NASA/TM-2016-219357

# Modeling Relationships Between Flight Crew Demographics and Perceptions of Interval Management

Benjamin Remy
*Florida Institute of Technology, Melbourne, Florida*

Sara R. Wilson
*Langley Research Center, Hampton, Virginia*

December 2016

# NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM-2016-219357

# Modeling Relationships Between Flight Crew Demographics and Perceptions of Interval Management

Benjamin Remy
*Florida Institute of Technology, Melbourne, Florida*

Sara R. Wilson
*Langley Research Center, Hampton, Virginia*

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

December 2016

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA  23681-2199
Fax: 757-864-6500

# Table of Contents

## List of Figures



## List of Tables

# Abbreviations and Acronyms

| | |
|---|---|
| AAL | American Airlines |
| ADS-B | Automatic Dependent Surveillance - Broadcast |
| ASTOR | Aircraft Simulation for Traffic Operations Research |
| ATD-1 | Air Traffic Management Technology Demonstration - 1 |
| ATOL | Air Traffic Operations Laboratory |
| CGD | Configurable Graphics Display |
| DAL | Delta Air Lines |
| DTS | Development and Test Simulator |
| EFB | Electronic Flight Bag |
| FDX | FedEx Express |
| IFD | Integration Flight Deck simulator |
| IM | Interval Management |
| IMAC | Interval Management Alternative Clearances experiment |
| KDEN | Denver International Airport |
| MCH | Modified Cooper-Harper workload scale |
| PTP | Planned Termination Point |
| RNAV | Area Navigation |
| RNP | Required Navigation Performance |
| SCX | Sun Country Airlines |
| SD | Standard Deviation |
| STAR | Standard Terminal Arrival Route |
| SWA | Southwest Airlines |
| UAL | United Airlines |

# Abstract

The Interval Management Alternative Clearances (IMAC) human-in-the-loop simulation experiment was conducted to assess interval management system performance and participants' acceptability and workload while performing three interval management clearance types. Twenty-four subject pilots and eight subject controllers flew ten high-density arrival scenarios into Denver International Airport during two weeks of data collection. This analysis examined the possible relationships between subject pilot demographics on reported perceptions of interval management in IMAC. Multiple linear regression models were created with a new software tool to predict subject pilot questionnaire item responses from demographic information. General patterns were noted across models that may indicate flight crew demographics influence perceptions of interval management.

# 1 Introduction

The Interval Management Alternative Clearances (IMAC) experiment was conducted at NASA Langley Research Center in July-August 2015 as part of the Air Traffic Management Technology Demonstration – 1 (ATD-1) sub-project [1]. The ATD-1 Concept of Operations integrates arrival scheduling, controller decision support tools, and flight deck avionics and procedures with the goal of improving efficiency and capacity in busy terminal airspace, primarily through speed control on optimized profile descents [2].

Interval management (IM) is the flight-deck based technology in ATD-1. The goal of IM is to improve airport throughput and arrival efficiency by increasing the precision of spacing between aircraft. IM encompasses the flight deck automation and crew procedures that allow a controller to issue a strategic clearance with a time or distance spacing interval based behind the preceding aircraft. The IM automation uses Ownship information (position, routing, etc.), IM clearance information (Target, spacing interval), and Target aircraft state data transmitted through Automatic Dependent Surveillance-Broadcast (ADS-B) to calculate an appropriate airspeed to achieve or maintain the assigned spacing goal.

## 1.1 Interval Management Alternative Clearances Experiment

The objective of the IMAC experiment was to assess IM system performance and participants' acceptability of three IM operations in a dynamic, high traffic density human-in-the-loop simulation. Flight crew acceptability and workload were a focus of the experiment to identify possible issues. The following three IM clearance types were investigated:

- **CROSS**

  When conducting a CROSS operation, the IM system will command airspeeds to achieve the spacing goal by the achieve-by point then maintain the spacing goal to the Planned Termination Point (PTP). The CROSS clearance can be used when aircraft are on separate, converging routes.

- **CAPTURE**

  When conducting a CAPTURE operation, the IM system will command airspeeds to achieve the spacing goal then maintain that spacing to the PTP.

- **MAINTAIN**

  When conducting a MAINTAIN operation, the IM system will command airspeeds to maintain the current spacing interval between the Ownship aircraft and the Target aircraft.

### 1.1.1 Simulation Environment

The simulated environment in the experiment was the airspace surrounding the Denver International Airport (KDEN). KDEN was chosen because all published Standard Terminal

Arrival Routes (STARs) into the airport connect directly to an instrument approach, a requirement for IM operations. North-flow and south-flow operations were simulated to reduce learning effects by controllers and pilots and to have a wider range of scenarios in the experiment. North-flow operations used runways 35L and 35R. South-flow operations used runways 17L and 35R.

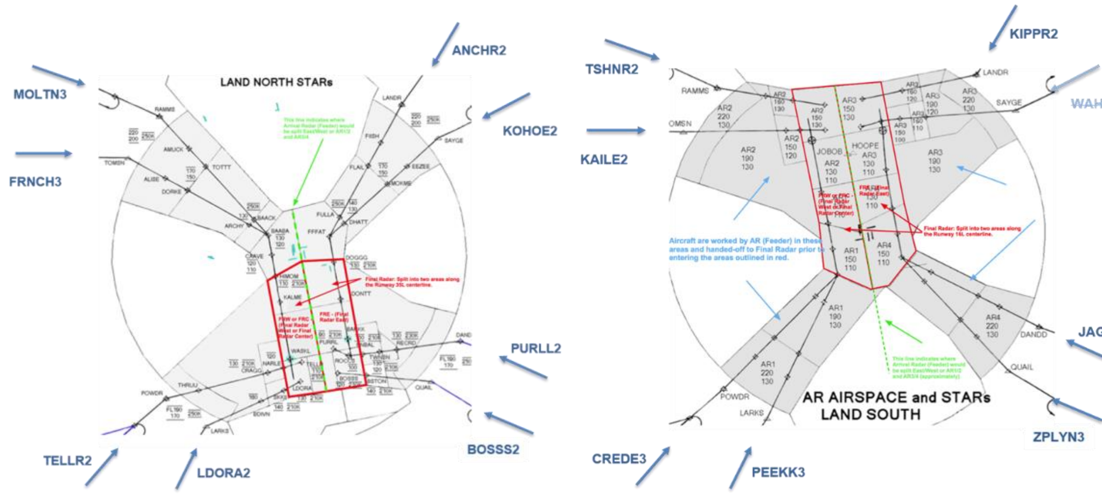


**Figure 1. North-flow (left) and south-flow (right) arrival routes into Denver.**

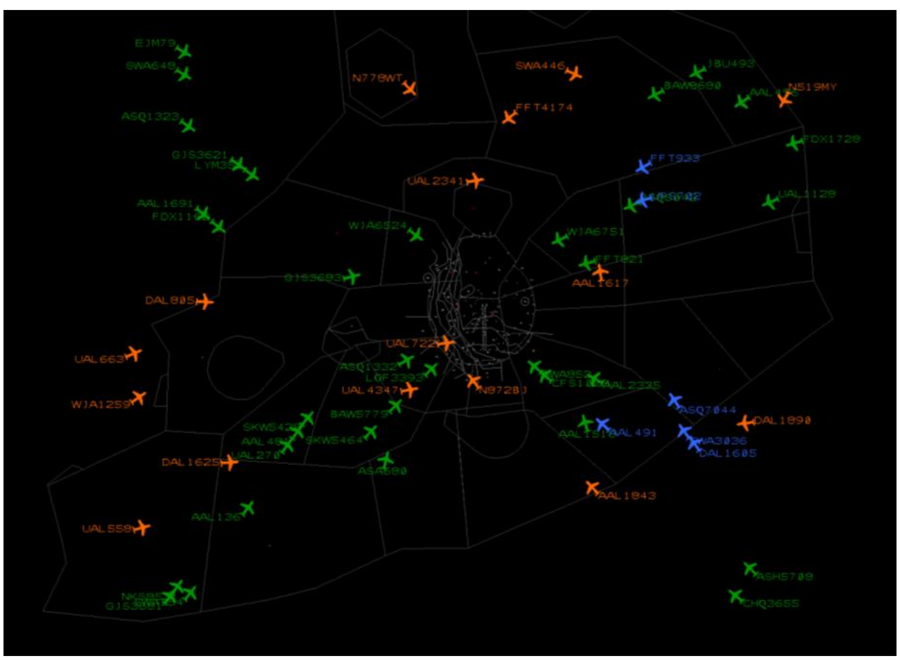

**Figure 2. Example of aircraft starting locations in Denver airspace.**

Each simulation scenario featured aircraft arriving on routes into Denver from all four directions (Figure 1). The traffic scenarios were based on real-world arrival rates, including the proportion of aircraft on each arrival route. Two types of human pilots flew simulated aircraft during the scenarios: research subject pilots (the source of the data) and confederate pilots to add realistic traffic to the arrival routes. Human pilots flew arrivals into either the east or west side of the airport, while approximately 35 computer-flown aircraft flew into the opposite side, to simulate a busy visual environment on traffic displays. Each scenario included six IM equipped aircraft flown by 12 subject pilots. Figure 2 gives an example of aircraft positioning at the start of a scenario. IM-equipped aircraft are shown in blue. All green colored aircraft on the same side of the airport are flown by confederate pilots, and those on the opposite side are computer-flown. The orange colored aircraft are aircraft flying over the area for additional visual complexity. The IM equipped aircraft began each scenario at a point prior to top-of descent.

Wind was not a variable in this experiment. One truth and one forecast wind field was used throughout. The wind field was designed from a one-year wind history to emulate realistic conditions at Denver and allow arrivals in both directions.

### 1.1.2 Experiment Design

The independent variable in the experiment was the type of IM operation: BASELINE (no IM operations), CAPTURE, CROSS, MAINTAIN, or MIXED (in which controllers could issue CAPTURE, CROSS, and MAINTAIN clearances based on their preference and judgment). Six categories of dependent variables were assessed in the experiment:

- IM algorithm performance,
- air traffic system performance,
- controller objective performance,
- flight crew objective performance,
- controller subjective assessments, and
- flight crew subjective assessments.

Each scenario had two replicates (see Table 1). Scenarios were conducted once with the Captain as the pilot flying (PF) and the First Officer as the pilot monitoring (PM), and once with the opposite roles. A within-subject design was used, and each flight crew flew ten scenarios. The run order for the BASELINE, CAPTURE, CROSS, MAINTAIN, and MIXED scenarios was partially counterbalanced using a Latin square design [3]. For additional details, see [1], [4], and [5].

**Table 1. IMAC Experiment Test Matrix**

| **BASELINE** | **CAPTURE** | **CROSS** | **MAINTAIN** | **MIXED** |
|---|---|---|---|---|
| *Replicate 1*<br>*Replicate 2* | *Replicate 1*<br>*Replicate 2* | *Replicate 1*<br>*Replicate 2* | *Replicate 1*<br>*Replicate 2* | *Replicate 1*<br>*Replicate 2* |

### 1.1.3 Participants

Two weeks of data collection took place. Each week, the participants included two confederate controllers, four subject controllers, six confederate pilots, and twelve subject pilots. A total of eight subject controllers and twenty-four subject pilots participated in the experiment. Section 1.2 contains information about subject pilot demographics.

### 1.1.4 Facilities

Air traffic controllers in the experiment operated from the Air Traffic Operations Laboratory (ATOL). Controller stations in the ATOL were configured to act as en route, terminal, and tower control facilities. Subject controllers acted in en route and terminal positions.

Subject flight crews conducted IM operations using one of three flight simulators: the Integration Flight Deck (IFD, Figure 3), the Development and Test Simulator (DTS, Figure 4) or the Aircraft Simulator for Traffic Operations Research (ASTOR, Figure **5**). The IFD and DTS are full scale flight deck simulators. The IFD is configured similarly to a Boeing 737-800 flight deck. The DTS flight deck is representative of a large twin-engine transport category aircraft. The ASTOR simulators are three-monitor desktop pilot stations configured for two-crew operations in this experiment. ASTOR pilots interact with simulated aircraft systems through the touch screen interface or with a mouse and keyboard.

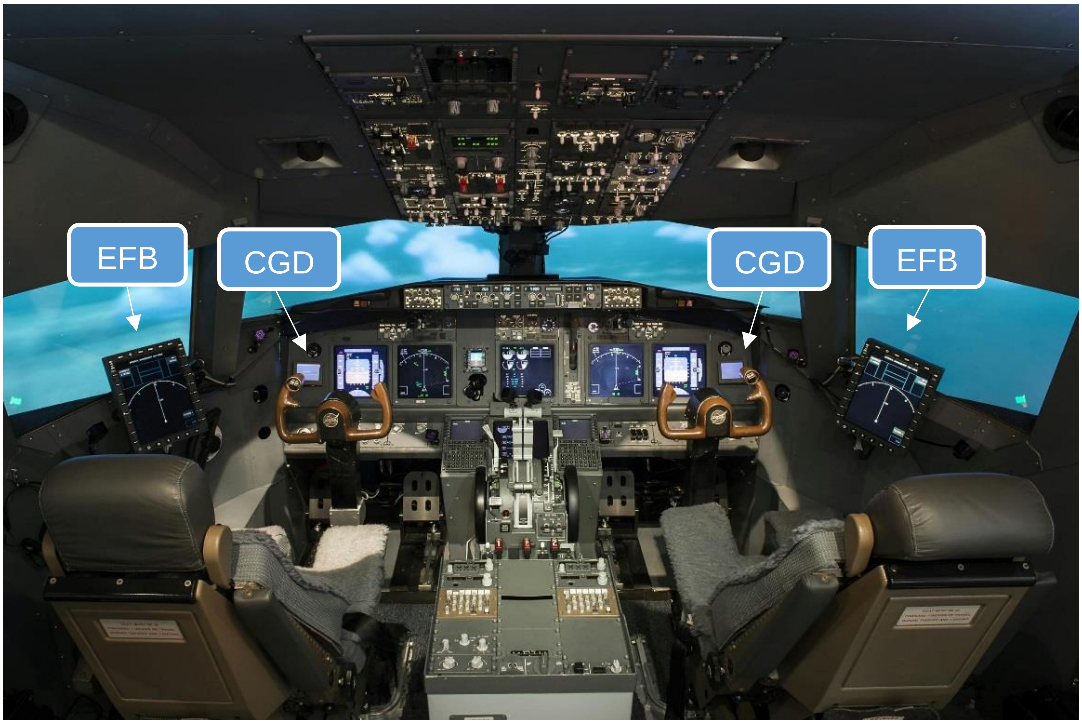


**Figure 3. Integration Flight Deck (IFD).**


**Figure 4. Development and Test Simulator (DTS).**


**Figure** 5**. Two-crew ASTOR station.**

### 1.1.5 IM Avionics Interfaces

Pilots of IM equipped aircraft interacted with the IM automation through the electronic flight bag (EFB) and configurable graphics display (CGD), seen in Figure 6. Ownship and IM clearance information was entered on the EFB which provided IM speed guidance and visual indications of IM state, as well as a display of other aircraft in reference to Ownship. The CGD provided a repeated display of IM speed commands and other visual indications of IM state in the forward field of view. These displays can be seen in context in Figure 3, Figure 4, and Figure **5**.


**Figure 6. IM EFB (left) and CGD (right).**

## 1.2 Flight Crew Demographic Data

In a pre-experiment questionnaire, data were collected on each participating pilot's age, airline affiliation, flight experience in years and in hours, types of aircraft flown, date most recently flown, qualifications like instructor or standards captain, experience with area navigation (RNAV) and required navigation performance (RNP) operations, and experience with previous IM experiments.

The pilot participants had a mean age of 57.5 years (SD = 7.1). Twelve of the 24 pilots had most experience with United Airlines, four pilots with Delta Air Lines, three pilots with American Airlines, two pilots with Southwest Airlines, two pilots with FedEx, and one pilot with Sun Country Airlines. The mean flight experience in years of all pilots was 34.4 (SD = 9.0). The mean commercial, multi-engine experience of all pilots was 17382 hours (SD = 6346). Pilot participants reported experience in a variety of aircraft, listing between three and 21 unique aircraft. All pilots listed experience in a Boeing 737 or 757, and most had experience in a variety of contemporary airliners. Eleven of 24 pilots listed experience in military aircraft. The reported date a pilot last flew a commercial aircraft was recoded as the number of months the pilot had flown prior to the experiment. Eighteen of 24 pilots had flown a commercial aircraft within two months before the experiment. Three pilots had not flown a commercial aircraft in the last four years. Twelve of 24 pilots had participated in a previous IM experiment. The type of simulator flown by a pilot in the experiment was also recorded.

From these biographic and administrative data, the following independent variables were extracted for analysis against flight crew subjective assessments:

1. Age (years)
2. Airline
3. Length of experience (years)
4. Flight time (hours)
5. Time since last flown (months)
6. Simulator type (ASTOR or IFD/DTS)
7. IM experience (Y/N)
8. Military flying experience (Y/N)

Two demographic variables (aircraft type and RNAV experience) were excluded from this analysis because there were no operationally significant differences among responses. All participants had experience in modern Boeing aircraft (737 through 787), and all participants had experience with RNAV operations. Experience with RNP was also excluded because responses could not be effectively categorized. Military experience was collected from a question which asked pilots what types of aircraft they flew; pilots were thorough in the aircraft they reported, and pilots who reported experience in military aircraft were assumed to have military experience.

A number of these demographic variables were categorical. Variables were coded for regression as zero for "No"/category absent and one for "Yes"/category present. The airline affiliation variable was split by airline, with the baseline as United Airlines. The simulator type variable was split into the categories ASTOR and IFD/DTS, with ASTOR as zero and IFD/DTS as one. The IFD and DTS crews were grouped together and compared against ASTOR crews in this analysis because the IFD and DTS are both full flight decks.

## 1.3 Post-Run and Post-Experiment Questionnaires

Flight crew subjective assessments were collected with two separate questionnaires, a post-run questionnaire and post-experiment questionnaire. The post-run questionnaire was administered to each pilot immediately after the last training run and after each experimental run. The post-experiment questionnaire was administered after the final post-run questionnaire. The primary goal of the post-run and post-experiment flight crew questionnaires was to assess participating pilots' subjective acceptability of IM operations and the cockpit workload associated with IM operations. The questionnaires also assessed concerns like the usefulness and intuitiveness of IM cockpit displays and situational awareness. Apart from the workload item measured on the Modified Cooper-Harper rating scale [6] (MCH, measured 1-10), all questionnaire items were either free response or an assessment on Likert scales from 1 to 7. Only numeric responses were included in this analysis.

This report focuses on the relationships between flight crew subjective assessment responses and flight crew demographic data. Multiple linear regression models were created to assess and predict these relationships.

# 2 Methodology

To examine the relationships between flight crew demographics and perceptions of IM, flight crew questionnaire item responses were predicted from pilot demographic data with multiple linear regression modeling. This was accomplished in R [7] with an application developed for the task in the Shiny web application framework [8]. Before beginning regression analysis, however, relationships within flight crew demographic data were examined to check for collinearity. Models were then created with the Selection of Linear Regression Models (SoLR-M) application using adjusted $R^2$ [9] and Mallows' $C_p$ statistics [10]. Adjusted $R^2$ represents the proportion of variance in the dependent variable that can be predicted by the independent variables, adjusted for the number of independent variables. In other words, Adjusted $R^2$ describes the ability of a model to explain the dependent variable. Mallows' $C_p$ describes the fit of the model, where a smaller value, close to the number of parameters, describes a better fit.

## 2.1 Relationships within Flight Crew Demographic Data

Relationships between IMAC flight crew demographic variables were investigated with Pearson product-moment correlations, one-way Analysis of Variance (ANOVA), and chi-square tests for independence [11]. See Table **2** for all results and $p$ values. Significant relationships ($p \leq .05$) in Table **2** are displayed in bold text, though note that a significant relationship does not necessarily imply a strong relationship. Age and years of experience were highly correlated. Flight time was moderately correlated with age and years of experience. These relationships were accounted for in the creation of regression models by avoiding models which contained both variables in a correlated pair. Future research could investigate other methods for addressing multi-collinearity, such as principal component analysis or ridge regression.

## 2.2 Selection of Linear Regression Models (SoLR-M) Application

The Selection of Linear Regression Models (SoLR-M) application was created for this analysis with the Shiny web application framework for R [8]. The application takes user data input as a .csv or .RData file and allows the user to select the appropriate independent and dependent variables, outputting a linear model and associated statistics. With its graphical interface, the user can quickly include or remove variables from a model while comparing coefficients and statistics, without repeatedly typing variable names and commands. SoLR-M also displays a plot of models found with an automated exhaustive search ordered by the desired statistic, which could be scaled by $R^2$, adjusted $R^2$, Mallows' $C_p$, or BIC (Bayesian information criterion). The graphical interface and model selection tools available in SoLR-M allowed a thorough selection of useful linear regression models more quickly and easily than when using the standard text-based R interface.

With this tool, multiple linear regression models were selected for each quantitative questionnaire item. To create the most useful predictive models of flight crew questionnaire responses, an appropriate balance was found between model statistics (adjusted $R^2$ and Mallows' $C_p$) and the subjective importance of each coefficient.

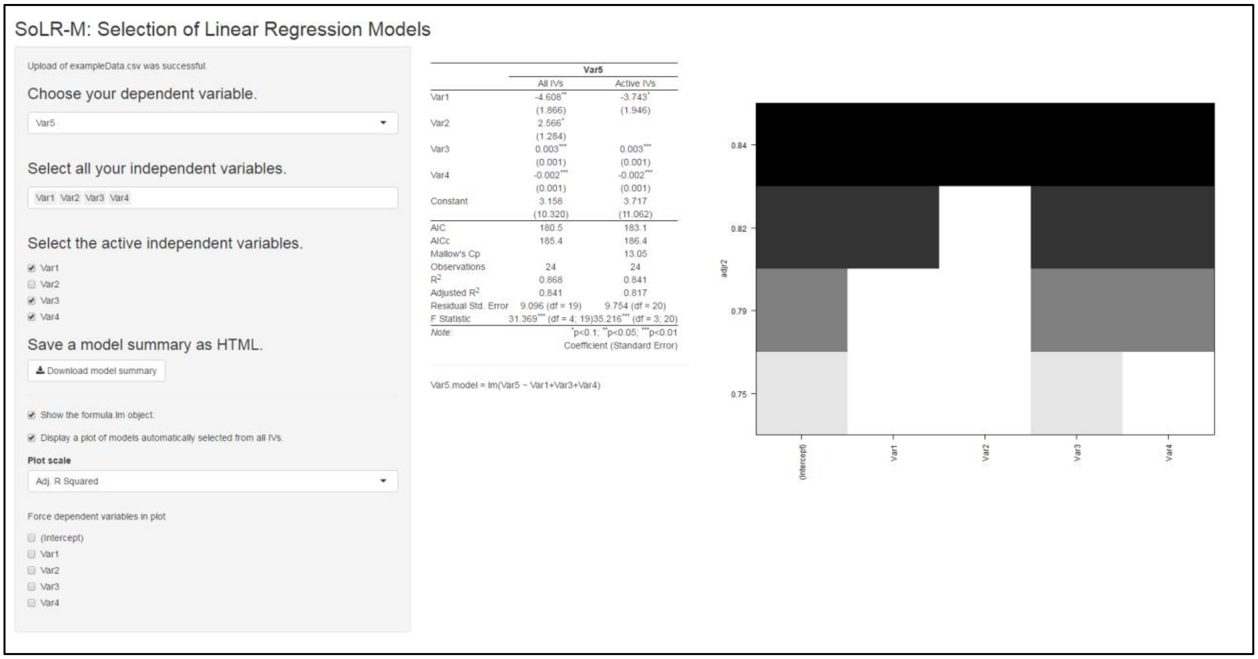


**Figure 7. Example of the Selection of Linear Regression Models (SoLR-M) application interface with generic data.**

## Table 2. Flight Crew Demographic Relationships.

| | Years of Experience | Flight Time | Months Since Last Flown | Airline | Military Experience | IM Experience | Simulator (by Type) |
|---|---|---|---|---|---|---|---|
| Current Age | $\boldsymbol{r = 0.842}$ **($\boldsymbol{p < .001}$)** | $\boldsymbol{r = 0.650}$ **($\boldsymbol{p = .001}$)** | $r = 0.298$ ($p = .158$) | $F(5,18) = 1.82$ ($p = .159$) | $F(1, 22) = 0.00$ ($p = .954$) | $\boldsymbol{F(1, 22) = 6.31}$ **($\boldsymbol{p = .02}$),** $R^2 = .22$ | $F(2, 21) = 0.43$ ($p = .653$) |
| Years of Experience | - | $\boldsymbol{r = 0.550}$ **($\boldsymbol{p = .005}$)** | $r = 0.089$ ($p = .678$) | $F(5, 18) = 1.86$ ($p = .152$) | $F(1, 22) = 0.01$ ($p = .926$) | $\boldsymbol{F(1, 22) = 7.22}$ **($\boldsymbol{p = .013}$),** $R^2 = .25$ | $F(2, 21) = 0.22$ ($p = .979$) |
| Flight Time | - | - | $r = 0.123$ ($p = .567$) | $F(5, 18) = 0.50$ ($p = .772$) | $F(1, 22) = 0.21$ ($p = .648$) | $F(1, 22) = 0.89$ ($p = .357$) | $F(2, 21) = 0.11$ ($p = .900$) |
| Months Since Last Flown | - | - | - | $F(5, 18) = 0.70$ ($p = .63$) | $F(1, 22) = 0.38$ ($p = .544$) | $F(1, 22) = 0.21$ ($p = .649$) | $F(2, 21) = 0.55$ ($p = .585$) |
| Airline | - | - | - | - | $\chi^2(5) = 1.51$ ($p = .912$) | $\chi^2(5) = 6.22$ ($p = .285$) | $\boldsymbol{\chi^2(10) = 30.63}$ **($\boldsymbol{p < .001}$)** $V = .79$ |
| Military Experience | - | - | - | - | - | $\chi^2(1) = 1.40$ ($p = .285$) | $\chi^2(2) = 0.08$ ($p = .959$) |
| IM Experience | - | - | - | - | - | - | $\chi^2(2) = 0$ ($p = 1$) |

# 3 Results

Most models of flight crew responses could not explain greater than half the variation in a questionnaire item (i.e. adjusted $R^2$ below 0.5), including the measures of MCH Workload and overall acceptability of IM operations. Some responses could be reasonably well predicted, however, including the usefulness of some EFB and CGD items. The presence or absence of a predictor across models was also noted, to search for trends indicating a broader relationship between flight crew characteristics and questionnaire responses.

It is important to note that the sample size used in this analysis for many pilot demographic categories is too small to be operationally conclusive. These results are intended to identify trends and areas for future research.

## 3.1 Post-Run Questionnaire Models

All models with post-run questionnaire results as the models' dependent variable resulted in adjusted $R^2$ values less than 0.38, and thus were not useful for prediction. Since these models are not considered useful, they are not included in this report. All of these models, however, were significant ($p < 0.05$), with many significant independent variables. While not useful for prediction, these models help the researchers understand the relative importance of demographics in the IMAC flight crew post-run questionnaire results. The flight crews' airline affiliation was present as a predictor in the most useful models of post-run questionnaire results. This may indicate differences in airline standard operating procedure have an influence on the acceptability and use of interval management.

Of the post-run models, the dependent variable best predicted was the acceptability of the heads-down time required of the pilot monitoring for IM clearance entry into the EFB. The model predicts this variable with airline affiliation, pilot flight time, prior IM experience, simulator type, and military experience and was significant ($p < 0.01$), with an adjusted $R^2$ of 0.38. Flight crew responses to this questionnaire item were generally favorable (Mean = 6.3, SD = 1.05). The model indicates United and Delta pilots found the heads-down time less acceptable than pilots from FedEx, Sun Country, and Southwest. Pilots flying the IFD and DTS rated the heads down time slightly less acceptable than pilots flying the ASTORs.

## 3.2 Post-Experiment Questionnaire Models

Two questionnaire items produced particularly useful models for predicting post-experiment questionnaire responses, with adjusted $R^2$ values above 0.7. These questions were about the usefulness of IM commanded speed on the CGD and the operational acceptability of the MAINTAIN clearance phraseology. Six post-experiment questionnaire item responses can be somewhat well predicted with models whose adjusted $R^2$ value fell between 0.5 and 0.7. These are the acceptability of responsibility for spacing, the intuitiveness of entering an IM clearance into the EFB, the usefulness of target route on the EFB, the usefulness of the merge point/waypoint on the EFB, the usefulness of the PTP on the EFB, and the predicted difficulty for a typical crew to learn IM procedures. All other models of post-experiment responses had adjusted $R^2$ values below 0.5, and were not considered useful and thus are not included in this report.

#### 3.2.1 Usefulness of IM Commanded Speed on CGD

The flight crew-rated usefulness of the IM commanded speed on the CGD was part of a group of questions about the usefulness of CGD elements, rated from 1 as “not at all useful (ignored)” to 7 as “very useful (essential).” Two models were created with this variable in the SoLR-M tool, seen in Table 3. The most useful model for prediction, included the following independent variables: airline affiliation, flight time, time since last flown, and military experience. The model, as a whole, is significant ($p < 0.01$) with an adjusted $R^2$ of 0.72. Within the airline affiliation variable, only Southwest had a significant coefficient ($p < 0.01$), indicating an increased rating of usefulness. Flight time and military experience were not significant, but were included in this model to improve the adjusted $R^2$ value. Both flight time and military experience, like a Southwest Airlines affiliation, predicted an increased usefulness rating.

**Table 3. Post-Experiment Ratings of Usefulness of IM Commanded Speed on CGD.**

| Regressors | Model |
|---|---|
| Age | - |
| Airline: AAL | 0.079 |
| Airline: DAL | 0.033 |
| Airline: FDX | 0.041 |
| Airline: SCX | 0.138 |
| Airline: SWA | $0.379^{***}$ |
| Years of experience | - |
| Flight time | 0.00001 |
| Sim type (IFD/DTS) | - |
| IM experience | - |
| Time since last flown | $-0.005^{***}$ |
| Military experience | 0.095 |
| Constant | $6.767^{***}$ |
| Observations | 24 |
| $R^2$ | 0.816 |
| Adjusted $R^2$ | 0.718 |
| Residual Std. Error | 0.150 (df = 15) |
| F Statistic | $8.332^{***}$ (df = 8; 15) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

### 3.2.2 Acceptability of MAINTAIN Clearance Phraseology

The flight crew-rated acceptability of the MAINTAIN clearance phraseology was part of a group of questionnaire items about the acceptability of phraseology in the IMAC experiment. Phraseology was rated from 1 as "Completely Unacceptable" to 7 as "Completely Acceptable." The model created in the SoLR-M tool, which can be seen in Table 4, includes airline affiliation, flight time, and time since last flown. The model is significant ($p$ < 0.01) with an adjusted $R^2$ of 0.86. Only Southwest Airlines had a significant coefficient within the airline affiliation variable, predicting a decreased rating. Flight time and time since last flown were not significant predictors but did improve adjusted $R^2$ and $C_p$. They also added value to the model in predicting flight crew responses from demographics, indicating pilots with more flight time, and pilots who had not flown recently would rate the MAINTAIN clearance phraseology lower.

**Table 4. Model of MAINTAIN clearance phraseology acceptability.**

| Regressors | Model |
|---|---|
| Age | - |
| Airline: AAL | 0.287 |
| Airline: DAL | -0.107 |
| Airline: FDX | 0.115 |
| Airline: SCX | -0.027 |
| Airline: SWA | $-4.491^{***}$ |
| Years of experience | - |
| Flight time | -0.00002 |
| Sim type (IFD/DTS) | - |
| IM experience | - |
| Time since last flown | -0.004 |
| Military experience | - |
| Constant | $7.255^{***}$ |
| Observations | 24 |
| $R^2$ | 0.903 |
| Adjusted $R^2$ | 0.86 |
| Residual Std. Error | 0.551 (df = 16) |
| F Statistic | $21.223^{***}$ (df = 7; 16) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
*$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

### 3.2.3 Acceptability of Responsibility for Spacing

The flight crew-rated acceptability of the responsibility they had for aircraft spacing was best predicted by years of experience, simulator type, time since last flown, and military experience, seen in Table 5. Of those predictors, simulator type, time since last flown, and military experience were significant ($p < 0.1$). Years of experience was included to improve the adjusted $R^2$ value and explanatory ability of the model. Overall, the model was significant ($p < 0.01$) with an adjusted $R^2$ of 0.624. More experienced pilots, IFD/DTS pilots, and pilots with military experience were associated with positive coefficients, indicating increased acceptability ratings of the responsibility for spacing. An increase in the time since pilots had last flown was associated with a decreased acceptability rating.

**Table 5. Model of acceptability of responsibility for spacing.**

| Regressors | Model |
|---|---|
| Age | - |
| Airline: AAL | - |
| Airline: DAL | - |
| Airline: FDX | - |
| Airline: SCX | - |
| Airline: SWA | - |
| Years of experience | 0.018 |
| Flight time | - |
| Sim type (IFD/DTS) | $0.587^{**}$ |
| IM experience | - |
| Time since last flown | $-0.010^{***}$ |
| Military experience | $0.426^{*}$ |
| Constant | $5.640^{***}$ |
| Observations | 24 |
| $R^2$ | 0.690 |
| Adjusted $R^2$ | 0.624 |
| Residual Std. Error | 0.508 (df = 19) |
| F Statistic | $10.559^{***}$ (df = 4; 19) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
*$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

### 3.2.4 Intuitiveness of Entering IM Clearance into EFB

The rated intuitiveness of entering an IM clearance into the EFB was best predicted by pilot age and the time since a pilot had last flown (Table 6). Overall, the model was significant ($p < 0.01$), with an adjusted $R^2$ value of 0.52. An increase in pilot age predicted an increased rating ($p < 0.05$), while an increase in time since last flown predicted a decreased rating ($p < 0.01$).

**Table 6. Model of intuitiveness of entering IM clearance into EFB.**

| Regressors | Model |
|---|---|
| Age | $0.031^{**}$ |
| Airline: AAL | - |
| Airline: DAL | - |
| Airline: FDX | - |
| Airline: SCX | - |
| Airline: SWA | - |
| Years of experience | - |
| Flight time | - |
| Sim type (IFD/DTS) | - |
| IM experience | - |
| Time since last flown | $-0.008^{***}$ |
| Military experience | - |
| Constant | $5.092^{***}$ |
| Observations | 24 |
| $R^2$ | 0.557 |
| Adjusted $R^2$ | 0.515 |
| Residual Std. Error | 0.393 (df = 21) |
| F Statistic | $13.213^{***}$ (df = 2; 21) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
*$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

### 3.2.5 Usefulness of EFB Items

A number of questions asked flight crews about the usefulness of EFB items, rated on a scale from 1 as “not at all useful (ignored)” to “very useful (essential). Models of these responses can be seen in Table 7.

The usefulness of the target route on the EFB can be best predicted by airline affiliation, flight time, simulator type, the time since a pilot has last flown, and military experience. Two airline affiliations were significant, Delta ($p < 0.01$) and Sun Country ($p < 0.05$), both predicting an increase in usefulness ratings. Also significant were flight time ($p < 0.05$) and time since last flown

($p < 0.05$), both predicting increases in usefulness rating. Overall the model was significant ($p < 0.05$), with an adjusted $R^2$ value of 0.54.

The usefulness of the merge point/waypoint on the EFB can be best predicted by airline affiliation and simulator type. In the airline affiliation variable, only FedEx had a significant coefficient ($p < 0.05$). The simulator type coefficient was also significant ($p < 0.01$). Overall the model was significant ($p < 0.01$), with an adjusted $R^2$ value of 0.69.

The usefulness of the PTP on the EFB can be best predicted by pilot age, airline affiliation, pilot flight time, and simulator type. Older pilots gave lowered ratings ($p < 0.1$) as did pilots who flew the IFD/DTS simulators ($p < 0.01$). Significant coefficients within the airline affiliation variable were for affiliation with Delta ($p < 0.05$), FedEx ($p < 0.01$), and Sun Country ($p < 0.01$). As a whole, the model was significant ($p < 0.01$), with an adjusted $R^2$ value of 0.60.

**Table 7. Models of EFB feature usefulness.**

| Regressors | Target Route Model | Merge Point / Waypoint Model | Planned Termination Point Model |
|---|---|---|---|
| Age | - | - | $-0.132^{*}$ |
| Airline: AAL | 0.585 | 1.105 | $2.250^{*}$ |
| Airline: DAL | $1.885^{***}$ | 0.079 | $2.513^{**}$ |
| Airline: FDX | 0.479 | $2.553^{**}$ | $6.962^{***}$ |
| Airline: SCX | $3.351^{**}$ | -1.947 | $7.826^{***}$ |
| Airline: SWA | -1.099 | -0.447 | $3.644^{*}$ |
| Years of experience | - | - | - |
| Flight time | $0.0001^{**}$ | - | $0.0002^{**}$ |
| Sim type (IFD/DTS) | $-1.229^{*}$ | $-2.947^{***}$ | $-4.961^{***}$ |
| IM experience | - | - | - |
| Time since last flown | $0.009^{**}$ | - | - |
| Military experience | 0.675 | - | - |
| Constant | $4.003^{***}$ | $5.895^{***}$ | $8.264^{**}$ |
| Observations | 23 | 21 | 22 |
| $R^2$ | 0.731 | 0.785 | 0.753 |
| Adjusted $R^2$ | 0.545 | 0.693 | 0.601 |
| Residual Std. Error | 0.893 (df = 13) | 1.066 (df = 14) | 1.583 (df = 13) |
| F Statistic | $3.932^{**}$ (df = 9; 13) | $8.530^{***}$ (df = 6; 14) | $4.953^{***}$ (df = 8; 13) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

### 3.2.6 Difficulty for Typical Crew to Learn IM Procedures

The pilot participant rated difficulty for a typical crew to learn IM procedures can be best predicted by airline affiliation, flight time, simulator type, and military experience (Table 8). Difficulty was rated from 1 as "very difficult" to 7 as "very easy." Overall the model was significant ($p < 0.01$), with an adjusted $R^2$ value of 0.58. In the airline affiliation variable, significant coefficients were associated with Delta ($p < 0.05$), FedEx ($p < 0.01$), and Sun Country ($p < 0.01$) – the positive coefficients indicated pilots from those airlines responded that it would be easier for the typical crew to learn those procedures. The simulator type coefficient was negative ($p < 0.01$), indicating pilots in the IFD and DTS thought it would be more difficult for the typical crew to learn IM procedures. The military experience coefficient was positive ($p < .01$), meaning pilots with military flying experience rated this item as less difficult.

**Table 8. Model of difficulty for typical crew to learn IM procedures.**

| Regressors | Selected Model |
|---|---|
| Age | - |
| Airline: AAL | -0.139 |
| Airline: DAL | $1.806^{**}$ |
| Airline: FDX | $4.936^{***}$ |
| Airline: SCX | $6.270^{***}$ |
| Airline: SWA | 0.568 |
| Years of experience | - |
| Flight time | $0.0001^{*}$ |
| Sim type (IFD/DTS) | $-2.769^{***}$ |
| IM experience | - |
| Time since last flown | - |
| Military experience | $1.707^{***}$ |
| Constant | $2.691^{***}$ |
| Observations | 24 |
| $R^2$ | 0.728 |
| Adjusted $R^2$ | 0.582 |
| Residual Std. Error | 1.103 (df = 15) |
| F Statistic | $5.006^{***}$ (df = 8; 15) |

*Notes: The baseline is UAL for airline affiliation and ASTOR for simulator type.*
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

## 3.3 General Observations

While most models could not explain the majority of variation in numeric flight crew subjective assessments, patterns of flight crew demographic coefficients were observed across models.

- Airline affiliations often appeared as significant predictors of flight crew response with relatively large coefficients.
- An increase in the time between the experiment and the time a pilot had last flown predicted decreased ratings in most models that contained that variable.
- The simulator type variable was generally associated with negative coefficients for ratings of EFB and CGD feature usefulness by flight crews who operated the IFD or DTS.
- Prior interval management experience did not significantly predict questionnaire responses in any of the selected models.
- In the selected models, military experience appeared to increase ratings of usefulness and acceptability.

# 4 Discussion and Recommendations

The results of this analysis indicate the possibility of flight crew demographic influence on questionnaire responses. The patterns of regression model coefficients across questionnaire items may suggest the relative importance of certain demographic variables in flight crew responses. In some characteristics, like airline affiliation, the pilot participants in IMAC were not representative of the current population of operational airline pilots. These relationships should be investigated further to determine not only the importance of pilot characteristics in interval management, but also the importance of accurately representing the current population of operational flight crews in interval management and other flight tech technology experiments. Sections 4.1 through 4.6 discuss these possible relationships in further detail to offer more specific recommendations.

## 4.1 Age, Length of Experience, and Flight Time

Age, length of experience, and flight time are related both linearly and conceptually. Age featured as a predictor in one useful model, significantly predicting the intuitiveness of entering an IM clearance into the EFB. Amount of experience predicted a small change in the acceptability of responsibility for spacing, but this was not significant. Flight time, however, featured in the majority of useful models and often as a significant predictor. Flight time significantly predicted increases in two usefulness ratings of EFB items.

## 4.2 Airline

Airline affiliation coefficients were often significant, indicating that airline affiliation may affect flight crew responses. This may result from differences in training or operating procedures. Procedures for configuration changes (lowering flaps and landing gear at certain altitudes, airspeeds, or points on an approach) may have affected flight crews' perceptions of IM operations. Generally, ratings given by United flight crews of acceptability and the usefulness of display items were lower than ratings by crews from other airlines. Flight crews from American and Delta gave ratings most similar to United crews. Ratings by participants from other airlines were further from the baseline. Notably, half of the flight crews in IMAC were affiliated with United Airlines, but United pilots only represent 15% of pilots employed by air carriers in the United States [12]. Further research could investigate this area in more detail with pilots from a more representative sample of airlines.

## 4.3 Military Experience

Military experience often predicted slightly increased ratings across questionnaire item categories, but this data had to be derived from the types of aircraft participants had flown. The authors strongly suggest military flying experience be explicitly included in pre-experiment questionnaires because of its potential influence on flight crews' subjective ratings.

## 4.4 Simulator Type

The simulator type variable was generally associated with negative coefficients for ratings of EFB and CGD feature usefulness by flight crews who operated the IFD or DTS. Simulator type also featured in the model of the acceptability for responsibility of spacing, suggesting pilots who operated in the full-cockpit IFD and DTS found this item more acceptable than those who operated the desktop ASTOR simulator. Further study may be able to expand on differences between desktop simulators and full-cockpit simulators in interval management research.

## 4.5 Recency of Experience

The recency of a pilot participant's experience commonly predicted questionnaire responses as a significant negative coefficient. A pilot with more time between their last flight and the experiment was more likely to provide a lower rating in most of the questionnaire items well predicted by the regression models. This may suggest recency of experience is an important factor in the flight crew acceptability of interval management, and could be a topic of further study.

## 4.6 Interval Management Experience

Interval management experience did not appear in any useful regression models of pilot questionnaire responses. Though half of the participating pilots had experience with interval management in prior experiments, the training in IMAC was designed to be thorough enough to minimize the difference between pilots with IM experience and pilots without IM experience. The thorough IM training in IMAC may be reflected in this result, but the sample size is too small to be conclusive. Further research may investigate this relationship further.

# References


[1] B. T. Baxley, S. R. Wilson, K. A. Swieringa, W. C. Johnson, R. D. Roper, C. E. Hubbs, P. A. Goess and R. F. Shay, "Human-in-the-Loop Assessment of Alternative Clearances in Interval Management Arrival Operations," NASA TP, 2016 (in preparation).

[2] B. T. Baxley, W. C. Johnson, H. N. Swenson, J. E. Robinson, T. Prevot, T. J. Callantine, J. Scardina and M. Greene, "Air Traffic Management Technology Demonstration-1 Concept of Operations (ATD-1 ConOps), Version 2.0," NASA TM-2013-218040, 2013.

[3] A. Dean and D. Voss, Design and Analysis of Experiments, New York, NY: Springer-Verlag, 1999.

[4] K. A. Swieringa, S. R. Wilson and B. T. Baxley, "System Performance of an Integrated Airborne Spacing," in *16th AIAA Aviation Technology, Integration, and Operations Conference*, Washington, D.C., 2016.

[5] B. T. Baxley, S. R. Wilson, K. A. Swieringa and R. D. Roper, "Flight Crew Responses to the Interval Management Alternative Clearances (IMAC) Experiment," in *16th AIAA Aviation Technology, Integration, and Operations Conference,* Washington, D.C., 2016.

[6] W. W. Wierwille and J. G. Casali, "A Validated Rating Scale for Global Mental Workload Measurement Applicatons," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1983.

[7] R Development Core Team, *R: A language and environment for statistical computing,* Vienna, Austria: R Foundation for Statistical Computing, 2008.

[8] W. Chang, J. Cheng, J. J. Allaire, Y. Xie and J. McPherson, *Shiny: Web Application Framework for R.,* 2016.

[9] R. H. Merys, Classical and Modern Regression with Applications, 2nd ed., Pacific Grove, CA: Duxbury, 1990.

[10] C. L. Mallows, "Some Comments on Cp," *Technometrics,* vol. 15, no. 4, pp. 661-675, 1973.

[11] M. Hollander and D. Wolfe, Nonparametric Statistical Methods, 2nd ed., New York, NY: John Wiley & Sons, 1999.

[12] Bureau of Transportation Statistics, *Air Carrier Employees,* United States Department of Transportation, 2015.

## REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** *(DD-MM-YYYY)*
01-12-2016

**2. REPORT TYPE**
Technical Memorandum

**3. DATES COVERED** *(From - To)*

**4. TITLE AND SUBTITLE**
Modeling Relationships Between Flight Crew Demographics and Perceptions of Interval Management

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**
330693.04.10.07.09

**6. AUTHOR(S)**
Remy, Benjamin; Wilson, Sara R.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
NASA Langley Research Center
Hampton, VA 23681-2199

**8. PERFORMING ORGANIZATION REPORT NUMBER**
L-20764

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
National Aeronautics and Space Administration
Washington, DC 20546-0001

**10. SPONSOR/MONITOR'S ACRONYM(S)**
NASA

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
NASA-TM-2016-219357

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Unclassified - Unlimited
Subject Category 65
Availability: NASA STI Program (757) 864-9658

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
The Interval Management Alternative Clearances (IMAC) human-in-the-loop simulation experiment was conducted to assess interval management system performance and participants' acceptability and workload while performing three interval management clearance types. Twenty-four subject pilots and eight subject controllers flew ten high-density arrival scenarios into Denver International Airport during two weeks of data collection. This analysis examined the possible relationships between subject pilot demographics on reported perceptions of interval management in IMAC. Multiple linear regression models were created with a new software tool to predict subject pilot questionnaire item responses from demographic information. General patterns were noted across models that may indicate flight crew demographics influence perceptions of interval management.

**15. SUBJECT TERMS**
Airport; Demographics; Flight crew; Human; Interval management; Simulation

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| | | | | | STI Help Desk (email: help@sti.nasa.gov) |
| U | U | U | UU | 27 | 19b. TELEPHONE NUMBER *(Include area code)* (757) 864-9658 |

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39.18