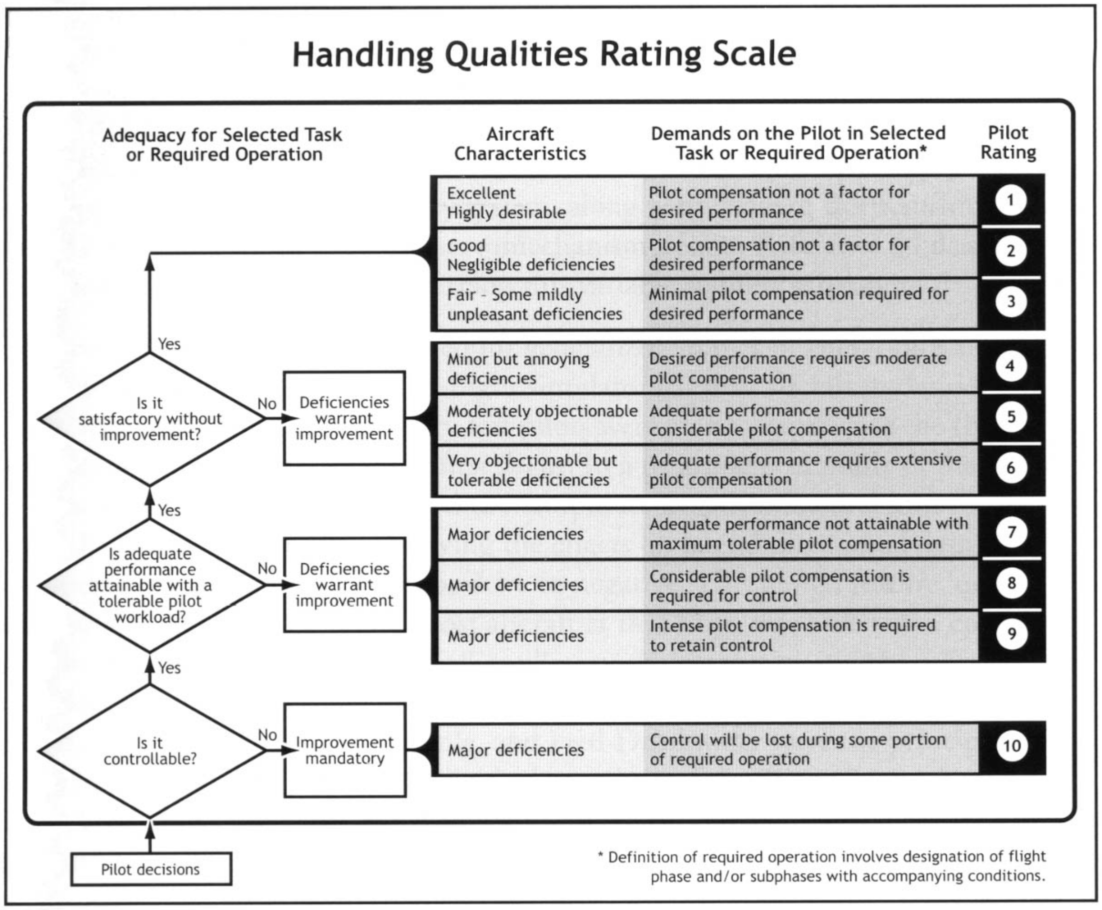


Figure 8. Cooper-Harper handling qualities rating scale. Adapted from "The use of pilot rating in the evaluation of aircraft handling qualities," by G.E. Cooper and R. P. Harper, 1969, NASA TN D-153.

The Cooper-Harper definition of pilot workload recognized mental workload to some extent, but acknowledged that it could not be directly quantified. When developing the handling qualities rating scale and initially validating the pilot ratings for workload, Harper and Cooper (1966) used pilot heart rates, and/or frequency and magnitude of control inputs as proxies for workload. Hence, the scale focused primarily on the physical effort expended by the pilot in moving or imposing forces on the mechanical flight controls during a specified task.

The HQR scale is still in use today for rating aircraft handling qualities, even though aircraft have evolved from manually operated mechanical flight and engine controls to fly-by-wire systems with fully automated digital engine controls and autopilots (Cooper & Harper, 1969; Harper & Cooper, 1986). The consensus in the literature is that workload is comprised of multiple factors, to include but not limited to physical, mental, psychological, and environmental. New technology in modern aircraft has transformed the primary role of the pilot from a direct manipulator of flight controls, to a systems operator where mental workload far exceeds physical demands from moving or imposing forces on mechanical flight controls. Hence, the prevalence and influence of these workload factors in modern aircraft has shifted away from the physical workload of physical force inputs to mechanical flight controls, demanding changes to the engineering metrics.

The Royal Aerospace Establishment, NASA, and others recognized the shortcomings in the Cooper-Harper workload assessment and developed modified and new scales, such as the Bedford rating scale, Subjective Workload Assessment Technique (SWAT), and the NASA Task Load Index (TLI) (Hart & Staveland, 1988; Reid & Nygren, 1988; Roscoe & Ellis, 1990).

For example, the Bedford scale focused exclusively on pilot workload, measured by the pilot's excess capacity to perform other tasks, while the NASA TLX model considered six factors; mental, physical, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988; Roscoe & Ellis, 1990). Although there have been modifications to the workload rating metrics, there has not been a consistent process for determining when new technology requires modifications to legacy metrics, and what modifications are necessary to ensure the legacy metrics are valid for the new technology.

The application of this methodology is also evident in non-aviation fields such as Internet applications where Russell (2009) suggested that "new metrics are needed, in part because the legacy metrics are outdated but also because the digital world is evolving at an accelerated speed". Although Russell (2009), Bughin et al. (2008), and others have recognized the increasing obsolescence of legacy or traditional metrics in non-aviation fields, no one has presented a methodology for evaluating and modifying legacy or traditional metrics for application to new technology in any field.

Summary

As SURF-IA technology is installed in aircraft, the FAA will determine the effectiveness and benefits based on the change in the actual number and rate of pilot deviation type runway incursion incidents using the existing FAA data gathering and reporting process for runway safety statistics, which is based on the RISC model. Hence, the SURF-IA model should be validated using profiles from specific actual incidents that have been categorized by the RISC model and recorded in the FAA ASIAS database, as was performed in this study, instead of a set of generic scenarios as has been previously accomplished. This study improved on shortcomings in previous research by using RISC model outcomes from specific runway incursion incidents recorded by the FAA in ASIAS to examine the outcomes from the SURF-IA alerting model. The literature review also indicated precedents in research and academia for utilizing two raters and measuring inter-rater reliability by percentage agreement and Cohen's Kappa.

CHAPTER III

METHODOLOGY

Video reenactments of runway incursion incidents and the associated ASIAS incident report data were reviewed by two expert raters. Although the raters may have been familiar with the types of incidents investigated in this study, the focus of the ratings was on evaluating the SURF-IA alerting outcomes. Thus, any prior familiarity with the incidents was not expected to influence their ratings of the SURF-IA technology. The raters applied the baseline minimally compliant implementation of the RTCA/DO-323 SURF-IA model to determine which incidents would have triggered a SURF-IA Warning or Caution *alert*. A minimum acceptable PABAK of ≥ 0.90 was used for this study. The runway incursion incidents were rated on a dichotomous scale that classified the incidents as either SURF-IA *alerting* or SURF-IA *non-alerting*. This research approach of using expert raters to validate legacy metrics for application to new technology is generalizable to other fields of study.

Research Approach

The analysis focused on whether runway incursion incidents classified with an outcome as *serious* (Category A or B) using the RISC model would trigger a SURF-IA model outcome to display an *alert* (Warning or Caution) to the pilot.

The video reenactments and ASIAS runway incursion report narratives were viewed by two expert raters who were the actual developers of the *aircraft states* and *alerting* outcomes for the SURF-IA model, as defined in Table 8. Both raters are internationally recognized SURF-IA subject matter experts and have been referenced extensively in the literature as well as the citations for this study. To avoid any real or perceived bias, the raters were selected from non-regulatory government agencies and independent government contractors, rather than from the FAA, which regulates/reports runway incursions and certifies new technology; the avionics industry, which profits from the certification of new technology, or; the airline industry, which is subject to pilot deviation reports from runway incursions. The minimum requirements for the SURF-IA alerting model were applied to each incident using the logic adapted from RTCA/DO-323 and the definitions from Table 8 (RTCA, 2010).

Table 8

Definition of aircraft states

Aircraft State	Explanation of Aircraft State
Entering/Crossing Runway (not lined-up)	Heading > 20º difference from runway heading
Take-off	From detection of take-off roll to lift-off
Approach to runway	Straight path segment toward a runway, 1000 ft. AFE and ≤ 3 nm from RTHRE
After Landing Roll-out on-runway	≥ 40 knots
Stopped or Taxiing along runway (lined-up)	Heading < 20º difference from runway heading and < 40 knots

Note: Adapted from “Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA)” by Radio Technical Commission for Aeronautics [RTCA/DO-323], 2010.

In addition to the SURF-IA model limitations mentioned in Chapter II, the model follows the logic diagram in Figure 9. Commencing from the “Start” arrow in Figure 8 and moving vertically down, the first two blocks represent SURF-IA design limitations where the logic path leads to the *no alert* conditions for: (a) wrong runway departure; and (b) less than 40 knots closure between ownship and traffic. The next block vertically down addresses whether or not one of the aircraft is airborne. No SURF-IA *alert* will be issued if the airborne traffic is above 1000 feet above field elevation (AFE), more than 3 nm from the runway threshold (RTHRE), or more than 35 seconds from the runway threshold.

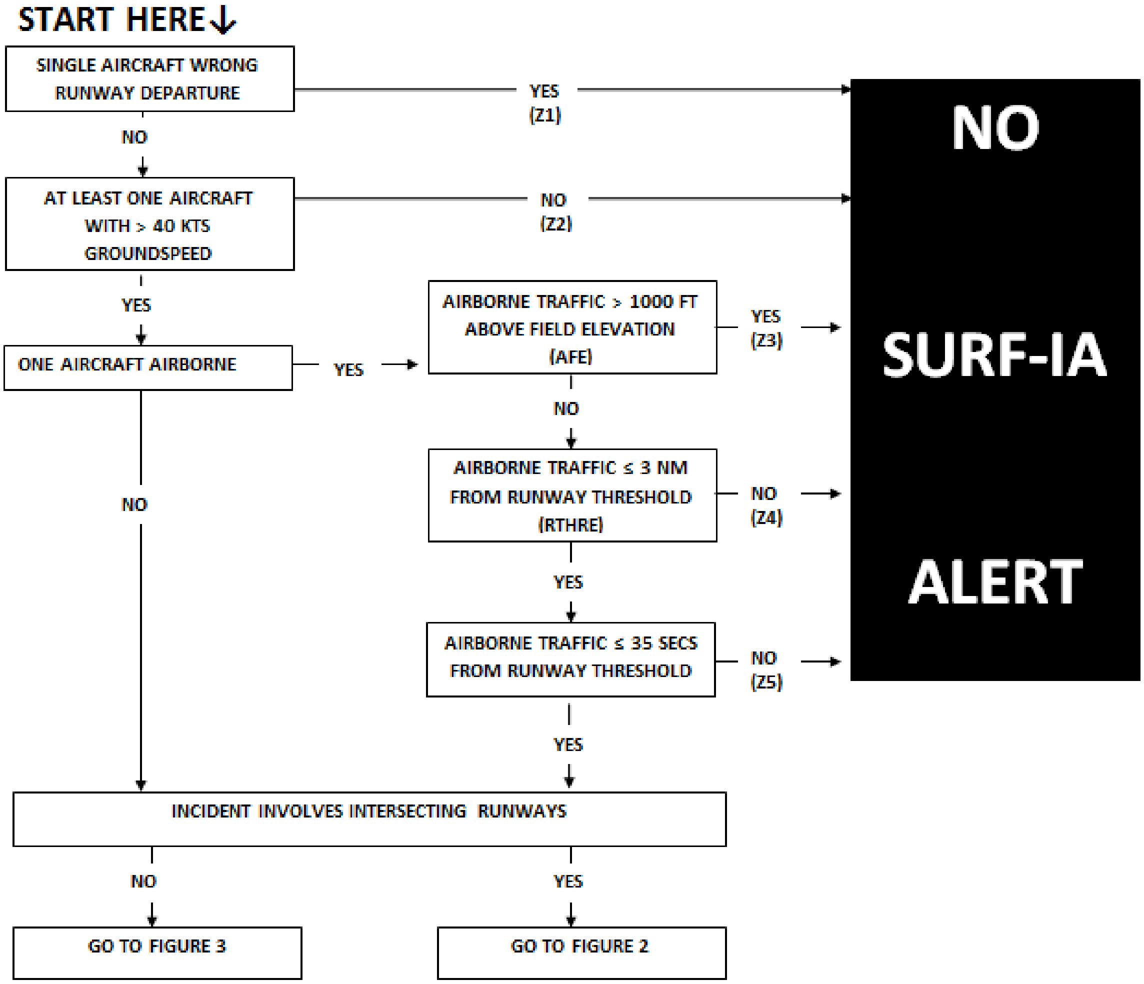


Figure 9. SURF-IA alerting logic. Adapted from “Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA),” by Radio Technical Commission for Aeronautics [RTCA/DO-323], 2010.

The SURF-IA model specifies a symmetric 4 x 4 matrix of ownship/traffic *aircraft states* for runway incursions involving intersecting runways (Figure 10). The four possible *aircraft states* are: (a) on approach and within 35 seconds of runway threshold; (b) on approach and within 15 seconds of runway threshold; (c) landing; and (d) taking off. All paired *aircraft states* are designed to provide an *alerting* outcome, except for two conditions: ownship will not have a SURF-IA alerting outcome during landing or take-off if the traffic aircraft on approach, and within 35 seconds of the runway, is trailing behind ownship as shown in blocks C1 and D1 of Figure 10. However, ownship will provide an *alert* once the trailing traffic on approach is within 15 seconds of the runway threshold, as shown in blocks C2 and D2 of Figure 10.

	OWNSHIP STATE					
		A	B	C	D	
TRAFFIC STATE		ON APPROACH AND WITHIN 35 SECS OF RUNWAY THRESHOLD	ON APPROACH AND WITHIN 15 SECS OF RUNWAY THRESHOLD	LANDING	TAKING OFF	
	ON APPROACH AND WITHIN 35 SECS OF RUNWAY THRESHOLD	ALERT (A1)	ALERT (B1)	NO ALERT (C1)	NO ALERT (D1)	1
	ON APPROACH AND WITHIN 15 SECS OF RUNWAY THRESHOLD	ALERT (A2)	ALERT (B2)	ALERT (C2)	ALERT (D2)	2
	LANDING	ALERT (A3)	ALERT (B3)	ALERT (C3)	ALERT (D3)	3
	TAKING OFF	ALERT (A4)	ALERT (B4)	ALERT (C4)	ALERT (D4)	4

Figure 10. SURF-IA alerts for intersecting runways. Adapted from "Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA)," by Radio Technical Commission for Aeronautics (RTCA/DO-323).

For all other possible two-*aircraft states* involving intersecting or same runway scenarios, the SURF-IA model specifies a 5 x 8 matrix of ownship/traffic states. Five *aircraft states* are common to ownship and traffic: (a) entering-crossing runway (not lined up), (b) take-off, (c) approach to runway (≤ 3 nm from runway), (d) alert landing roll-out on runway, and (e) stopped or taxiing along runway (lined up). Three exclusive *aircraft states* for traffic are identified for intersecting runways: (a) approach to intersecting runway (≤ 3nm from runway), (b) landing rollout intersecting runway, and (c) take-off on intersecting runway. All of the 25 paired states not involving intersecting runways are designed to provide a SURF-IA alert, with the following four *aircraft state* pairs as exceptions (Figure 11): (a) ownship and traffic simultaneously entering or crossing the same runway (not lined-up), as shown in block E1; (b) one aircraft entering or crossing the runway (not lined up) and the other aircraft stopped or taxiing on the same runway (lined up), as shown in block, I1, and E12; (c) ownship and traffic simultaneously stopped or taxiing on same runway (not lined-up), as shown in block I12. All of the 15 paired states involving intersecting runways are designed to provide a SURF-IA alert, except for six aircraft state pairs where ownship is either stopped/taxiing along a runway (not lined up), or entering-crossing a runway (not lined up) with the traffic aircraft on approach, landing rollout, or taking off from an intersecting runway, as shown in blocks E14, E16, E18, I14, I16, and I18 of Figure 11.

		OWNSHIP STATE					
		E	F	G	H	I	
	▢ Alert ■ No Alert	ENTERING/CROSSING RUNWAY (Not Lined Up)	TAKE-OFF < 80 KTS	APPROACH TO RUNWAY (≤ 3 NM FROM RWY)	AFTER LANDING ROLL-OUT ON RUNWAY	STOPPED OR TAXIING ALONG RUNWAY (Lined Up)	
TRAFFIC STATE	ENTERING/CROSSING RUNWAY[1] (Not Lined Up)	NO ALERT (E1)	TRAFFIC AHEAD (F1)	TRAFFIC AHEAD ≤ 15 secs (G1)	TRAFFIC AHEAD (H1)	NO ALERT (I1)	1
				TRAFFIC AHEAD≤ 35 secs (G2)			2
	TAKE-OFF FROM SAME RUNWAY	CONVERGENCE[2] (E3)	TRAFFIC AHEAD (F3)	TRAFFIC BEHIND (G3)	TRAFFIC AHEAD (H3)	TRAFFIC BEHIND (I3)	3
			TRAFFIC HEAD-ON (F4)	TRAFFIC AHEAD or TRAFFIC HEAD-ON≤15 secs (G4)	TRAFFIC BEHIND W/CONVERGENCE (H4)	TRAFFIC HEAD-ON (I4)	4
				TRAFFIC AHEAD or TRAFFIC HEAD-ON≤35 secs (G5)			5
	APPROACH TO SAME RUNWAY (≤ 3 NM FROM RWY)	CONVERGENCE ≤15 secs (E6)	TRAFFIC AHEAD (F6)	TRAFFIC HEAD-ON≤15 secs (G6)	TRAFFIC AHEAD (same direction)≤ 15 secs (H6)	TRAFFIC BEHIND or TRAFFIC HEAD-ON≤15 secs (I6)	6
		CONVERGENCE ≤35 secs (E7)	TRAFFIC HEAD-ON (F7)	TRAFFIC HEAD-ON≤35 secs (G7)	TRAFFIC HEAD-ON (opposite direction)≤ 15 secs (H7)	TRAFFIC BEHIND or TRAFFIC HEAD-ON ≤ 35 secs (I7)	7
					TRAFFIC Behind≤ 15 secs (H8)		8
					TRAFFIC Behind≤ 35 secs (H9)		9
	AFTER LANDING ROLL-OUT ON RUNWAY	CONVERGENCE (E10)	TRAFIC AHEAD (F10)	TRAFFIC AHEAD or TRAFFIC HEAD-ON≤15 secs (G10)	TRAFFIC AHEAD (H10)	TRAFFIC BEHIND (I10)	10
			TRAFFIC HEAD-ON (F11)	TRAFFIC AHEAD or TRAFFIC HEAD-ON≤35 secs (G11)	TRAFFIC BEHIND W/ CONVERGENCE (H11)	TRAFFIC HEAD-ON (I11)	11
	STOPPED OR TAXIING ON SAME RUNWAY (Lined Up)	NO ALERT (E12)	TRAFFIC AHEAD (F12)	TRAFFIC AHEAD ≤ 15 secs G(12)	TRAFFIC AHEAD (H12)	NO ALERT (I12)	12
			TRAFFIC HEAD-ON (F13)	TRAFFIC AHEAD ≤ 35 secs G(13)	TRAFFIC BEHIND W/ CONVERGENCE (H13)		13
	APPROACH TO INTERSECTING RUNWAY (≤ 3 NM FROM RWY)	NO ALERT (E14)	CONVERGENCE ≤ 15 secs NO LAHSO[3] (F14)	CONVERGENCE≤15 secs NO LAHSO (G14)	CONVERGENCE ≤15 secs NO LAHSO (H14)	NO ALERT (I14)	14
				CONVERGENCE≤35 secs NO LAHSO (G15)			15
	LANDING/ROLLOUT INTERSECTING RUNWAY	NO ALERT (E16)	CONVERGENCE NO LAHSO (F16)	CONVERGENCE≤15 secs NO LAHSO (G16)	CONVERGENCE NO LAHSO (H16)	NO ALERT (I16)	16
				CONVERGENCE≤35 secs NO LAHSO (G17)			17
	TAKE-OFF ON INTERSECTING RUNWAY	NO ALERT (E18)	CONVERGENCE (F18)	CONVERGENCE≤15 secs NO LAHSO (G18)	CONVERGENCE NO LAHSO (H18)	NO ALERT (I18)	18
				CONVERGENCE≤35 secs NO LAHSO (G19)			19

Figure 11. SURF-IA alerts for all possible two-*aircraft state* combinations. Adapted from “Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF-IA),” by Radio Technical Commission for Aeronautics [RTCA/DO-323].

Each incident was dichotomously categorically coded whether or not the incident would have triggered an *alert* (Warning or Caution) in accordance with the RTCA (2010) SURF-IA model. This study did not evaluate the physical location of the SURF-IA annunciation on the CDTI in the pilot's field-of-view, navigational positional accuracy, or aircraft deceleration/braking performance. Hence, the raters only rated whether or not a SURF-IA *alert* would have been triggered, without considering whether or not a runway incursion would have been avoided.

Pilot Study. A pilot study (i.e., feasibility study) was conducted to establish the validity of the instrument and the inter-rater reliability of the raters. The sample for the pilot study consisted of FAA video reenactments and ASIAS reports from nine *serious* runway incursion incidents, which was the minimum required sample size to determine inter-rater reliability according to the Sim and Wright (2005) guidance. The population of *serious* runway incursions incidents, for which video reenactments were developed by the FAA Office of Runway Safety, consisted of 58 incidents; however, only nine were PD type. Consequently, the pilot study utilized Operational Error/Deviation/ Incident (OE/D/I) type incidents and not any of the PD type incidents, which were reserved for the main study. The SURF-IA software alerting algorithm made no distinction between aircraft alerting states caused by the actions of a pilot or ATC; hence, OE/D/I incidents used in the pilot study were considered surrogates for PD type incidents. The pilot study incidents were validated against a true score, which was expected to be a SURF-IA *alert* for all nine incidents based on the outcomes from similar scenarios that alerted during the ACSS (2010) and Honeywell (2010) SURF-IA flight demonstrations.

The data from the two raters in the pilot study were consolidated to show a side-by-side comparison, and then used to develop a 2 x 2 contingency table with descriptive statistics. The pilot study was used to: (a) check that the instructions given to the raters in Appendix E were comprehensible; (b) verify the raters were skilled in viewing the video reenactments and associated narratives; (c) ensure the document and video files on the media storage device provided to the raters were readable; and (d) evaluate the forms, procedures, and data analysis approach to identify any necessary modifications. The same raters were used in the pilot study and the main study, as suggested by Thabane et al. (2010), and van Teijlingen and Hundley (2001). A Human Subjects Protocol application was submitted to the ERAU Institutional Review Board (IRB) which granted the study an exemption for both the pilot study and the main study since the research used existing data and posed no risks to the raters (Appendix A). Pilot studies are often used to estimate the sample size required for a main study to be statistically significant (Thabane et al., 2010). However, this study established a sample size of nine cases for the main study a priori based on the Sim and Wright (2005) guidance. Although pilot study sample sizes are typically smaller than those for a main study, matching sample sizes were used to provide the raters familiarity and training with the rater instructions, instruments, and data collection devices (Hertzog, 2008; Thabane et al., 2010). Both raters had previously utilized and cited ASIAS reports and the SURF-IA alerting logic in their own research (Jones et al., 2012; Moertl et al., 2012), hence no additional training beyond the pilot study was deemed necessary. The identical sample sizes also provided some insight for the time that would be required for the raters to complete the main study. The raters were mailed a media storage device with document and video files as listed in

Appendix E. The raters were then allowed to rate the nine runway incursion incidents at a time and place of their convenience, under the condition that they abide by the provisions of their signed Informed Consent Form (Appendix A), which required an independent assessment to be completed within four weeks using no other materials. Both raters reported that they required approximately four hours to rate all nine incidents, and both returned the data collection sheet (Appendix F) via email within two weeks of receipt of the rater data package. The data from the pilot study were not merged with the main study due to modifications that were incorporated into the procedures as a result of the pilot study. Furthermore, the data for the pilot study consisted of OE/D/I type incidents that constituted a different sampling frame from the PD type incidents used in the main study. The pilot study was analyzed to ensure the *reason codes* matched the rated outcomes. For example, if a *reason code* of I12 was selected by the rater from Figure 10, it should have been recorded in the *no alert* block of the data collection device in Table B1. In some cases the raters assigned multiple *reason codes* as a result of confusion with some of the definitions, such as when an aircraft was considered to be *on-runway* or at what speed did an aircraft on the runway transition from a “taxiing” state to a “take-off” state. Consequently, prior to the main study the raters where provided a table of definitions for all *aircraft states*, derived from the SURF-IA RTCA (2010) requirements document and presented in Table 8.

Design and Procedures. A Human Subjects Protocol Application was approved by the ERAU IRB, which included an Informed Consent Form for the raters (Appendix A). Each rater was provided a media storage device with runway incursion reenactment videos and document files with narratives from the associated ASIAS runway incursion reports. The raters reviewed video reenactments of less than five minutes duration each and read the associated FAA runway incursion report for the incident extracted from the ASIAS integrated online database of safety data. The SURF-IA alerting logic from RTCA/DO-323, as shown in Figures 9, 10, and 11, was then followed to determine if the incident would have triggered an alerting outcome (Warning or Caution) from the SURF-IA model. The results for each incident were recorded on a data collection device (Appendix B). The following instructions were provided to the raters as part of the Informed Consent Form, along with the additional instructions in Appendix E:

1. Rate 18 runway incursion incidents within four weeks of receipt of a rater package by mail which will include a media storage device (flash drive) with videos and document files, as well as paper copies of the documents.
2. This study does not evaluate the location of the SURF-IA annunciation on the CDTI in the pilot's field-of-view, navigational positional accuracy, or aircraft deceleration/braking performance. Hence, the raters should only rate whether or not a SURF-IA *alert* would have been triggered, without considering whether or not a runway incursion would have been avoided.
3. Not reproduce or share any of the items and will return them to this investigator along with a completed rater matrix.
4. Rate the incidents independently without discussion with any other person or reference to any other information.

5. You will not be expected to travel to any location but will require a personal computer with word processing (.doc and .docx) and the ability to view video files (.swf, .exe, Adobe® Flash® Player, Internet Explorer®).

The SURF-IA logic for this study only considered *alerts* and did not include or consider SURF-IA *indications*. Conflicts were defined as any movement between two aircraft that potentially could lead to a high speed collision on the runway surface. The conflict prediction was based on the relative speed and track between the two aircraft unless ownship was on the surface and the conflicting traffic was airborne on approach, or when ownship was airborne on approach and the conflicting traffic was intruding on the runway. In both of the latter cases, which involved one aircraft on the surface and one airborne aircraft, the *alert* logic was based on predicted time for the airborne aircraft on approach to reach the runway threshold (RTHRE). A Caution *alert* was issued if the predicted time to conflict was less than 35 seconds. A Warning *alert* was issued if the predicted time to conflict was less than 15 seconds (RTCA, 2010). There was no difference in the expected crew response to a Warning *alert* or a Caution *alert*: the difference between the alerts being solely a matter of urgency (Honeywell, 2010).

Apparatus and Materials. The SURF-IA model, as depicted in Figures 9-11, represented the minimum recommended output specifications that SURF-IA should provide. (RTCA, 2010). Reenactment videos of runway incursions along with the associated ASIAS narrative reports were provided to each rater. Examples of screenshots from a runway incursion video reenactment produced by the FAA Office of Runway Safety and an ASIAS runway incursion incident report are depicted in Appendix D.

Population/Sample

The FAA Office of Runway Safety produced reenactment videos of runway incursion incidents that were of high interest to the public, FAA, or the NTSB (R. Motzko, personal communication, July 11, 2012). The database consisted of 58 video reenactments from runway incursion incidents of all types and categories that occurred between CY 2005-CY 2012, of which nine were PD type *serious* (Category A or B) incidents. The sample set consisted of the entire population of *serious* pilot deviation type runway incursion incidents recorded by the FAA Office of Runway Safety in the ASIAS database, for which video reenactments were produced using actual surveillance data from the incidents (Table 9).

Table 9

ASIAS reports of pilot deviation type serious (Category A or B) runway incursion incidents

		ASIAS ID		
8173	5826	4828	7167	11322
10923	10675	10969	3374	-------

Note: Adapted from ASIAS database website (http://www.asias.faa.gov/)

An FAA (2002) analysis of 719 PD type runway incursions recorded from FY 1997-FY 2001 indicated that 624 of the incidents (87%) were associated with an aircraft that entered the runway or crossed the hold short line after acknowledging hold short instructions, landed over aircraft in position, or landed/departed on a closed/wrong runway. The FAA FY 2000-FY 2003 Runway Safety Report (2004) identified the following common errors in PD type runway incursion incidents: (a) pilots read back controller's instructions correctly but did not comply with the instructions; (b) pilots failed to hold short of the runway as instructed and crossed or taxied into position on the runway; and (c) pilots accepted clearances issued to an aircraft other than their own. Cardosi, Chase, and Eon (2010) had similar findings in an analysis of 637 PD type runway incursions reported in FY 2008. Another analysis by RTCA (2010) indicated that 84% of all runway incursions and 75% of the most severe runway incursions involved an aircraft entering a runway ahead of an aircraft departing or landing.

A more recent study by Joslin, Goodheart, and Tuccio (2011) analyzed 70 Aviation Safety Reporting System (ASRS) reports and also concluded that the primary event leading to runway incursion incidents was aircraft entering the runway after being instructed to hold short. The sample set used for this present study was comprised of nine incidents involving the common errors reported by Joslin et al. (2011), Cardosi et al. (2010), and the FAA (2004), which were representative of the most pervasive types of incidents that would be classified by the RISC model as a *serious* (Category A or B) incident.

A minimum sample size of eight was recommended by Sim and Wright (2005) for a 2-rater study to detect a statistically significant kappa ($p \leq .05$) for a one-tailed test with a null value of .00, kappa to detect of .90, and 80% statistical power. The Sim and Wright (2005) recommendation was based on Donner and Eliasziw (1992) goodness-of-fit Chi-square calculations for two raters with a dichotomous outcome. The minimum sample size was also predicated on a proportion of positive ratings between .10 to .90, calculated by Sim and Wright (2005) using the notation from Table 5 as follows:

$$\textit{Proportion of positive ratings} = (f_1 + g_1)/2n$$

Sources of the Data

Two archival sources of data were used: (a) ASIAS reports, and (b) FAA runway incursion video reenactments. ASIAS data are publically available without need for permission and were collected and posted on the FAA website http://www.asias.faa.gov/. FAA runway incursion video reenactments are also periodically posted on the FAA Office of Runway Safety website (http://www.faa.gov/airports/ runway_safety/videos/), and publically available without need for permission. The FAA video reenactments were developed through precise surveillance data from Airport Surface Detection Equipment, Model X (ASDE-X) and the Performance Data Analysis and Reporting System (PDARS), along with the narrative information from the ASIAS runway incursion reports.

The ASDE-X data came from surface movement radar located on the ATC control tower, multi-lateration sensors, ADS-B sensors, terminal automation system, and aircraft transponders. By fusing that data from these sources, ASDE-X determined the position and identification of aircraft on the airport surface as well as aircraft flying within five miles of the airport. The Performance Data Analysis and Reporting System, developed and maintained by the ATAC Corporation for the FAA, is a comprehensive set of software tools for gathering aviation performance data. PDARS data includes all flight planning data, speeds, headings, and altitudes along with a dynamic measuring tool built into the program to monitor the separation distance between selected aircraft. The audio files for the incidents were time synched with the available ASDE and/or PDARS files and matched with index points on the video frames. These data from the surveillance tapes were tied to the audio files, which were what the graphic artists worked with to match the index points, taking into account aircraft gross weight, initial roll speeds, acceleration, and deceleration rates during the incident. Drafts of the video were then sent to the observers of the incident, pilots, companies, and ATC facilities for a verification of the video reenactment (R. Motzko, personal communication, February 20, 2013).

Archival reports of actual runway incursion incidents were already classified as *serious* (Category A or B) by the FAA Office of Runway Safety by applying the legacy Runway Incursion Severity Classification model. These data, posted on the FAA ASIAS on-line database, were extracted and matched with the entire population of video reenactments of PD type incidents that were produced by the FAA Office of Runway Safety. Neither the FAA nor the NTSB considers the video reenactments to be part of any official investigation or official report.

Event summaries and airport diagrams showing the *aircraft states* for each of the nine PD type *serious* (Category A or B) runway incursion incidents are provided in

Figures 12 through20). The complete detailed ASIAS reports for the runway incursions, as extracted from the ASIAS on-line database, were provided to the raters.

Event 1 (ASIAS ID 8173): An airplane was instructed by ATC to hold short of Runway 33 on Taxiway F. The clearance was read back correctly; however, the taxiing airplane did not hold short and entered Runway 33, thereby conflicting with another aircraft on take-off from Runway 33 (Figure 12).

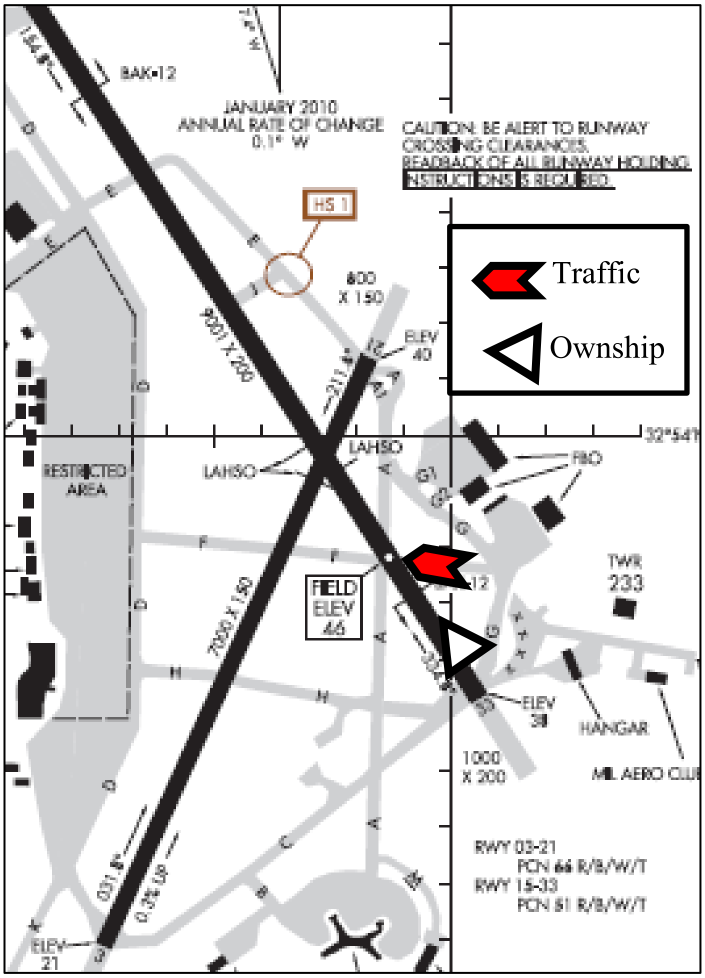


Figure 12. Event 1 (ASIAS ID 8173) with CHS airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 2 (ASIAS ID 5826): An airplane was instructed by ATC to taxi and hold short of Runway 7L at Taxiway N5; however, the airplane did not hold short and entered Runway 7L without clearance, thereby conflicting with an aircraft on take-off from the same runway (Figure 13).

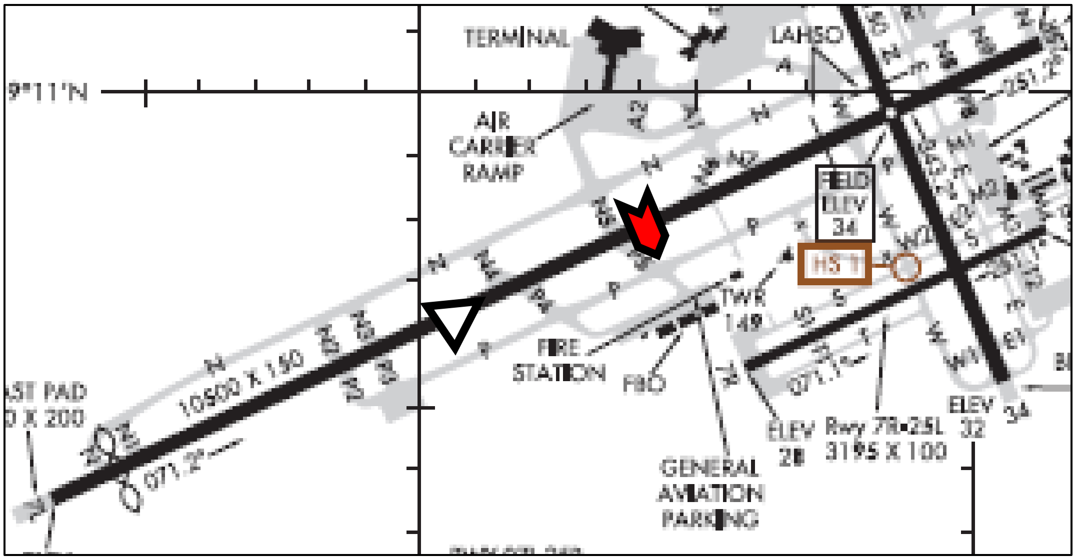


Figure 13. Event 2 (ASIAS ID 5826) with DAB airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 3 (ASIAS ID 4828): An airplane pilot was instructed by ATC to taxi from the ramp to Taxiway M; however, the pilot missed the left turn to Taxiway M and entered Runway 35L, thinking it was Taxiway M. The aircraft then proceeded to taxi on the active runway without clearance. Another airplane approximately one-half mile from the runway threshold on an approach to Runway 35L saw the airplane on the runway and executed a go-around to avoid a high speed ground collision upon landing (Figure 14).

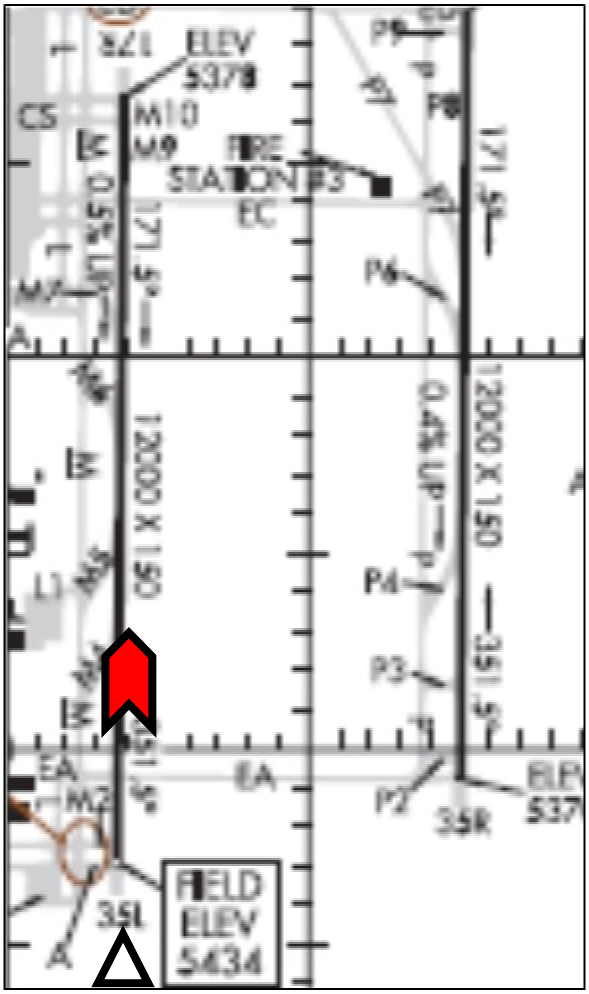


Figure 14. Event 3 (ASIAS ID 4828) with DEN airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 4 (ASIAS ID 7167): An airplane, holding short of Runway 7R at Taxiway B9, erroneously executed the ATC instructions given to another airplane (Call Sign 922 vs. 229), taxied onto Runway 7R without clearance, and was overflown by another airplane that had just lifted off from the same runway (Figure 15).

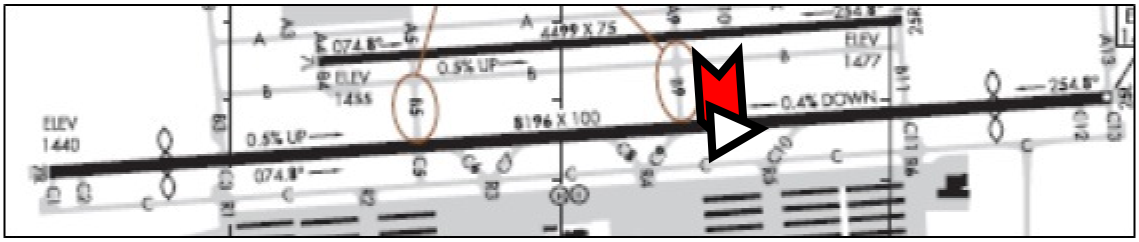


Figure 15. Event 4 (ASIAS ID 7167) with DVT airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 5 (ASIAS ID 11322): An airplane landed on Runway 35L that was occupied by another aircraft, after being instructed (twice) by ATC to execute a go-around due to insufficient separation from a preceding airplane that was still on the runway performing a stop-and-go maneuver (Figure 16).

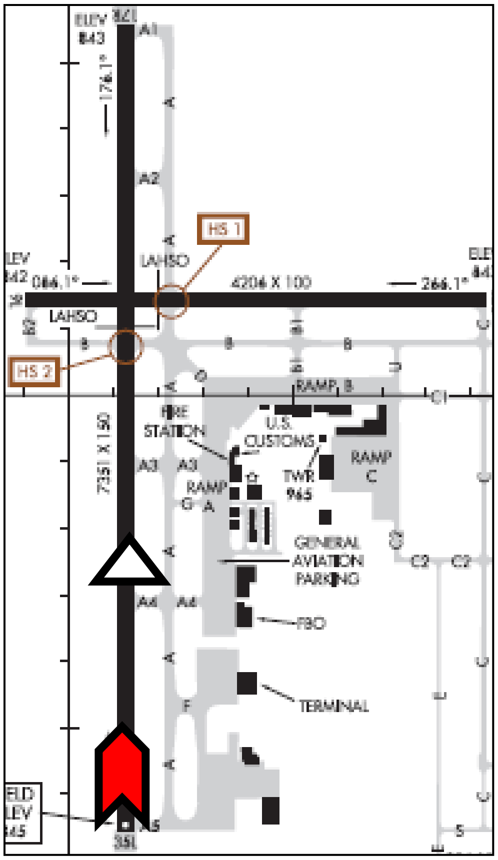


Figure 16. Event 5 (ASIAS ID 11322) with GFK airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 6 (ASIAS ID 10923): An airplane was instructed by ATC to taxi on Taxiway L and hold short of Runway 36R. The airplane read back the clearance correctly but then proceeded to cross Runway 36R/18L without a clearance and was overflown by an airplane that was taking off from Runway 18L (Figure 17).

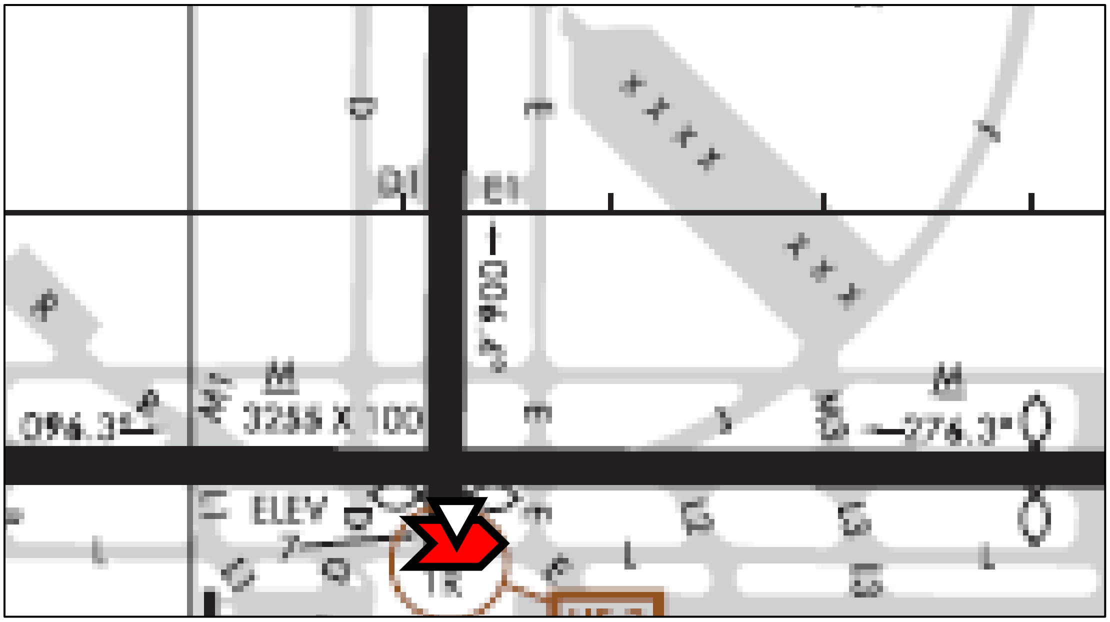


Figure 17. Event 6 (ASIAS ID 10923) with HWO airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 7 (ASIAS ID 10675): A helicopter was instructed to taxi to and hold short of Runway 16. The pilot read back the hold short clearance correctly but then continued to air taxi across Runway 16 and overflew an airplane that had just landed on Runway 16 (Figure 18).

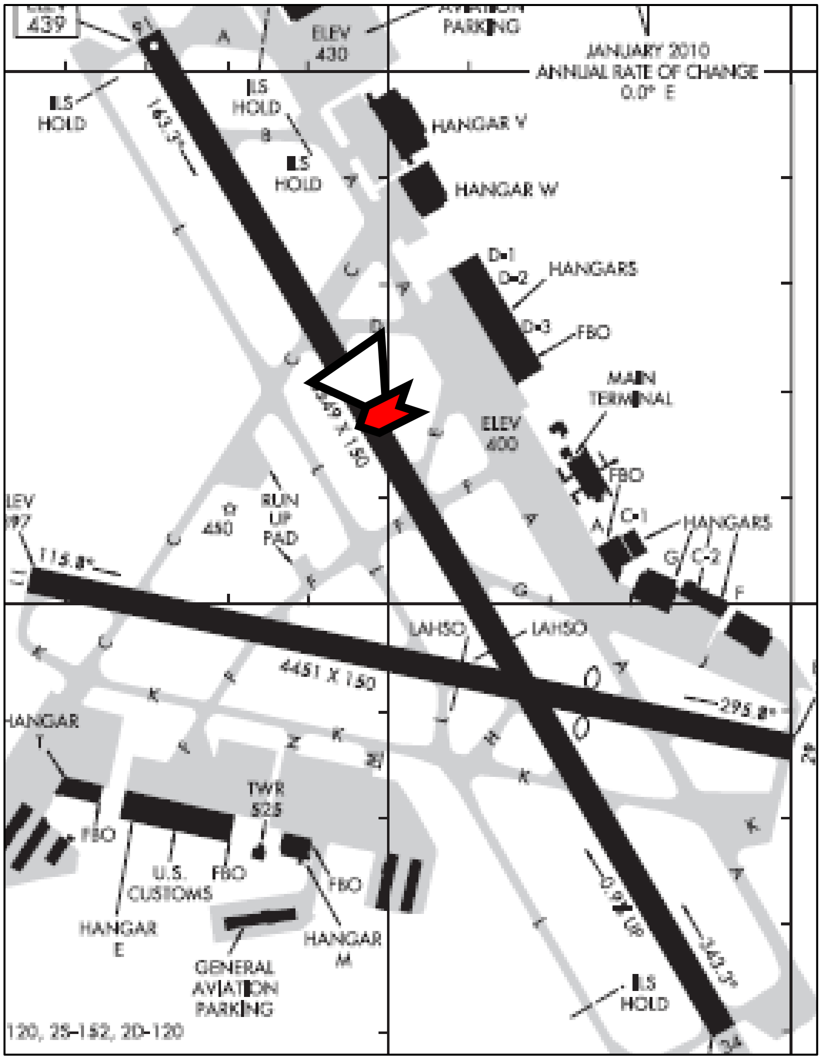


Figure 18. Event 7 (ASIAS ID 10675) with HPN airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 8 (ASIAS ID 10969): An airplane was given taxi instructions for a take-off on Runway 25L with intermediate instructions to hold short of Runway 19L. The airplane pilot called ATC ready for take-off, while still holding short of Runway 19L, which intersected with Runway 25L. ATC issued a take-off clearance for a takeoff on Runway 25L; however, the airplane initiated its take-off acceleration from Runway 19L, which was the wrong runway (Figure 19).

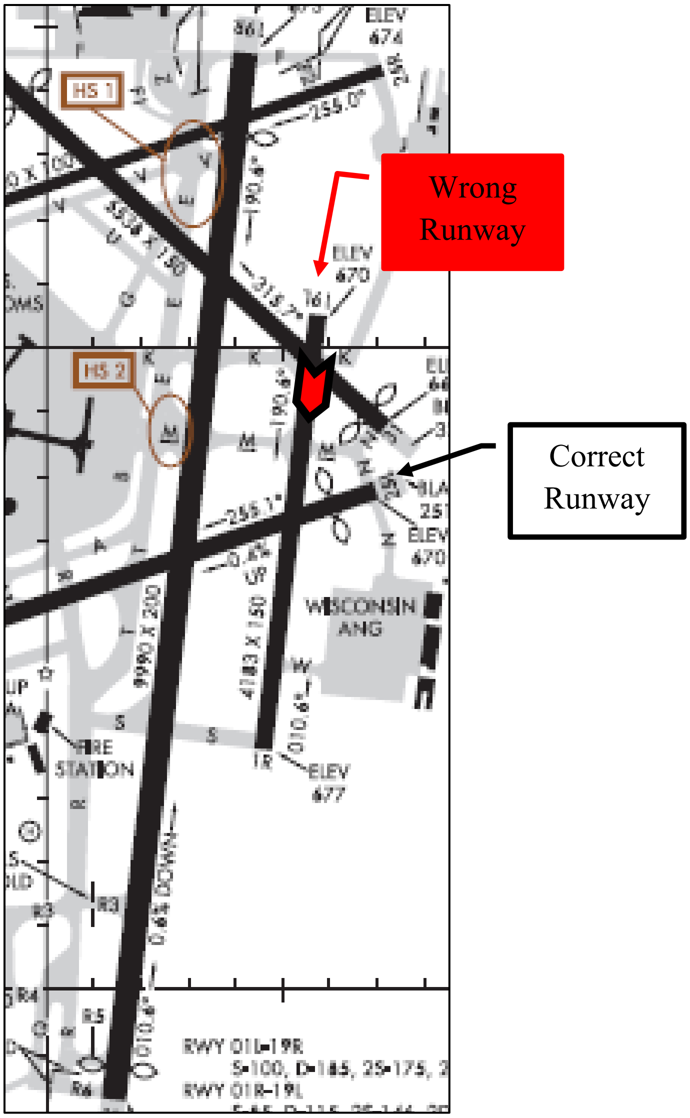


Figure 19. Event 8 (ASIAS ID 10969) with MKE airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Event 9 (ASIAS ID 3374): An airplane was instructed by ATC to taxi to Runway 22R via Taxiway H and turn left onto Taxiway B, which parallels Runway 22R. However, the airplane did not turn left onto Taxiway B as instructed, crossed Runway 22R via taxiway hotel, and was overflown by an airplane departing on Runway 22R (Figure 20).

Figure 20. Event 9 (ASIAS ID 3374) with JFK airport diagram. Adapted from FAA website (http://www.faa.gov/airports/runway_safety/diagrams/)

Data Collection Device

Rating data were collected from each rater for each incident and recorded in a table with two choices (*alert* or *no alert*), and incident number/name/location (Appendix B). The raters determined a SURF-IA *alerting* or *non-alerting* outcome for the aircraft that was cited with the pilot deviation in the runway incursion incident report. The list of incidents on the data collection devices were developed based on the entire population of PD type runway incursion incident video reenactments recorded in the FAA Office of Runway Safety database.

Instrument reliability. The data collection form was populated based on the raters' evaluation of ASIAS data and FAA Office of Runway Safety runway incursion video reenactments from archival data that constituted fixed data in the public record. The reliability of the instrument for the main study was tested in the pilot study and established a statistically significant ($p \leq .05$) inter-rater reliability ($k = 1.0$) with all cases alerting for a 100% true score agreement corresponding to a 100% proportion of positive ratings.

Instrument validity. The raters populated the data collection sheet by application of the SURF-IA model to nine *serious* runway incursion incidents already categorized by the RISC model and archived in the ASIAS database. The SURF-IA *aircraft states* for ownship and traffic were tabulated in SURF-IA logic diagrams with lettered columns and numbered rows (Figures 8, 9, and 10). The raters utilized the video reenactments and ASIAS report narratives to determine the *aircraft state* of ownship (e.g., Column E- entering/crossing runway-not lined up) and traffic (e.g., Row 3- take-off from same runway) and then followed the row and column intersection to obtain the *reason code* for either an *alerting* or *non-alerting* outcome (e.g., convergence-*reason code* E3).

The SURF-IA model was developed by subject matter experts from a broad range of aviation disciplines through the internationally recognized Radio Technical Commission for Aeronautics (RTCA). The RISC model was developed by subject matter experts from NASA and VOLPE National Transportation Systems Center, and was internationally recognized and utilized by the FAA and ICAO. The runway incursion reenactment videos were developed by the FAA Office of Runway Safety using precise position and timing data from mature surveillance systems routinely utilized and recognized by the NTSB for accident and incident reconstruction. All the instruments used in this study were pre-established by nationally and internationally recognized organizations (i.e., FAA, ICAO, RTCA) and had face validity. The pilot study discussed in Chapter 3 was conducted to establish the validity of the instrument.

Treatment of the Data

Statistical analysis was performed using the IBM® Statistical Package for the Social Sciences (*SPSS*®) Statistics computer software, and hand calculations where simplicity allowed. Kappa (*k*) was computed manually and with *SPSS*. The manual computations were adjusted for prevalence and bias to provide a prevalence-adjusted bias-adjusted kappa to assess inter-rater reliability. The calculation of kappa (*k*) for the proportion of agreement beyond what was expected by chance was manually computed from the data from the contingency table using equations (*1a*), (*1b*), and (*1c).* Prevalence index and bias index were obtained using equations (*1d*) and (*1e*).

$$P_o = (a + d)/n = (9 + 0)/9 = 1 \qquad (1a)$$

$$P_c = \left[\frac{(f1\ x\ g1)}{n} + \frac{(f2\ x\ g2)}{n}\right] / n = \left[\frac{(9\ x\ 9)}{9} + \frac{(0\ x\ 0)}{9}\right] / 9 = 1 \qquad (1b)$$

$$k = \frac{(Po - Pc)}{1 - Pc} = \frac{(1-1)}{1-1} = 1.0 \qquad (1c)$$

$$Prevalence\ Index = |a - d|/n = \ = |9 - 0|/9 = 1 \qquad (1d)$$

$$Bias\ Index = |b - c|/n = |0 - 0|/9 = 0 \qquad (1e)$$

Descriptive statistics. Counts and percentages of runway incursions with an outcome classified as *serious* by the RISC model were cross-tabulated with the rater's outcome for triggering a Warning or Caution *alert* from the SURF-IA model. Agreement between the raters was calculated for all paired ratings as an overall percentage agreement.

Reliability testing. Inter-rater reliability for the categorical variable of *alerting* or *non-alerting* was used to determine the consistency among raters by overall percentage agreement and then by Cohen's kappa statistic to calculate agreement beyond that expected by chance, as suggested by Leech, Barrett, and Morgan (2008). The magnitude of kappa was influenced by prevalence of the attribute of SURF-IA *alerting* for the runway incursion incidents as well as bias, which is the extent to which the raters disagree on the proportion of *alerting* or *non-alerting* cases. To assist in interpreting Cohen's kappa, a prevalence index and a bias index were computed using the formulas from Sim and Wright (2005), and Byrt, Bishop and Carlin (1993).

Qualitative data. The SURF-IA logic diagrams were annotated with a grid of ownship and traffic *aircraft states* that determined a specific *reason code* for each incident. A column was provided in the data collection device for the raters to indicate the *reason code* that determined whether or not the SURF-IA model alerted. The data collection device also included a column for rater free-form comments that were used to further explain the raters' determination of whether or not the SURF-IA model alerted.

The quantitative data only identified the *aircraft state* (e.g., after landing, take-off, approach) of ownship and traffic aircraft when the outcome of the SURF-IA model was not validated by matching a SURF-IA *alert* with a corresponding *serious* outcome from the RISC model. The qualitative comments from the free-form column on the data collection device were supplemented by a follow-up questionnaire (Appendix G) that was emailed to the raters, who typed in their responses to each question and then returned the completed document by email. The follow-up questionnaire was used to: (a) explain the raters' interpretation of ownship and traffic *aircraft states,* as defined by the SURF-IA algorithm, which led to their outcome assessments of either *alerting* or *non-alerting*; (b) identify the factors in the model(s) that influenced the different outcomes; (c) identify lessons learned; (d) identify other applications that might benefit from the methodology of using expert raters to validate legacy metrics; and (e) provide recommendations for modifying one or both of the models to harmonize the outcomes, and validate the SURF-IA model.

CHAPTER IV

RESULTS

The pilot study quantitative and qualitative results were consolidated for a side-by-side comparison, as depicted in Appendix F, and used to develop a 2 x 2 contingency table with descriptive statistics. The results from the pilot study identified modifications to the methodology that were incorporated into the main study. The rater data from the main study were consolidated in a similar manner (Appendix F). The descriptive statistics and reliability testing were in agreement in both the pilot study and the main study, which when analyzed in conjunction with the qualitative data, identified specific alerting and non-alerting *aircraft states.*

Pilot Study

In the pilot study there was 100% agreement between the two raters for the OE/D/I type *serious* runway incursion incidents that triggered a Warning or Caution *alert* from the SURF-IA model (Table 10).

Table 10

Pilot study rater cross-tabulation

Rater A * Rater B Cross-tabulation

			Rater B	
			Alert	Total
Rater A	Alert	Count	9	9
		% of Total	100.0%	100.0%
Total		Count	9	9
		% of Total	100.0%	100.0%

The calculation of kappa (*k*) for the proportion of agreement beyond what was expected by chance resulted in a kappa (*k*) of 1.0, as shown in equations *2a*, *2b*, and *2c*, indicating perfect inter-rater reliability (Table 11).

$$P_o = (a + d)/n = (9 + 0)/9 = 1 \qquad (2a)$$

$$P_c = \left[\frac{(f1\ x\ g1)}{n} + \frac{(f2\ x\ g2)}{n}\right] / n = \left[\frac{(9\ x\ 9)}{9} + \frac{(0\ x\ 0)}{9}\right] / 9 = 1 \qquad (2b)$$

$$k = \frac{(Po - Pc)}{1 - Pc} = \frac{(1-1)}{1-1} = 1.0 \qquad (2c)$$

A prevalence index of 1.0 and bias index of 0.0 informed the Kappa value, yielding a PABAK of 1.0 (Table 12). The kappa value for the main study was not computed using *SPSS* due to the expected misleadingly low kappa, which would have resulted from the influence of the high prevalence index and low bias index (Cunningham, 2009).

Table 11

Contingency table for pilot study of runway incursion alerting

		Rater A		
		Alert	No Alert	Total
Rater B	Alert	$a = 9$	$b = 0$	$g_1 = a + b = 9$
	No Alert	$c = 0$	$d = 0$	$g_2 = c + d = 0$
	Total	$f_1 = a + c = 9$	$f_2 = b + d = 0$	$N = f_1 + f_2 = 9$

Table 12

Prevalence and bias adjusted contingency table for pilot study. (Prevalence-adjusted bias-adjusted kappa (PABAK) of 1.0)

		Rater A		
		Alert	No Alert	Total
Rater B	Alert	*5*	*0*	$g_1 = a + b = 5$
	No Alert	*0*	*4*	$g_2 = c + d = 4$
	Total	$f_1 = a + c = 5$	$f_2 = b + d = 4$	$N = f_1 + f_2 = 9$

Although both raters agreed that all nine incidents considered in the pilot study would have triggered a SURF-IA alert, the *reason codes* were only identical in four of the nine cases. The raters' free-form comments provided insight into the explanation behind the differences for the other five cases, which after analysis were determined to be equivalent (Table C2). Specifically, the definition for the *aircraft states* for landing rollout versus taxing on-runway (lined-up) accounted for the different *reason codes* selected for Case 1 (F12 v. F10), and Case 3 (G13 v. G10). The *reason code* differences (A2 v. A1) for Case 5 was attributed to the raters' temporal assessment of the aircraft time to cross the runway threshold (RTHRE) during an approach; Warning *alert* if ≤ 15 seconds to RTHRE, or a Caution *alert* if ≤ 35 seconds to RTHRE. Although the raters selected different *reason codes* for Case 6 (C4 v. F16) and Case 9 (D4 v. F18), an examination of the SURF-IA logic diagrams derived from RTCA (2010) indicated that the *aircraft state* codes for intersecting runways were redundant, hence, interchangeable. One rater provided successive SURF-IA alerting *reason codes* for Cases 3, 5, and 8 as the incidents progressed from a Caution to a Warning *alert* condition or from a landing rollout to a taxiing on-runway (lined-up) aircraft state.

Based on the results of the pilot study and comments from the raters, the following modifications to the methodology were identified: (a) clarification of the difference between a landing rollout aircraft state and a taxiing on-runway (lined-up) aircraft state; (b) correction of a typographical error in one of the SURF-IA logic diagrams; and (c) additional instructions for the raters to only provide the *reason code* that triggered the first *alert*. Analysis of the pilot study results established a statistically significant ($p \leq .05$) inter-rater reliability ($k = 1.0$) with 100% proportion of positive ratings, which validated the instrument per the Sim and Wright (2005) guidance in Figure C1, with all cases alerting for a 100% true score agreement. The remainder of the results presented in this Chapter were from the main study.

Descriptive Statistics

There was 100% agreement between the two raters for the outcomes from the SURF-IA model for the nine pilot deviation type runway incursion incidents categorized as *serious* (Category A or B) by the RISC model (Table 13). Both raters agreed that six (66.7%) of the nine incidents would have a SURF-IA outcome of an *alert*, and three (33.3%) would not have alerted.

Table 13

Rater cross-tabulation

Rater A * Rater B Cross-tabulation

			Rater B		
			No Alert	Alert	Total
Rater A	No Alert	Count	3	0	3
		% of Total	33.3%	.0%	33.3%
	Alert	Count	0	6	6
		% of Total	.0%	66.7%	66.7%
Total		Count	3	6	9
		% of Total	33.3%	66.7%	100.0%

Reliability Testing

The overall agreement (P_o) was calculated as 1.0 by equation *(3a)*. The agreement expected by chance alone (P_c) was calculated as .55 using equation *(3b)*. The calculation of kappa (k) for the proportion of agreement beyond what was expected by chance was computed with *SPSS*, as shown in Table 14, and then manually computed from the data in the contingency table using equation (*3c*) (Table 15). Both calculations resulted in a kappa of 1.0. The calculated value of kappa can range from -1.0 to +1.0; however, for practical purposes only the range from 0.0 to 1.0 is of interest, where zero indicates no agreement beyond chance and 1.0 indicates perfect agreement. The standard error (*SE*) and the confidence interval for the 2 x 2 contingency table were manually calculated using equations (*3d*) and (*3e*), as suggested by Kundel and Polansky (2003), and yielded a value of zero, which agreed with the *SPSS* calculated *SE* value in Table 14. The kappa for this study was 1.0, with a corresponding *SE* of zero and a confidence interval of zero. A prevalence index of .33 and bias index of 0.0 were obtained using equations (*3f*) and (*3g*) to characterize the kappa value, yielding a prevalence-adjusted bias-adjusted kappa (PABAK) of 1.0 indicating complete agreement between the raters (Table 16).

Table 14

Symmetric measures

		Value	Std. Error^a	Approx. Sig.
Measure of Agreement	Kappa	1.000	.000	.003
N of Valid Cases		9		

Table 15

Contingency table for runway incursion alerting

		Rater A		
		Alert	No Alert	Total
Rater B	Alert	$a = 6$	$b = 0$	$g_1 = a + b = 6$
	No Alert	$c = 0$	$d = 3$	$g_2 = c + d = 3$
	Total	$f_1 = a + c = 6$	$f_2 = b + d = 3$	$N = f_1 + f_2 = 9$

$$P_o = (a + d)/n = (6 + 3)/9 = 1 \tag{3a}$$

$$P_c = \left[\frac{(f1\ x\ g1)}{n} + \frac{(f2\ x\ g2)}{n}\right] \Big/ n = \left[\frac{(6\ x\ 6)}{9} + \frac{(3\ x\ 3)}{9}\right] \Big/ 9 = .55 \tag{3b}$$

$$k = \frac{(Po - Pc)}{1 - Pc} = \frac{(1 - .55)}{1 - .55} = 1.0 \tag{3c}$$

$$SE \sim \sqrt{\frac{Po(1 - Po)}{n(1 - Pc)^2}} = \sqrt{\frac{1(1 - 1)}{9(1 - .55)^2}} = 0 \tag{3d}$$

$$CI_{95\%} = k \pm 1.96\ x\ SE = 1.0 \pm 1.96\ x\ 0 = 0 \tag{3e}$$

$$\text{Prevalence Index} = |a - d|/n = |6 - 3|/9 = .33 \qquad (3f)$$

$$\text{Bias Index} = |b - c|/n = |0 - 0|/9 = 0 \qquad (3g)$$

Table 16

Prevalence and bias adjusted contingency table. (Prevalence-adjusted bias adjusted kappa (PABAK) of 1.0)

		Rater A		
		Alert	No Alert	Total
Rater B	Alert	5	0	$g_1 = a + b = 5$
	No Alert	0	4	$g_2 = c + d = 4$
	Total	$f_1 = a + c = 5$	$f_2 = b + d = 4$	$N = f_1 + f_2 = 9$

Qualitative Data

Both raters agreed on the *reason codes* for all nine incidents, of which six were rated as providing an *alerting* SURF-IA outcome (Table 17). The free-form comments provided insight into the explanation for the *aircraft states* for the three incidents that did not provide a SURF-IA *alert* outcome. The factors that precluded a SURF-IA outcome of a Warning or Caution *alert* were: (a) SURF-IA model did not *alert* for a single aircraft wrong runway departure; (b) SURF-IA model did not *alert* for helicopter runway incursions; and (c) SURF-IA model did not *alert* for ownship entering or crossing the runway and being overflown by another aircraft taking off on the same runway when the aircraft on take-off had already lifted off prior to ownship entering the runway (Table 18). All three of the aforementioned factors that precluded a SURF-IA outcome of a Warning or Caution *alert* would have been rated as *serious* by the RISC model.

Table 17

Ownship/Traffic pairs from serious (Category A or B) runway incursion incidents with an alerting SURF-IA outcome

		Ownship Aircraft State			
		Entering/Crossing Runway (Not Lined up)	Take-off < 80 knots	Approach to Runway (≤ 3 nm from Runway)	Stopped or Taxiing on Same Runway (Lined Up)
Traffic Aircraft State	Take-off from Same Runway	ASIAS ID 5826 ASIAS ID 7167 ASIAS ID 3374			
	Entering or Crossing Runway (Not Lined Up)		ASIAS ID 8173		
	Stopped or Taxiing on Same Runway (Lined Up)			ASIAS ID 11322	
	Approach to Runway (≤ 3 nm from Runway				ASIAS ID 4828

Table 18

Ownship/Traffic pairs from serious (Category A or B) runway incursion incidents without an alerting SURF-IA outcome

ASIAS ID	Ownship State	Traffic State	Factor
10923	Entering/Crossing Runway (Not Lined Up)	Take-off from same Runway	Traffic lifted off prior to Ownship entering the runway
10675	Entering/Crossing Runway (Not Lined Up)	After Landing Roll-out on Runway	Ownship was a helicopter
10969	Take-off < 80 knots	Not Applicable	Single aircraft wrong runway departure

A common theme in the raters' qualitative comments to the follow-up questionnaire shown in Appendix G was the lack of clear or harmonized definitions for the *aircraft states* used in the models for the SURF-IA technology and the legacy RISC metrics that were used to validate the benefit of the new technology. The *aircraft state* definition mentioned by the raters as being most troublesome was *on-runway, as* highlighted in Case 6. Both raters indicated that neither the legacy RISC definition of *on-runway* nor the SURF-IA definition was appropriate. Hence, in this case the raters' recommendation was not to adopt either model's definition, but rather develop an entirely new definition through further study. The raters were emphatic about the criticality of appropriate and harmonized definitions used in the models. Consequently, a specific step was included in the step-by-step methodology that was developed by this study for assessing the validity of legacy/traditional metrics for application to new technology, as presented in Chapter 5. The raters also suggested that the methodology of using expert raters, as presented in this study, could be applied in the systems engineering for any new technology: (a) during the concept development stage, this could help expert raters define the desired performance of the system; (b) during engineering development, expert raters could validate the requirements; and (c) during post implementation, expert ratings could be used to identify shortcomings and validate the functioning of the system. One of the delimitations of the study was that the raters only assessed whether or not a SURF-IA *alert* would have been issued, without considering late *alerts* or nuisance *alerts* when assessing the effectiveness of the system. However, both raters commented on the importance of considering late *alerts* and nuisance *alerts* when assessing the effectiveness of the system. Specifically, one of the raters recommended further analysis through a longitudinal study involving both normal operations and runway incursion incidents at multiple airports to assess correct alerting as well as nuisance and late *alerts* associated with non-serious (Category C and Category D) runway incursion incidents

> To define the actual requirements for an incursion alerting system, a large-scale validation study with empirical surveillance data from multiple airports and over extended periods of time needs to be performed to determine the overall system performance in terms of nuisance and true alerts. Not only runway incursion events should be considered but especially, normal operations where the system should not provide alerts. Making design decisions based on single events or incidents can be misleading because while the system may be optimized for these events, the overall system that needs to run under a much wider set of conditions may be operationally unacceptable. (Rater B, 2013)

This study revealed how different outcomes from the RISC and SURF-IA models may result in misleading information when using the reduction in runway incursion incidents classified as *serious* by the RISC model, as a metric for the benefit of SURF-IA technology. Expressly, the study revealed that the SURF-IA model did not yield an outcome of a Warning or Caution *alert* for all runway incursion incidents classified as *serious* (Category A or B) by the FAA/ICAO RISC model. There were specific *aircraft states* in the baseline SURF-IA model that precluded an outcome of a Warning or Caution *alert* for runway incursion incidents classified as *serious* (Category A or Category B) by the FAA/ICAO RISC model.

CHAPTER V

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

Discussion

The raters were in complete agreement that six of the nine *serious* (Category A or B) PD type runway incursion incidents would have triggered a SURF-IA *alert* outcome and three incidents would not have triggered an *alert*, resulting in a bias index of 0 and an overall 100% proportion of rater agreement. The mid-range prevalence index of .33 characterized a balance of *alerting* and *non-alerting* ratings, indicating that the 100% rater agreement attained was not by chance alone, which further supported the kappa value of 1.0 for perfect inter-rater reliability. The kappa statistic alone without adjustments for prevalence was appropriate and sufficient to assess inter-rater reliability for this study because the marginal totals for the 2 x 2 contingency table were relatively balanced (Cunningham, 2009).

This study, which only considered nine PD type *serious* (Category A or B) runway incursion incidents, identified three incidents (33%) with *aircraft states* that did not trigger a SURF-IA *alerting* outcome. Three of the 11 known limitations of the baseline SURF-IA model presented in Table 3 manifested themselves as factors in the *non-alerting* incidents: (a) wrong runway departures, (b) no alert if traffic enters runway after ownship lift-off from same runway, and (c) helicopter operations. The limitations of the SURF-IA model were known by RTCA SC-186, which developed the SURF-IA SPR. However, prior to this study, the validity of using RISC derived runway incursion statistics to measure SURF-IA effectiveness had not been explicitly identified or considered.

Seven of the nine incidents in this study involved the most prevalent *aircraft state* for runway incursions, which was when ownship aircraft entered a runway ahead of traffic aircraft departing or landing. All were rated with a SURF-IA *alerting* outcome except for the incident that involved ownship entering the runway after the traffic on take-off from the same runway had lifted-off; this is a known limitation of the SURF-IA model where it does not *alert* for an incident classified as *serious* by the RISC model. The two other incidents that were rated with a SURF-IA *non-alerting* outcome also involved aircraft states that were known limitations of the SURF-IA model: wrong runway departures, and helicopter runway incursions. Potential issues and solutions for the *aircraft states* outside the capability of the SURF-IA *alerting* model had been previously identified by Moertl et al. (2012, June), but they had not been considered in the context of harmonizing the SURF-IA outcomes with the RISC model.

The on-runway condition for the SURF-IA model only extended the aircraft's position to the runway shoulder, while the RISC model on-runway condition extended past the runway shoulder to the taxiway hold line (Figure 7). A broader SURF-IA definition of an on-runway aircraft condition/state would provide a SURF-IA *alerting* outcome for an ownship aircraft entering a runway occupied by traffic aircraft on take-off, even though the traffic may lift off prior to ownship actually crossing the runway shoulder. The SURF-IA definition of on-runway is when the aircraft crosses the runway shoulder, while the RISC model defines an on-runway condition as anytime the aircraft is beyond the hold line for the runway (Figure 7).

Single aircraft wrong runway departures were addressed by Moertl et al. (2012, June) who suggested that an ADS-B transponder could be placed on the approach and departure ends of runways that were inactive, closed, or otherwise not in use, as *mock* traffic which would trigger a SURF-IA *alert* for an aircraft departing/arriving on the wrong runway. The modified SURF-IA model, proposed by Moertl et al. (2012, June), interpreted a wrong runway departure as traffic on the runway.

A wrong runway departure involving a single aircraft is classified as a *serious* runway incursion by the RISC model; however, the SURF-IA model only provides *alerts* for incidents involving two aircraft. Hence, all runway incursion incidents from wrong runway departures, even if the aircraft were SURF-IA equipped, would reflect as an increase in rate and number of runway incursions. The potentially misleading statistical analysis of the benefit of SURF-IA for runway incursion data when it fails to *alert* for wrong runway departures classified as *serious* by the RISC model, was estimated by looking at historical data for the number of wrong runway departures. An FAA (2007, July) report on U.S. domestic wrong runway departures indicated that from CY 1981-CY 2006 there were 696 incidents or accidents involving wrong runway operations. These data were collected prior to the FAA adopting the ICAO definition of runway incursions that added wrong runway departures. From FY 2008-FY 2013 (January), under the expanded definition of runway incursion, the ASIAS database recorded 23 wrong runway incidents. All of the aforementioned runway incursions, which involved single aircraft wrong runway departures with a subsequent loss of separation from another aircraft, would have been classified by the RISC model as *serious.* However, none would have resulted in a SURF-IA *alerting* outcome because the SURF-IA model does not provide *alerts* for incidents involving one aircraft. Hence, these wrong runway departure incidents would have been interpreted as a failure of the SURF-IA technology.

The baseline version of SURF-IA was not intended for installation on helicopters; however, the incorporation of helicopter operations into the SURF-IA model was addressed by RTCA (2010) as follows:

> The performance and safety analysis did not analyze helicopter installations. Therefore, the assumption excludes this configuration from the scope of the application. However, due to the similarity of surface operations of helicopters and airplanes while operating on standard airport surface elements such as runways and taxiways (helicopters on, or hovering above taxiways, holding short of runways, taking off and landing on runways, on approach to runways/pad) helicopters may actually be able to safely operate. However, it is expected that a simple add-on safety analysis may be able to show that helicopters could safely operate SURF IA. (RTCA, 2010, p. A-40)

Joslin (2013) identified the following considerations for the SURF-IA helicopter add-on safety analysis recommended by RTCA (2010) SC-186: (a) air/ground determination for an airborne helicopter in a hover or hover taxiing; (b) helipads as a surface identified for take-off/landing, but not located on an actual runway surface; (c) helicopters entering/crossing runways from areas other than known taxiways; and (d) helicopters on approach to a runway, but not lined up with the runway centerline.

Runway incursions were not confined to airplanes, and neither the number nor the rate of helicopter runway incursions has shown any appreciable reduction (Figure 21 and Figure 22). However, the overall statistics for rates and number of runway incursions, as published by the FAA Office of Runway Safety, do not distinguish between airplanes and helicopters. Hence, the exclusion of valid SURF-IA *alerting* outcomes in ownship helicopters may influence the validity of runway incursion statistics, derived from the RISC model, as a measure of the effectiveness of SURF-IA technology for the reduction PD type runway incursions.

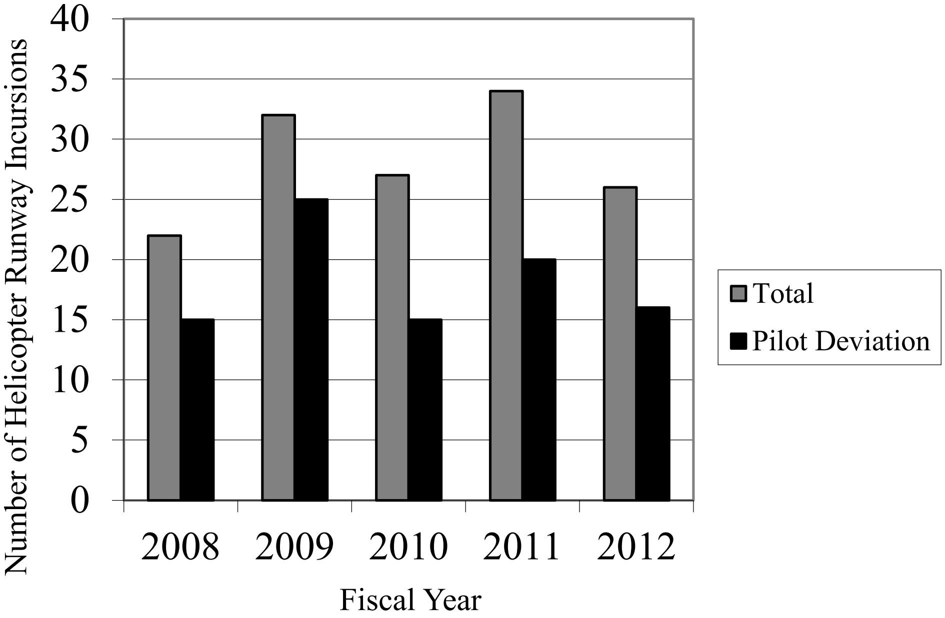


Figure 21. Number of helicopter runway incursions FY 2008-FY 2012. Adapted from http://www.asias.faa.gov/

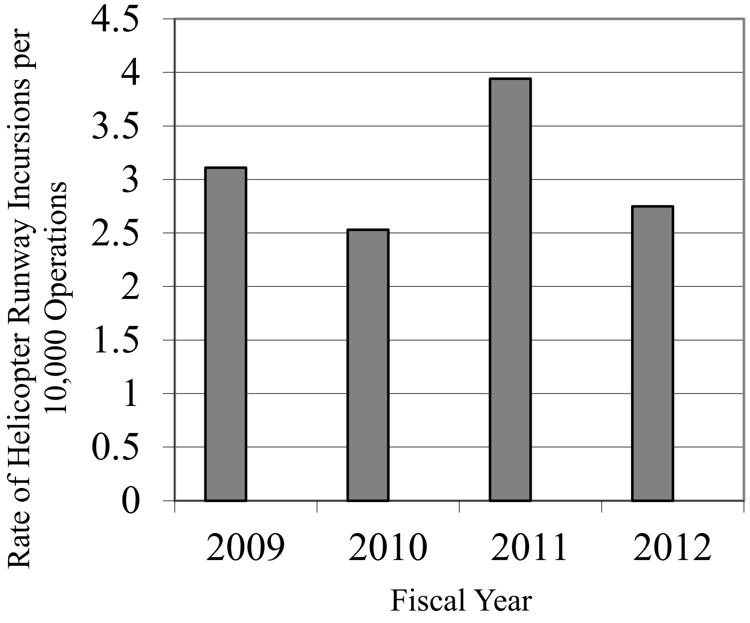


Figure 22. Rate of helicopter runway incursions per 10,000 operations, FY 2009-FY 2012. Adapted from https://aspm.faa.gov/tfms/ and ASIAS- http://www.asias.faa.gov/

Whereas the existing literature and previous studies did not define or follow a formal rigorous and repeatable process for validating metrics, this study developed a step-by-step methodology that filled the gap for assessing the validity of legacy/ traditional metrics for application to new technology (Table 19).

Table 19

Step-by-step methodology for validating metrics for new technology

1	Determine the intended function of the new technology. For this study, the intended function of the new technology was the reduction in serious runway incursions.
2	Identify the model for the traditional or legacy metric used to measure the outcome from the intended function of the new technology. For this study, the RISC model was the legacy metric used for categorizing the severity of runway incursions.
3	Identify the model for the new technology (e.g., SURF-IA).
4	Identify the limitations of the technical capabilities of the new technology (e.g., SURF-IA model will not alert for single aircraft wrong runway departures).
5	Identify any differences in definitions between the models (e.g., on-runway condition was defined differently in the RISC model versus the SURF-IA model).
6	Identify expert raters in the field of the new technology.
7	Gather archival data from actual cases of interest that have already been classified by the legacy/traditional metric. For this study the cases of interest were pilot deviation type runway incursion incidents classified as serious (Category A or B) by the RISC model.
8	Select a sufficient number of cases of interest to establish a statistically significant sample size.
9	Have the expert raters apply the model for the new technology (e.g., SURF-IA) to the cases of interest, and determine the outcome from the new technology (e.g., alerting or non-alerting).
10	Gather qualitative comments from the raters to: • explain why or how they determined their rating • provide lessons learned • identify which cases were most troublesome • recommend modifications to the model(s)
11	Calculate the inter-rater reliability, and descriptive statistics (e.g., percentage agreement, counts.
12	Identify the cases and conditions where the outcome from the metric used to measures the benefit of the new technology does not match the outcome from the new technology, as assessed by the expert raters. The cases identified in this study were those where the SURF-IA model did not yield an outcome of a Warning or Caution alert for runway incursion incidents classified as serious (Category A or B) by the RISC model.
13	Identify modifications to the model(s) that would harmonize the metrics with the outcome of the new technology.

Conclusions

The study answered both research questions. (1) Does the SURF-IA model yield an outcome of a Warning or Caution *alert* for runway incursion incidents classified as *serious* (Category A or B) by the FAA/ICAO RISC model? The study revealed that the SURF-IA model did not yield an outcome of a Warning or Caution *alert* for all runway incursion incidents classified as *serious* (Category A or B) by the FAA/ICAO RISC model. (2) What are the *aircraft states* in the SURF-IA model that preclude an outcome of a Warning or Caution *alert* for runway incursion incidents classified as serious (Category A or Category B) by the FAA/ICAO RISC model? There were specific *aircraft states* in the baseline SURF-IA model that precluded an outcome of a Warning or Caution *alert* for runway incursion incidents classified as *serious* (Category A or Category B) by the FAA/ICAO RISC model: (a) wrong runway departures, (b) traffic entering the runway after ownship lift-off from same runway, and (c) helicopter operations. This study also revealed how different outcome severities from the RISC and SURF-IA models may result in misleading information when using the reduction in *serious* runway incursion incidents, classified by the RISC model, as a metric for the benefit of SURF-IA technology.

In FY 2012 there were 10 *serious* (Category A or B) pilot deviation type runway incursions, which was a tenfold increase over the one (1) runway incursion of this type and category reported in FY 2011. This study used four of the ten incidents recorded in FY 2012, of which three were rated as *non-alerting* by the SURF-IA model. If this study had assumed that all aircraft involved in the FY 2012 incidents had SURF-IA equipment installed, and then used the change in PD type runway incursions classified as *serious* by the RISC model as a metric to assess the effectiveness and benefit of SURF-IA, at least three of the ten FY 2012 incidents would not have provided a SURF-IA alert. Hence, the FY 2012 runway incident data would have been misleading by indicating that the SURF-IA model was at best only 70% effective and beneficial in providing an *alert* to mitigate the hazard from runway incursions.

However, the three incidents that were *non-alerting* involved three *aircraft states* identified in this study as not providing a SURF-IA alerting outcome for an incident classified as *serious* by the RISC model. Even if the SURF-IA model had performed to design, the best it could have achieved would have been a 70% alerting outcome for incidents classified as *serious* by the legacy RISC model metric.

This study demonstrated an innovative method of utilizing expert raters and actual high-risk incidents to identify the shortcomings of using legacy metrics to measure the effectiveness of new technology designed to mitigate hazardous incidents. The expansion of the methodology used in this study to other areas lies in first identifying the known limitations and capabilities in the actual design of any new technology and then using expert raters to see if, and how, the outcomes from legacy metrics were affected. If the model differences yield outcomes that do not match, the design of the new technology and/or the design of the metric for measuring the benefit of the new technology may need adjustment. The overall implication from this study is that the implementation of new technology designed to mitigate a legacy hazard demands a concurrent re-assessment of the legacy metrics. The methodology is generalizable and can be applied to other high-risk areas, such as medicine, nuclear power plants, and other modes of transportation.

Recommendations

Prior to the certification of SURF-IA for use on aircraft, it is recommended that further study with a larger number of PD type runway incursion incidents classified as *serious* (Category A or B) by the RISC model is conducted to identify other *aircraft states* and associated factors that do not trigger a SURF-IA *alerting* outcome. It is also recommended that prior to using the ASIAS runway incursion data as a metric for the benefit of SURF-IA, the FAA develop a process for identifying and tracking ASIAS reported PD type *serious* runway incursion incidents which will not trigger an *alerting* outcome in the baseline SURF-IA. Data from the runway incursion incidents for Cases 6, 7, and 8 involving runway *aircraft states* not designed to trigger an *alert* by the baseline SURF-IA model should not be considered when assessing the effectiveness and benefit of the new SURF-IA technology for reducing runway incursion incidents. However, consideration should be made to improving the SURF-IA model technical capabilities to accommodate all possible *aircraft states* that the RISC model would classify as *serious* (Category A or B) runway incursion incidents.

REFERENCES

Arany, J. G. (2012). *The reliability of a coding system for capturing the content of family mediation agreements* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3517074)

Automatic Dependent Surveillance Broadcast (ADS-B), 76 Fed. Reg. 71430 (2011).

Aviation Communication and Surveillance Systems (ACSS) (2010). *Flight test demonstration report for enhanced situational awareness on the airport surface with indications and alerts* (8008170-001). Retrieved from http://www.acss.com/videos/faa-surface-indicating-alerting-demonstration/#

Barach, P., & Small, S. D. (2000). Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems. *British Medical Journal, 320(7237),* 759-763.

Bughin, J., Shenkan, A. G., & Singer, M. (2009). How poor metrics undermine digital marketing. *McKinsey Quarterly*, *4*, 38-61.

Byrt, T., Bishop, J., & Carlin, J. (1993). Bias, prevalence, and kappa. *Journal of Clinical Epidemiology, 46*(5), 423-429.

Cardosi, K., Hannon, D., Sheridan, T., & Davis, W. (2005). A method for rating the Severity of runway incursions. *Proceedings of the USA/Europe 6th Air Traffic Management Research and Development Seminar*.

Cardosi, K., Chase, S., & Eon, D. (2010). Runway Safety. *Air Traffic Control Quarterly*, *18*(3), 303.

Cassell, R., Evers, C., Esche, J., & Sleep, B. (2002). *NASA runway incursion prevention* *system Dallas-Fort Worth demonstration performance analysis (*NASA/CR-2002-211677). Hampton, VA: NASA.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement (20)*1, 37-46.

Collision Avoidance Systems (CAS), 68 Fed. Reg. 15884-15904 [2003] (to be codified at 14 C.F.R. pt. 121,125, 129)

Cooper, G.E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities* (NASA TN D-153). Washington, DC: NASA.

Cunningham, M. (2009). More than just the kappa coefficient: A program to fully Characterize inter-rater reliability between two raters. In *SAS Global Forum* (pp. 242-2009).

D'Amato, B. (2008). *A study of threats against the dream self in dreams of adult adopted women as manifestations of psychic loss* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3350567)

Donner, A., & Eliasziw, M. (1992). A goodness-of-fit approach to inference procedures for the kappa statistic: Confidence interval construction, significance-testing and sample size estimation. *Statistics in Medicine*, *11*(11), 1511-1519.

European Aviation Safety Agency (EASA) (2010). *Annual safety review 2010.* Retrieved from http://www.easa.eu.int

Federal Aviation Administration (FAA) (1989). *Airport design* (FAA Advisory Circular 150/5300-13). Retrieved from http://faa.rgl.gov/

Federal Aviation Administration (FAA) (2002). *Runway incursion causal analysis*.

Retrieved from http://www.faa.gov/airports/runway_safety/ace/training_ presentations/ index.cfm?print=g

Federal Aviation Administration (FAA) (2004). *FAA runway safety report: Runway incursion trends and initiatives at towered airports in the United States, fy 2000– fy 2003*. Retrieved from http://www.faa.gov/airports/runway_safety /media/pdf/report4.pdf

Federal Aviation Administration (FAA) (2006, August). *Memorandum from ATO safety services; Use of runway incursion severity classification (risc) calculator*. Washington, DC: ATO Office of Safety.

Federal Aviation Administration (FAA) (2007a). *Flight Standards Information Management Systems [FSIMS]* (FAA Order 8900.1). Retrieved from http://fsims.faa.gov/

Federal Aviation Administration (FAA) (2007b). *Economic Values for FAA Investment and Regulatory decisions: A Guide* (Report No. DTFA-01-02-C00200). Washington, DC: Government Printing Office.

Federal Aviation Administration (FAA) (2007c). *FAA compliance and enforcement program* (FAA Order 2150.3B). Retrieved from http://rgl.faa.gov/

Federal Aviation Administration (FAA) (2007, July). *Wrong runway departures*. Retrieved from http://www.asias.faa.gov/pls/apex/f?p=100:8:0::NO ::P8_STDY_VAR:3

Federal Aviation Administration (FAA) (2009). *National runway safety plan 2009-2011*. Retrieved from http://www.faa.gov/airports/runway_safety/news/ publications/media/runwaysafetyreport-kh10-plan.pdf

Federal Aviation Administration (FAA) (2010a). *Runway safety program* (FAA Order 7050.1A). Retrieved from http://rgl.faa.gov

Federal Aviation Administration (FAA) (2010b). *Flight crew alerting* (FAA Advisory Circular 25.1322-1). Retrieved from http://rgl.faa.gov/

Federal Aviation Administration (FAA) (2010c). *Air traffic organization aircraft accident and incident notification, investigation, and reporting* (JO 8020.16A). Retrieved from http://rgl.faa.gov

Federal Aviation Administration (FAA) (2011a, March). *AJS-4 Runway Incursion Assessment Process* (AJS-RSO-RRI-T03-F-001).Washington, DC: ATO Office of Safety.

Federal Aviation Administration (FAA) (2011b, March). *Runway Incursion Severity Assessment Guidance* (AJS-RSO-RRI-T03). Washington, DC: ATO Office of Safety.

Federal Aviation Administration (FAA) (2011a). *FAA aerospace forecast FY 2012-2032.* Retrieved from http://www.faa.gov/about/office_org/headquarters _offices/apl/aviation _forecasts/aerospace_forecasts/2012-2032/

Federal Aviation Administration (FAA) (2011b). *Airworthiness approval for ADS-B(In) systems and applications* (FAA Advisory Circular 20-172). Retrieved from http://faa.rgl.gov/

Federal Aviation Administration (FAA) (2011c). *National Runway Incursion Totals.* Retrieved from http://www.faa.gov/airports/runway_safety/statistics/

Federal Aviation Administration (FAA) (2012a). *FAA NextGen implementation plan*. Retrieved from http://www/faa.gov/about/initiatives/nextgen/media/

Federal Aviation Administration (FAA) (2012b). *Destination 2025*. Retrieved from http://www.faa.gov/about/plans_reports/media/Destination2025.pdf

Federal Aviation Administration (FAA) (2012c). *Air Traffic Organization Occurrence Reporting* (JO 7210.632). Retrieved from http://rgl.faa.gov

Federal Aviation Administration (FAA) (2012d). *Report on the risk mitigation study related to ads-b surface detection performance issues in support of future ads-b surface operations*. Washington, DC: FAA Surveillance and Broadcast Services Program.

Federal Aviation Administration (FAA) (2012e). *Fact Sheet- Aviation Safety Information and Analysis Sharing System.* Retrieved from http://www.faa.gov/news/fact_sheets/news_ story.cfm?newsId=13732

Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382.

Government Accountability Office [GAO] (2007). *Aviation-runway and ramp safety: Sustained efforts to address leadership, technology, and other challenges needed to reduce accidents and incidents* (GAO-08-29). Washington DC: Government Printing Office.

Green, A. (1997). Kappa statistics for multiple raters using categorical classifications. *Proceedings of the 22nd Annual Conference of SAS Users Group.*

Green, D. (2002). *Runway safety monitor algorithm for detection and alerting* (NASA/CR-2002-211416). Hampton, VA: NASA.

Green, D. (2006). *Runway safety monitor algorithm for single and crossing runway* *incursion detection and alerting* (NASA/CR-2006-214275). Hampton, VA: NASA.

Green, D., Otero, S., Barker, G., & Jones, D. (2009). *Initial concept for terminal area conflict detection, alerting, and resolution capability on or near the airport surface* (NASA/TM-2009-215696). Hampton, VA: NASA.

Gwet, K. L. (2001). *Handbook of inter-rater reliability*. Advanced Analytics, LLC.

Harper, R. P., & Cooper, G. E. (1966). *A Revised Pilot Rating Scale for the Evaluation Of Handling Qualities*. Cornell Aeronautical Laboratory, Incorporated.

Harper, R. P., & Cooper, G. E. (1986). Handling qualities and pilot evaluation. *Journal Of Guidance, Control, and Dynamics*, *9*(5), 515-529.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, *1*, 139-183.

Hendriksen, I., & Holewijn, M. (1999). Degenerative changes of the spine of fighter pilots of the Royal Netherlands Air Force (RNLAF). *Aviation Space Environmental Medicine*, 70, 1057–1063.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing & health*, *31*(2), 180-191.

Honeywell International (2010). *SURF-IA Flight Evaluation Data Report* (FAA agreement DTFAWA-09-A-00001).

International Civil Aviation Organization [ICAO] (2007). *Manual on the prevention of runway incursions.* Retrieved from http://legacy.icao.int/fsix/_Library/ Runway%20 Incursion%20Manual-final_full_fsix.pdf\

Jones, D. R., Chartrand, R.C., Wilson, S.R., Commo S. A., Otero, S.D., & Barker G.D. (2012). SURF-IA conflict detection and resolution algorithm evaluation. *Proceedings of the AIAA/IEEE 31st Digital Avionics Systems Conference.*

Jones, D. R. (2002). Runway incursion prevention system simulation evaluation *Proceedings from the AIAA/IEEE 21st Digital Avionics Systems Conference.*

Jones, D. R. (2004). Runway incursion prevention system testing at the Wallops Flight Facility. *Proceedings from the SPIE Defense and Security Symposium*.

Jones, D. R., Quach, C.C., & Young, S. D. (2001). Runway incursion prevention system demonstration and testing at the Dallas-Fort Worth international airport. *Proceedings of the AIAA/IEEE 20th Digital Avionics Systems Conference.*

Jones, D. R., & Prinzel, L. J. (2006). Runway incursion prevention for general aviation operations. *Proceedings of the AIAA/IEEE 25th Digital Avionics Conference*.

Jones, D. R., Prinzel, L. J., Otero, S. D., & Barker, G. D. (2009). Collision avoidance for airport traffic concept evaluation. *Proceedings of the AIAA/IEEE 25th Digital Avionics Conference.*

Jones, D. R., Prinzel, L. J., Shelton, K. J., Bailey, R. E., Otero, S. D., & Barker, G. D. (2010). Collision avoidance for Airport traffic simulation evaluation. *Proceedings of the AIAA/IEEE 29th Digital Avionics Conference.*

Joslin, R. E., Goodheart, B. J., & Tuccio, W. A. (2011). A mixed method approach to Runway incursion rating. *International Journal of Applied Aviation Studies*, *11*(2), 13-36.

Joslin, R. E. (2013). NextGen flight deck technology for runway incursion awareness, *Helicopter Association International, Heli-Expo 2013-Rotor Safety Challenge*, [PowerPoint Slides], Las Vegas, NV.

Kilpikoski, S., Airaksinen, O., Kankaanpää, M., Leminen, P., Videman, T., & Alen, M. (2002). Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine 27*(8), E207-E214. Retrieved from http://www.ncbi .nlm .nih.g ov/pubmed/11935120

Kundel, H. L., & Polansky, M. (2003). Measurement of observer agreement. *Radiology*, *228*(2), 303-308.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data coefficient of agreement for nominal scales. *Biometrics,(33)*1, 159-174.

Lascara, B. J. , Moertl, P. (2012). *Surface surveillance analysis platform for ads-b surface applications.* McLean, VA: MITRE Corporation.

Latimer, B. A. (2012). *Conceptual model or collision detection and avoidance for Runway incursion prevention* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3516876)

Laupacis, A., Feeny, D., Detsky, A. S., & Tugwell, P. X. (1992). How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ: Canadian Medical Association Journal*, *146*(4), 473.

Laura, D. G., & William, L. G. (1999). Research design and methodology section: Measurement myths and misconceptions. *School Psychology Quarterly, 14*(4), 408-408.

Leech, N.L., Barrett, K. C., &, Morgan, G. A. (2008). *SPSS for Intermediate Statistics*, New York, NY: Psychology Press.

Mata, S. (1993). *Chest x-rays as a diagnostic tool for diffuse idiopathic skeletal Hyperostosis* (Master's Thesis). Available from ProQuest Dissertations and Theses. (UMI No. MM98153)

McKenzie, R. (1981). *The Lumbar Spine: Mechanical Diagnosis and Therapy*. Waikanae, New Zealand: Spinal Publications LTD.

Moertl, P. M., & Nickum, J. D. (2008). Flight deck based indications and alerting to increase runway safety. *Proceeding from the 2008 ICNS Conference*.

Moertl, P. M., Lascara, B., Higgins, M.,& Baker, J. (2012, June). *Results of an analysis to estimate safety benefits of enhanced situation awareness on the airport surface with indications and alerts (SURF IA).* [PowerPoint slides]. Retrieved from http://www.mitre.org/

National Transportation Safety Board (NTSB) (2000). *Safety recommendation letter to the FAA administrator* (A-00-66). Retrieved from http://www.ntsb.gov/doclib /recletters /2000/a00_66_71.pdf

National Transportation Safety Board (NTSB) (2007). *Attempted takeoff from wrong runway Comair flight 5191 Bombardier CL-600-2B19, N431CA Lexington, Kentucky-accident report* (NTSB/AAR-07-05). Retrieved from http://www/ntsb.gov

National Transportation Safety Board (NTSB) (2012). *Most wanted list of safety improvements*. Retrieved from http://www/ntsb.gov/safety/mwl.html

Office of Special Counsel (OSC) (2008). *Analysis of disclosures, agency investigation and report, whistleblower comments and comments of the associate special counsel* (OSC File DI-08-1708). Washington, DC: OSC.

Prinzel, L. J., & Jones, D. R. (2006). *Cockpit technology for the prevention of general aviation-runway incursions.* Hampton, VA: NASA.

Radio Technical Commission for Aeronautics (RTCA) (2010). *Safety, performance and interoperability requirements document for enhanced traffic situational awareness on the airport surface with indications and alerts (SURF-IA)*. Washington, DC: RTCA.

Reid, G. B.,& Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam, The Netherlands: Elsevier.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use* (Tech. Rep. No. RAE-TR-90019; RAE-FM-6; BR114794; ETN-91-98364). Farnborough, England: Royal Aerospace Establishment.

Russell, M. G. (2009). A call for creativity in new metrics for liquid media. *Journal of Interactive Advertising, 9*(2).

Sheridan, T. (2004). An interpolation method for rating the severity of runway incursions. *Proceedings of the Symposium on Human Performance, Situational Awareness and Automation.*

Sheridan, T., Cardosi, K., & Hannon, D. (2004). Rating the severity of close-call events. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.*

Shoukri, M. (2004). *Measures of inter-observer agreement*. Boca Raton, FL: Chapman & Hall.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Journal of the American Physical Therapy Association, 85*(3), 257-268.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L., ... & Goldsmith, C. (2010). A tutorial on pilot studies: the what, why and how. *BMC medical research methodology*, *10*(1), 1.

van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social research update*, (35), 1-4.

Wrisley, D. M. (1998). *Reliability of the dynamic gait index in vestibular disorders* (Master's Thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1393363).

Young, S. D., & Jones, D. R. (2001). Runway incursion prevention: A technology solution. *Proceedings of the Flight Safety Foundation's 54th Annual International Air Safety Seminar.*

Zuschlag, M. (2005). *Violations of temporary flight restrictions and air defense identification zones: An analysis of airspace violations and pilot report data* (NASA/CR-2005-213923). Cambridge, MA: VOLPE.

APPENDIX A

Permission to Conduct Research

Embry-Riddle Aeronautical University

Application for IRB Approval

Determination Form

13-144

Principle Investigator: Robert Edward Joslin
Other Investigators:

Project Title: ***Examination of SURF-IA Alerting Outcomes for Serious Runway Incursion Incidents***

Submission Date: January 8, 2013

Determination Date: January 25, 2013

Review Board Use Only

Initial Reviewer: Teri Vigneau/Bert Boquet

Exempt: X Yes __ No

Approved: X Yes ___ No

Comments: The purpose of this study will be to compare the outcomes of runway incursion severity classifications of serious derived from the FAA/ICAO RISC model and flight-deck alerts from SURF-IA flight deck equipage. The research will utilize publically available reports and video re-enactments from the FAA Runway Safety Office and the Aviation Safety Information Analysis and Sharing (ASIAS) system. Since this research is using existing data and there will be no risks to participants it may be determined to be **exempt.** [Teri Vigneau 1-10-13]

This protocol is **exempt**. [Bert Boquet 1-17-13]

CONSENT FORM
Embry-Riddle Aeronautical University

You are invited to participate in a research study related to runway incursions, which continue to be a source of risk to air transportation; the issue is well documented by the FAA, NTSB, ICAO, and other international organizations. You were selected as a possible rater because you are an aviation professional with experience in flight-deck technology, and runway incursions. We ask that you read this form and ask any questions you may have before agreeing to be a rater for this study. The study is being conducted by Robert E. Joslin, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Aviation.

STUDY PURPOSE

The purpose of this study will be to compare the outcomes of runway incursion severity classifications of *serious* derived from the FAA/ICAO RISC model and flight-deck *alerts* from SURF-IA flight-deck equipage. The research will utilize publically available reports and video re-enactments from the FAA Runway Safety Office and the Aviation Safety Information Analysis and Sharing (ASIAS) system.

NUMBER OF PEOPLE TAKING PART IN THE STUDY:

If you agree to participate, you will be one of a minimum of 2 raters who will be participating in this research by rating alerting outcomes from flight-deck equipage for 18 runway incursion incidents.

PROCEDURES FOR THE STUDY:

If you agree to be a rater for this study, you will do the following things:

1. Rate 18 runway incursion incidents within four weeks of receipt of a rater package by mail which will include a media storage device(flash drive) with videos and document files, as well as paper copies of the documents.
2. Not reproduce or share any of the items and will return them to this investigator along with a completed rater matrix (Appendix B).
3. Rate the incidents independently without discussion with any other person or reference to any other information.
4. You will not be expected to travel to any location but will require a personal computer with word processing (.doc and .docx) and the ability to view video files (.swf, .exe, flashplayer, internet explorer).

RISKS OF TAKING PART IN THE STUDY:

There are no foreseeable risks associated with this project. Your individual identity will not be included in the research study.

BENEFITS OF TAKING PART IN THE STUDY:

There are no direct benefits for taking part in the study.

CONFIDENTIALITY

Efforts will be made to keep your personal information confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Your identity will be held in confidence in reports in which the study may be published and databases in which results may be stored.

PAYMENT

You will not receive payment for taking part in this study.

CONTACTS FOR QUESTIONS OR PROBLEMS

For questions about the study, contact the principal investigator Robert Joslin at (682) XXX-XXX.

For questions about your rights as a research participant or to discuss problems, complaints or concerns about a research study, or to obtain information, contact the ERAU Senior Grants Analyst,Teri Vigneau, Corsair Hall Room 203C,(386) 226-7179 or at hollerat@erau.edu.

VOLUNTARY NATURE OF STUDY

Taking part in this study is voluntary. You may choose not to take part or may withdraw consent and discontinue participation at any time without prejudice. Your decision whether or not to participate in this study will not affect your current or future relations with the principal investigator.

SUBJECT'S CONSENT

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. I have also read and fully understand the consent form and sign it freely and voluntarily.

In consideration of all of the above, I give my consent to participate in this research study.

I will print a copy of this informed consent document to keep for my records and mail the original to the investigator. My signature below indicates that I agree to take part in this study.

First Name:__________________Last Name:___________________Title:________

Organization/Agency Name:____

Tel:______________________Email:____________________________

Signature/Date:____

Principal Investigator Signature/Date:__________________________________

APPENDIX B

Data Collection Devices

Table B1

Pilot study data collection sheet

Date: Total Time to Complete:

Rater Name:				√ Check One Box Only. Indicate Grid Reason Code (e.g. E3)		Comments	True Rating
Event #	Airport ID	Date of Event	Ownship Model & Call Sign	SURF-IA Alert	SURF-IA No Alert		Alert or No Alert
Example	*XYZ*	*Date*	*C172* *N333RJ*		*Z1*	*Additional comments on reason for rating*	*Alert*
1	ABE	19-Sep-08	CRJ700 AS7138				Alert
2	SAN	16-Jan-08	B737 SW1626				Alert
3	FAT	28-Aug-08	CRJ200 SKW69R				Alert
4	CLT	29-May-09	CRJ200 JIA390				Alert
5	ATW	24-Jul-11	EMB145 BTA6131				Alert
6	SFO	26-May-07	ERJ170 RPA4912				Alert
7	TUS	2-Jun-06	F-16 Banshee1				Alert
8	HNL	14-Aug-09	B767 HAL9				Alert
9	UGN	24-Jul-04	C172 N405ES				Alert

Table B2

Data collection sheet

Date: Total Time to Complete:

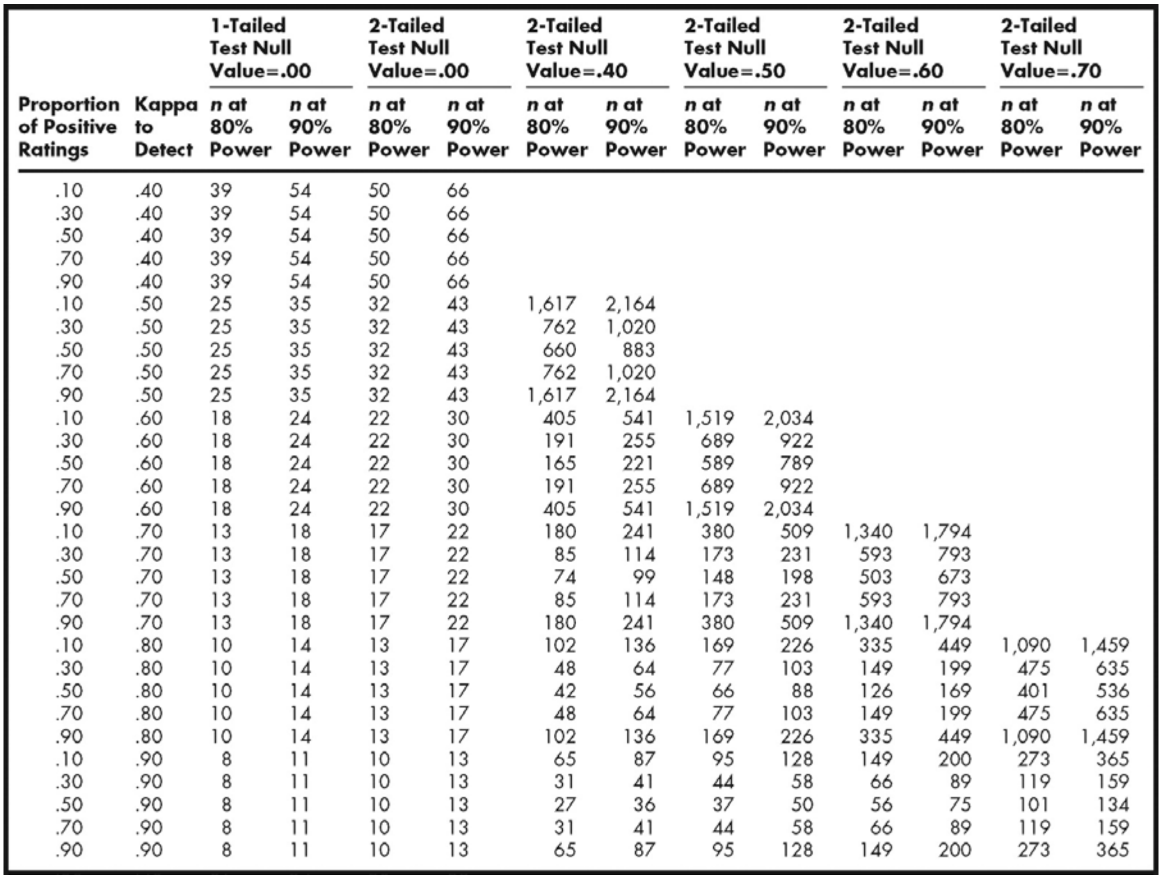
Rater Name:				√ Check One Box Only. Indicate Grid Reason Code (e.g. E3)		Comments
Event #	Airport ID	Date of Event	Ownship Model & Call Sign	SURF-IA Alert	SURF-IA No Alert	
Example	*XYZ*	*Date*	*C172*	*E3*		*Additional comments on reason for rating*
1	CHS	18-Dec-09	CRJ200 AS5510			
2	DAB	24-Nov-07	C182 N2438F			
3	DEN	5-Jan-07	SW4 LYM4216			
4	DVT	18-Jan-09	PA28 Trans922			
5	GFK	4-Jun-12	C210 N777JK			
6	HWO	29-Feb-12	C172 N64238			
7	HPN	4-Dec-11	B407 N408TD			
8	MKE	11-Mar-12	C750 FIV702			
9	JFK	6-Jul-05	B767 ISRAIR 102			

APPENDIX C

Tables

Table C1

Number of subjects required in a 2-rater study to detect a statistically significant k (p ≤ .05).



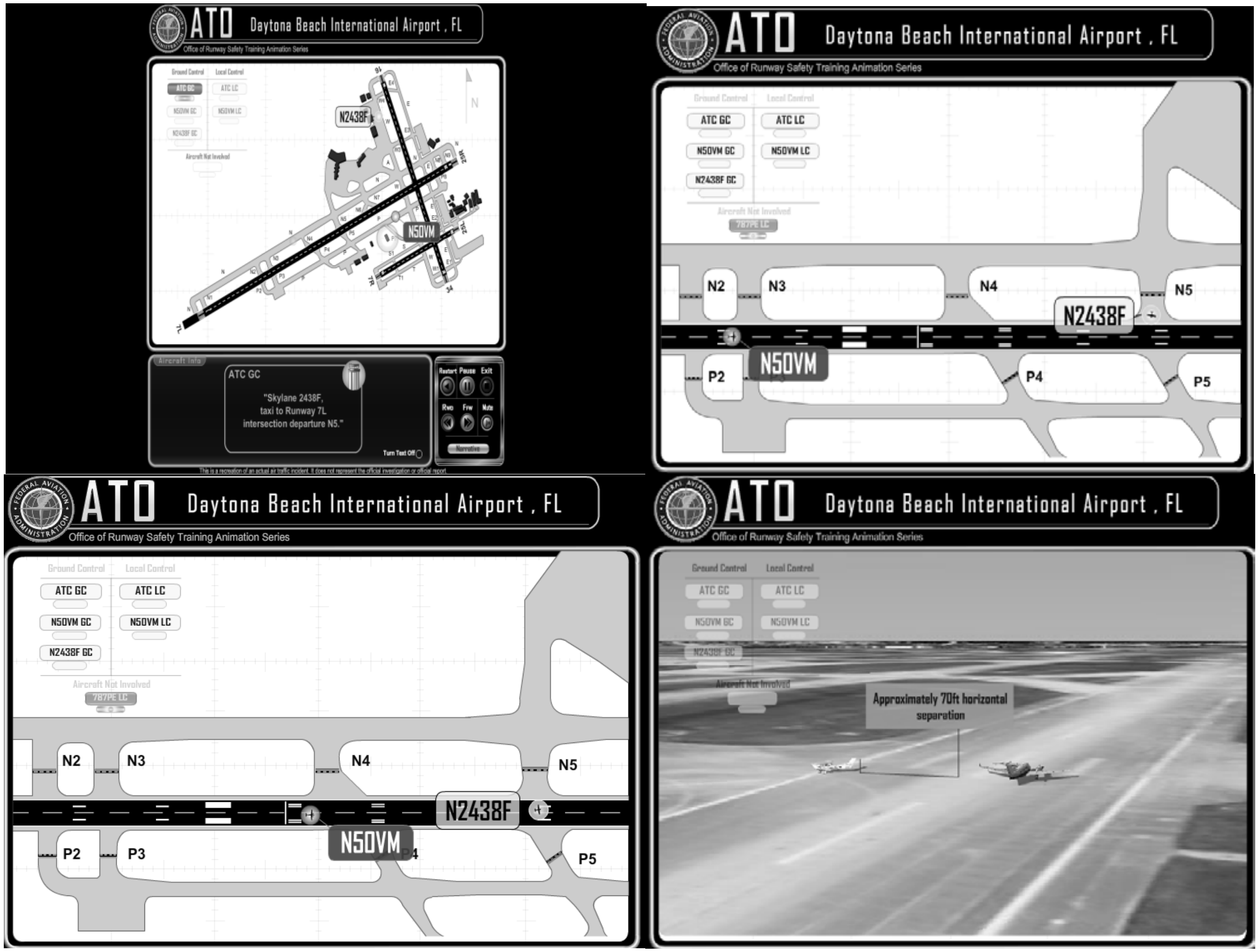
		1-Tailed Test Null Value=.00		**2-Tailed Test Null Value=.00**		**2-Tailed Test Null Value=.40**		**2-Tailed Test Null Value=.50**		**2-Tailed Test Null Value=.60**		**2-Tailed Test Null Value=.70**	
Proportion of Positive Ratings	**Kappa to Detect**	***n* at 80% Power**	***n* at 90% Power**	***n* at 80% Power**	***n* at 90% Power**	***n* at 80% Power**	***n* at 90% Power**	***n* at 80% Power**	***n* at 90% Power**	***n* at 80% Power**	***n* at 90% Power**	***n* at 80% Power**	***n* at 90% Power**
.10	.40	39	54	50	66								
.30	.40	39	54	50	66								
.50	.40	39	54	50	66								
.70	.40	39	54	50	66								
.90	.40	39	54	50	66								
.10	.50	25	35	32	43	1,617	2,164						
.30	.50	25	35	32	43	762	1,020						
.50	.50	25	35	32	43	660	883						
.70	.50	25	35	32	43	762	1,020						
.90	.50	25	35	32	43	1,617	2,164						
.10	.60	18	24	22	30	405	541	1,519	2,034				
.30	.60	18	24	22	30	191	255	689	922				
.50	.60	18	24	22	30	165	221	589	789				
.70	.60	18	24	22	30	191	255	689	922				
.90	.60	18	24	22	30	405	541	1,519	2,034				
.10	.70	13	18	17	22	180	241	380	509	1,340	1,794		
.30	.70	13	18	17	22	85	114	173	231	593	793		
.50	.70	13	18	17	22	74	99	148	198	503	673		
.70	.70	13	18	17	22	85	114	173	231	593	793		
.90	.70	13	18	17	22	180	241	380	509	1,340	1,794		
.10	.80	10	14	13	17	102	136	169	226	335	449	1,090	1,459
.30	.80	10	14	13	17	48	64	77	103	149	199	475	635
.50	.80	10	14	13	17	42	56	66	88	126	169	401	536
.70	.80	10	14	13	17	48	64	77	103	149	199	475	635
.90	.80	10	14	13	17	102	136	169	226	335	449	1,090	1,459
.10	.90	8	11	10	13	65	87	95	128	149	200	273	365
.30	.90	8	11	10	13	31	41	44	58	66	89	119	159
.50	.90	8	11	10	13	27	36	37	50	56	75	101	134
.70	.90	8	11	10	13	31	41	44	58	66	89	119	159
.90	.90	8	11	10	13	65	87	95	128	149	200	273	365

Note: Adapted from Sim, J., & Wright, C. C. (2005). “The kappa statistic in reliability studies: Use, interpretation, and sample size requirements,” by J. Sim and C.C. Wright, 2005, *Journal of the American Physical Therapy Association*, 85(3), 257-268.

Table C2

Pilot study rater reason code equivalencies

Event # Airport ID	Rater A Reason Code Description	Rater B Reason Code Description	Explanation of equivalency
1 ABE	F12-Traffic taxiing on centerline of same runway	F10-Traffic after landing rollout on same runway	The aircraft state for an aircraft that has just landed would be considered landing rollout until the groundspeed is ≤ 40 knots at which time the aircraft state becomes taxiing on-runway (lined up).
3 FAT	G13- Traffic after landing rollout on same runway	G10- Traffic taxiing on centerline of same runway	The aircraft state for an aircraft that has just landed would be considered landing rollout until the groundspeed is ≤ 40 knots at which time the aircraft state becomes taxiing on-runway (lined up).
5 ATW	A1- Traffic on Approach to same Runway as Ownship, with ≤ 35 secs to runway threshold	A2- Traffic on Approach to same Runway as Ownship, with ≤ 15 secs to runway threshold	Temporal assessment of the aircraft time to cross the runway threshold (RTHRE) during an approach; warning alert if ≤ 15 seconds to RTHRE, or a caution alert if ≤ 35 seconds to RTHRE.
6 SFO	C4-Ownship landing and Traffic taking off from intersecting runways	F16-Ownship taking off and Traffic landing from intersecting runways	An examination of the SURF-IA logic diagrams derived from RTCA (2010) DO-323 indicated that the aircraft state codes were redundant, hence interchangeable.
9 UGN	D4-Traffic and Ownship taking off from intersecting runways	F18-Ownship and Traffic taking off on intersecting runways	An examination of the SURF-IA logic diagrams derived from RTCA (2010) DO-323 indicated that the aircraft state codes were redundant, hence interchangeable.



ATC GC

"Skylane 2438F, taxi to Runway 7L intersection departure N5."

This is a recreation of an actual air traffic incident. It does not represent the official investigation or official report.

Figure D1. Example screen shots from runway incursion video reenactment data. Adapted from FAA RI Database http://www.faa.gov/airports/runway_safety/videos/

ASIAS BRIEF REPORT

GENERAL INFORMATION

Data Source:	RUNWWAY SAFETY OFFICE - RUNWAY INCURSIONS
Event Id:	5826
FAA Event Type:	PD
Event LCL Date:	24-NOV-07
Event LCL Time:	1556
Event State:	FL
RI Category Rank:	A
Airport Id:	DAB
Event Location:	DAYTONA BEACH INTL, FL
Event Lndg/Tkoff Surface:	RWY 7L
Aircraft 1 Type:	C182
Aircraft 2 Type:	BE20
Aircraft 1 FAR:	91
Aircraft 2 FAR:	91
Weather Condition:	9 SM SCT 026TCU BKN 055 BKN 100 CALM

ANALYSIS DESCRIPTION

A CESSNA C182 WAS TAXIED TO RUNWAY 7L INTERSECTION N5 FOR DEPARTURE. THE C182 THEN CROSSED 7L AT N5 WITHOUT CLEARANCE AND CONFLICTED WITH A BEECH BE20 ON DEPARTURE ROLL 7L FROM ABEAM P2, APPROXIMATELY 3,000 FEET WEST OF N5. ATCT ADVISED THE C182 OF THE IMPROPER CROSSING AND THE PILOT MADE A 180 TO EXIT. THE BE20 CONTINUED DEPARTURE ROLL AND AS IT PASSED N5 AND THE C182 WAS JUST ABOUT OFF THE RUNWAY. THE BE20 STATED HE WAS COMMITTED TO DEPART WHEN HE SAW THE C182. THE C182 PILOT SAID HIS BRAKES STUCK. CLOSEST HORIZONTAL PROXIMITY REPORTED WAS 70 FEET.

END REPORT

Figure D2. ASIAS brief report example.

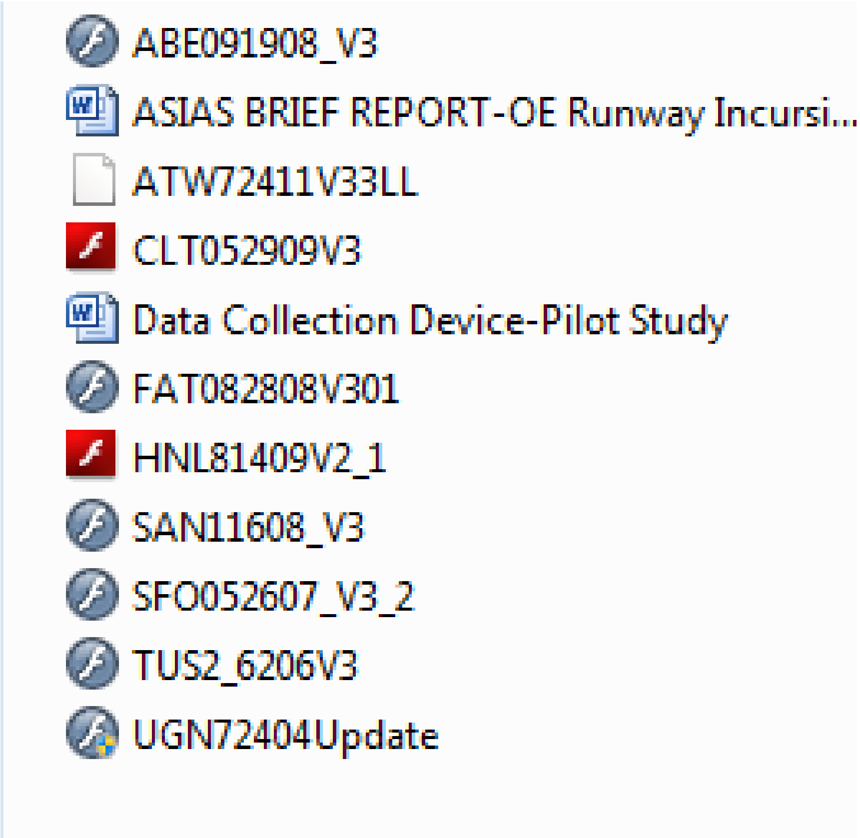
APPENDIX E

Instructions to Raters

INSTRUCTIONS TO RATERS

The following files are included on the attached flash drive under the folder name “Rater Package-Pilot Study”. The “ASIAS Brief Report” file is provided to supplement the video reenactments. After observing each video reenactment and reading the associated ASIAS report, please rate each event as either alerting or non-alerting following the SURF-IA logic diagrams, starting with Figure 1. The SURF-IA logic diagrams and a data collection sheet are provided under the files named “SURF-IA Logic Diagrams”. and “Data Collection Device” respectively. A sample data entry is provided on the data collection sheet. Please do not open the folder named “Rater Package-Final Study” until notified by the researcher. All of the other procedures on the Consent Form apply unless otherwise stated.

Thank you for your participation!



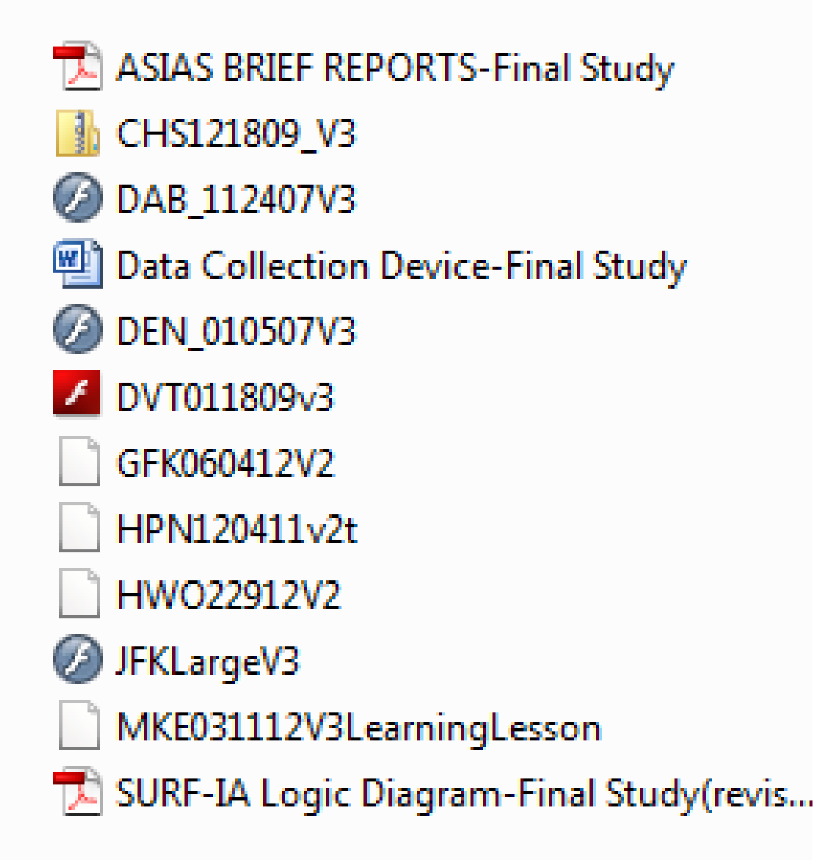
ABE091908_V3
ASIAS BRIEF REPORT-OE Runway Incursi...
ATW72411V33LL
CLT052909V3
Data Collection Device-Pilot Study
FAT082808V301
HNL81409V2_1
SAN11608_V3
SFO052607_V3_2
TUS2_6206V3
UGN72404Update

Figure E1. Instructions to raters (pilot study).

INSTRUCTIONS TO RATERS

The following files are included on the attached flash drive under the folder name "Rater Package-Final Study". The "ASIAS Brief Report" file is provided to supplement the video reenactments. After observing each video reenactment and reading the associated ASIAS report, please rate each event as either alerting (with reason code) or non-alerting following the SURF-IA logic diagrams, starting with Figure 1. The SURF-IA logic diagrams and a data collection sheet are provided under the files named "SURF-IA Logic Diagrams" and "Data Collection Device" respectively. A sample data entry is provided on the data collection sheet. For this study the aircraft state for an aircraft that has just landed should be considered *landing rollout* until the groundspeed is ≤ 40 knots at which time the aircraft states *becomes taxiing on runway (lined up)*. Also a typographical error was corrected in the last column header for Ownship in the SURF-IA Logic Diagram (Figure 3) to read *lined-up* instead of *not lined-up*. All of the other procedures on the Consent Form apply unless otherwise stated.

Thank you for your participation!



ASIAS BRIEF REPORTS-Final Study
CHS121809_V3
DAB_112407V3
Data Collection Device-Final Study
DEN_010507V3
DVT011809v3
GFK060412V2
HPN120411v2t
HWO22912V2
JFKLargeV3
MKE031112V3LearningLesson
SURF-IA Logic Diagram-Final Study(revis...

Figure E2. Instructions to raters.

APPENDIX F

Rater Data

Table F1

Pilot study rater data

Data Collection Date(s) January 25, 2013 –February 3, 2013				Alert/No Alert Reason Code		True Score	Comments
Event #	Airport ID	Date of Event	Ownship Model & Call Sign	Rater A	Rater B		Rater A and Rater B
1	ABE	19-Sep-08	CRJ700 AS7138	Alert F12	Alert F10	Alert	No comments from raters
2	SAN	16-Jan-08	B737 SW1626	Alert F12	Alert F12	Alert	Rater B- Conflict aircraft was stopped prior to B737 initiating take-off roll
3	FAT	28-Aug-08	CRJ200 SKW69R	Alert G13	Alert G10/H12	Alert	Rater B- Assumed that the landing state is defined as > 40 knots, per example in RTCA/DO-323, page A-16, therefore H12.
4	CLT	29-May-09	CRJ200 JIA390	Alert F1	Alert F1	Alert	Rater A-Cannot determine GS of JIA390 when PC12 enters runway. If GS was >80 kts, no alert would be issued
5	ATW	24-Jul-11	EMB145 BTA6131	Alert A1	Alert A2/B3/C3	Alert	Rater B-EMB145 would have triggered these three alert cells consecutively
6	SFO	26-May-07	ERJ170 RPA4912	Alert C4	Alert F16	Alert	No comments from raters
7	TUS	2-Jun-06	F-16 Banshee1	Alert E3	Alert E3	Alert	No comments from raters
8	HNL	14-Aug-09	B767 HAL9	Alert G13	Alert G13/G12/H12	Alert	Rater B-B767 would receive alert while on approach (G13, then G12) and also after touch down (H12)
9	UGN	24-Jul-04	C172 N405ES	Alert D4	Alert F18	Alert	No comments from raters

Table F2

Rater data

Data Collection Date(s) February 12, 2013 –February 15, 2013				Alert/No Alert Reason Code		Comments
Event #	Arpt ID	Date of Event ASIAS ID	Ownship Model & Call Sign	Rater A	Rater B	Rater A and Rater B
1	CHS	18-Dec-09 8173	CRJ200 AS5510	Alert F1	Alert F1	No comments from raters
2	DAB	24-Nov-07 5826	C182 N2438F	Alert E3	Alert E3	No comments from raters
3	DEN	5-Jan-07 4828	SW4 LYM4216	Alert I7,I6	Alert I7, I6	Rater A-Alert first due to I7, then later due to I6 Rater B-I7 followed by I6
4	DVT	18-Jan-09 7167	PA28 Trans922	Alert E3	Alert E3	Rater A-This was close to no alert, looked as if Trans415 lifted off as 922 enters runway
5	GFK	4-Jun-12 11322	C210 N777JK	Alert G13/G12	Alert G12	Rater A- Hard to tell when N777JK within SURF IA approach corridor, if not in corridor until <15 sec to threshold then G12
6	HWO	29-Feb-12 10923	C172 N64238	No Alert Z4	No Alert (1)	Rater A-No alert because departing aircraft lifted off before N64238 crosses hold line (1)Rater B- A basic SURF-IA implementation per DO-323 minimum would not identify this part of the crossing taxiway as part of the runway because it is located beyond the runway threshold.
7	HPN	4-Dec-11 10675	B407 N408TD	No Alert (2)	No Alert (3)	(2)Rater A- E10 based in information given. Based on DO-323, there would be no alert on helicopter because SURF IA not approved for operation on helicopters. (3) Rater B- A basic SURF-IA implementation per DO-323 minimum would not have been installed on helicopters as DO-323 did not include safety and performance requirements for helicopter installations. However, an advanced SURF-IA installation could have been installed on helicopters. In that case, the alert would have triggered cell E10.
8	MKE	11-Mar-12 10969	C750 FIV702	No Alert Z1	No Alert Z1	No comments from raters
9	JFK	6-Jul-05 3374	B767 ISRAIR 102	Alert	Alert	No comments from raters

APPENDIX G

Follow-up Questionnaire for Raters

Follow-up Questionnaire for Raters

Rater Name:		Date:		Time to Complete:	

Please provide a written response to each of the questions below by typing in your response immediately below each question and return via email to joslinr@my.erau.edu. Your Consent Form will extend to this Questionnaire, unless otherwise requested.

1. In Event 6 with ASIAS ID 10923 you rated the incident as *non-alerting* because the traffic aircraft was not in an "on-runway" state, based on the SURF-IA model definition, until after the aircraft on take-off from the runway had lifted off. (ASIAS ID 10923 report and video reenactment file are attached for your reference)

   (a) How would you have rated the incident using the RISC model definition of "on-runway" as depicted in Figure 1 below? Please explain why.
   RESPONSE:

   (b) How would you define an "off-runway" condition for an aircraft clearing the running on to a taxiway?
   RESPONSE:

2. What, if any, lessons learned did you gain while rating the incidents?
   RESPONSE:

3. What other applications, aviation or otherwise, might benefit from using expert raters to rate the outcomes from new technology?
   RESPONSE:

4. What other applications, aviation or otherwise, might benefit from validating their legacy metrics for assessing the benefits or performance of new technology?
   RESPONSE:

5. What scenarios created the most doubt in your response?
   RESPONSE:

6. What recommendations would you make for methodologies of future studies seeking similar expert ratings?
   RESPONSE:

7. What recommendations do you have for modifying one or both of the model(s) (RISC and/or SURF-IA) to harmonize the outcomes?
   RESPONSE:

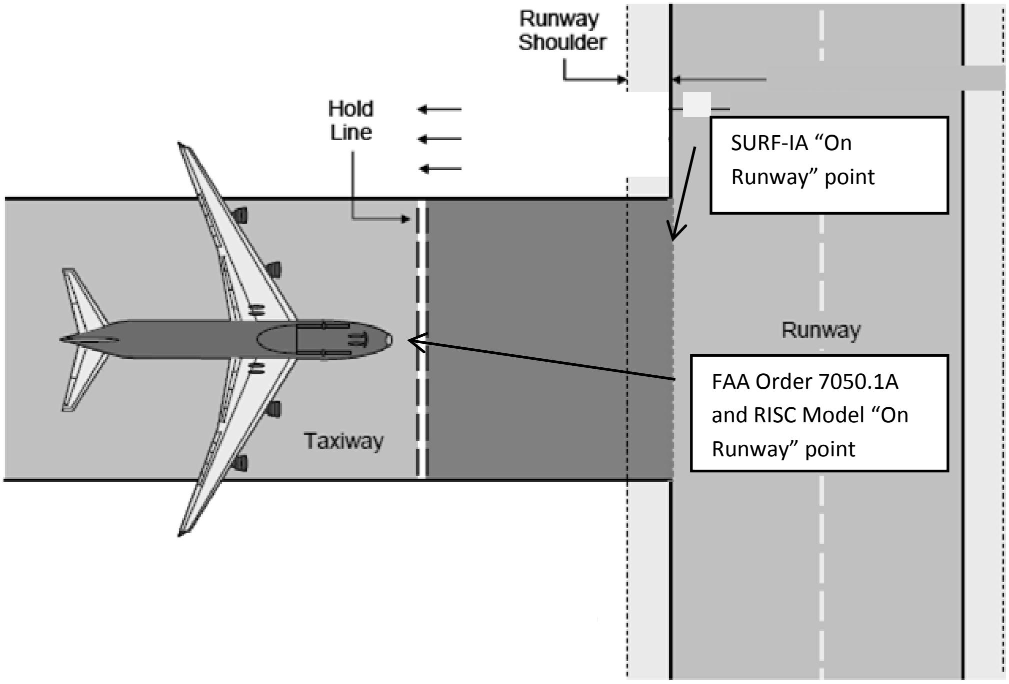


Figure 1. U.S. Airport Surface Geometry On-runway points. Adapted from “Safety, performance and interoperability requirements document for enhanced traffic situational awareness on the airport surface with indications and alerts (SURF-IA),” by RTCA, 2010, “Runway Safety Program,” by FAA, 2010.“Airport Design,” by FAA, 1989.

THANK YOU FOR YOUR CONTRIBUTION TO THIS STUDY

References:

Federal Aviation Administration [FAA] (1989). *Airport design* (FAA Advisory Circular 150/5300-13). Retrieved from http://faa.rgl.gov/

Federal Aviation Administration [FAA] (2010a). *Runway safety program* (FAA Order 7050.1A). Retrieved from http://rgl.faa.gov/