5-2011

# Re-estimating and remodeling general aviation operations

Ryan Black
*University of Arkansas, Fayetteville*

# Re-Estimating and Remodeling General Aviation Operations

an undergraduates honors thesis submitted to the

**Department of Industrial Engineering**

**University of Arkansas**

by

**Ryan Black**

March 14, 2010

**Mentor: Dr. Justin Chimka**

**Reader: Dr. Heather Nachtmann**

Keywords: statistical quality-control, general aviation, transportation security

## Introduction

Since September 11, 2001, countless steps have been taken to improve the security for commercial aviation flights. However, many of these steps were not taken and applied for general aviation. Furthermore, the security in place for general aviation is not sufficient enough to predict when future terrorist attacks in the general aviation industry may occur. The general aviation industry is still highly susceptible to terrorist attacks. By analyzing the data and information that is normally taken with general aviation plane flights, we can determine what normal or "acceptable" general aviation flight characteristics are. Using these characteristics and their acceptable variability we can then build a quality control model that can detect when there may be a high risk for a terrorist attack.

On November 17, 2008 the Department of Homeland Security released an article on strengthening general aviation security. The article concentrates on effort to minimize vulnerability of general aviation flights used to deliver illicit materials, transport dangerous weapons or people, or utilize aircrafts as weapons. This article illuminates the need for improved GA security and explains possible steps to reach this goal. One idea that the DHS is implementing is the Electronic Advance Passenger Information System, also known as eAPIS. This system will mandate general aviation operations to have more detailed information about arriving and departing planes, and the passengers and crew onboard. The Advance Information on Private Aircraft Arriving and Departing the United States final rule states that pilots must send CBP electronic manifest data relative to all people traveling on aircraft. This mandated information is sent through eAPIS or an approved alternate system one hour prior to departure for flights arriving into or departing from the United States. This data includes the following four things:

- Advance notice of arrival information ;
- Advance notice of departure information;
- Aircraft information to foster aircraft identification; and
- Complete passenger and crew manifest data.

Using this information, the Transportation and Security Administration plans to establish baseline standards of security for general aviation operations to enhance international and domestic general aviation security (DHS Press Office, 2008).

Another company taking steps to improve general aviation security is a Canadian business by the name of Transport Canada. Phase II of their Electronic Collection of Air Transportation Statistics (ECATS) allows general aviation planes to submit their air transportation data through web interfaces. This new data integration system will improve the timeliness and availability of air transportation data to be interpreted and analyzed. Transport Canada uses current and secure information technology to collect and distribute data. A collaboration of general aviation entities and a partnership between the government and industry has allowed this high security information to be shared and interpreted in order to improve general aviation security (Round Table Discussions, 2007).

In the United States, NASA has been working on constructing an "Aviation Data Integration System" also known as ADIS. This system provides rapid access to various data sources such as the following: weather data, airport operation condition reports, radar data, runway visual data, navigational charts, radar track point records and track deviation, aircraft conditions, and jeppensen charts. All of this information and data is integrated and then analyzed. There is also a data recorder in the cock pit of the aircraft that transfers data such as time since flight start, latitude, longitude, and altitude to a binary file (Iza, 2003). This data can then be used to determine when an aircraft is acting abnormally, and the threat of a terrorist attack can be examined immediately. The United States must use a combination of the above innovations in order to integrate a variety of data formats, "and transform raw data into useful and understandable information that enables productive and efficient analysis" (IDS University Affiliate Center for Multimodal Information Access and Synthesis). There must be strong and sufficient communication of information between the ground operations at an airport facility, the airport towers, the flights in progress, and the government.

The objective of this project is to understand the usual or "acceptable" characteristics of general aviation flights and airports so that unusual activity can be detected, analyzed, and resolved. To do this, a quality control model may be built to determine whether an airport is at high risk for a terrorist attack. For an appropriate model to be built there must be an adequate amount of data that describes the general aviation airport and its operations. In order to facilitate this data the general aviation industry must "continue to improve intelligence and information sharing (Homeland Security Advisory Council 2008)." This exploratory project should support work by the Center for Dynamic Data Analysis, and Purdue University Regional Visualization & Analytics Center. We will base our methods specifically on generalized linear models, and the context will be limited to GA security. This opportunity to specialize on a smaller scale should afford us unique insights. Our goal is to analyze integrated data and information using statistical quality control in order to improve standards of security in place for the general aviation industry.

## 2. Improving Estimation of General Aviation Operations:

The Federal Aviation Administration releases a terminal area forecast summary each year. This summary predicts the number of enplanements for future years to come for commercial aviation airports. Currently, this model has not been applied to the general aviation industry. To do this, historical relationships between airport passenger demand and/or activity measures and local and national factors that influence aviation activity are examined. The FAA also used regression analysis to reforecast the series. Regression models including variables that characterize airports and their activities can be used to accurately forecast the number of operations at an airport. This data can aid in building terminal area forecast models for general aviation airports (Schaufele, 2007). Predicting the annual number of operations at a general aviation airport will also aid in identifying unusual behavior at a GA airport.

The FAA administers a general aviation survey each year to assure safe operation of all aircraft in the National Airspace System. To do this the FAA classifies general aviation aircraft's in seven different categories, which include fixed wing piston, fixed wing turboprop, fixed wing turbojet, rotorcraft, other aircraft, experimental, and light-sport. The survey requests aircrafts owner to provide the following information:

1) Number of total hours flown in previous year
2) Airframe hour reading and the most common place the aircraft was flown in survey year
3) Hours flown by flight plan and flight conditions
4) Type of landing gear and number of landings
5) Fuel type and average fuel consumption
6) Percentage of hours flown by person or company other than primary owner
7) Avionics equipage

(FAA, 2007).

Due to adjustments to the general aviation survey and the way that it is administered, the response rate has been increasing for the past eight years. The collection of this data is vital in understanding baseline general aviation operations. The information obtained by these surveys can be used to build a regression model that estimates the annual number of operations at a general aviation airport.

In 2000, Hoekstra developed a methodology for estimating the annual number of general aviation operations at an airport and the annual number of general aviation operations per based aircraft at an airport. In July 2001, the FAA modified Hoekstra's model to more accurately estimate the number of general aviation operations for non-towered airports based on data from towered airports. To do this many of the same independent variables were used, however, several were added. The variables used for the regression analysis are listed below (GRA, Inc, 2001). A more descriptive list of the variables and their abbreviations is displayed in figure 2 of the appendix.

- Total based aircraft;
- Total base aircraft squared (since the number of operations tends to increase at a slower rate as the number of total based aircraft increases);
- Per capita income in the county where the airport is;
- Non-agricultural employment in the airport's county;
- Region where the airport is located;
- Population within 25 miles, 50 miles, and 100 miles;
- Airport prominence (proportion of based aircraft in region);
- Complexity of airport's based aircraft (ratio of single engine based aircrafts to total aircrafts);
- Presence of certified flight school;
- Population densities;
- Dummy variable to distinguish between towered and non-towered airports;
- Number of certified pilot schools on airport;
- Number of employees of FAR141 certified pilot schools on airport;

## 3. Steps to Create an Improved Model for Estimating General Aviation Operations

*Step 1: Model recreation*
*Step 2: Create new variables and recreate main effects model*
*Step 3: Compare GLM to choose distribution and link*
*Step 4: Create 2nd order terms associated with select continuous independent variables*
*Step 5: Eliminate noise (statistically insignificant variables) and model selection*

**Step 1: Model Recreation**
To more accurately understand the relationships between the characteristics of an airport and the annual number of airport operations, models previously constructed by GRA, Inc. were recreated using regression analysis on Stata data analysis software. An equation summary analysis is provided in figure 1 of the appendix. The main differences between equations and a description of the tables are also given in figure 1 of the appendix. The first nine equations found in table 3 of the text were remodeled using the same data set of 127 towered airports. One new variable is added to the equation to formulate the models for equations 1 through 7. The text provides $R^2$ values, but does not include $R^2_{adj}$ values or explanation for the change in variables for equations 7-10. To gain a better understanding of the text and decisions that were made, $R^2_{adj}$ values were calculated in addition to $R^2$. The $R^2$ values from the text, the $R^2$ values that we found, and the $R^2_{adj}$ values that we determined, are provided in figure 1 in the appendix.

In equation eight the independent variable describing the percent of based aircraft within 50 miles is removed, which in turn reduces the r-squared value. The reason for eliminating this variable is not explained in the text. This decision is unclear unless the $R^2_{adj}$ values are determined and analyzed. The $R^2_{adj}$ value that we determined in equation eight increased from 0.7284 in equation seven to 0.7291, which furthermore justifies their decision to remove this variable. Equation nine adds the independent variable that describes whether an airport has certified pilot schools, which increases both the $R^2$ and the $R^2_{adj}$ values. The $R^2$ values and variable coefficients found for these nine equations are identical to those found in the text. The equation with the highest $R^2_{adj}$ value is equation nine with an $R^2_{adj}$ value of 0.7334. The r-squared value for equation nine is 0.7482.

To develop a model that can accurately predict towered GA airport operations and non-towered GA airport operations, a non-towered data set would need to be included. To do this a new set of variables were used and applied to the following three data sets: towered airport data, non-towered airport data, and both towered and non-towered data combined as one data set. Equations ten, eleven, and twelve all include the same variables, but with different data sets. In equation ten, a model is created using both the non-towered and the towered data. This model does not include a dummy variable that distinguishes the data sets. The $R^2$ values found with the data and variables used in text were different than those that we found for equation ten. The $R^2$ value that GRA, Inc. found for equation ten was 0.7170, while the one we found was 0.7107 (GRA, Inc, 2001). These differences were also found for equation twelve. The $R^2$ value that we found was 0.6448, while the one that they found was 0.6480. The coefficients we found for this equation also differed from the text. The $R^2_{adj}$ values for these three models were slightly lower than that of equations 6-10. The chow test result rejects the hypothesis that the non-towered data and towered data come from similar distributions as represented by equation 10. Furthermore, these lower values are caused by the addition of the non-towered data.

To limit the effect caused by the differences between data sets, a dummy variable is introduced in equation 13. The dummy variable takes the value 1 if the data is from a towered airport and 0 if the data is from a non-towered airport. This variable helps the model distinguish between towered and non-towered data. The $R^2$ value and variable coefficients in equation thirteen from the text differ from those that we found in our model. The $R^2$ value that they calculated was 0.7430, while the $R^2$ value that we determined for the model was 0.7386 (GRA, Inc, 2001). The model that we developed is a more accurate estimation of the annual number of general aviation airport operations for the towered and non-towered data sets. Equations fourteen and fifteen used all of the same variables except the dummy variable. In equation fourteen the variables were applied to the towered data set and in equation 15 the variables were applied to the non-towered data set. The $R^2$ value and variable coefficients in equation 15 also differed from those in the text. The data and variables that we used in equation 16 and equation 13 were identical, however, in equation sixteen fifteen non-towered airports were randomly excluded from the data. These airports were excluded in order to assess the accuracy of the model for estimating the annual number of airport operations for these fifteen excluded airports. The regression equation represented by the data set and variables used in equation 13 of the text best estimates the number of annual airport operations. Although equation fourteen's $R^2$ value and $R^2_{adj}$ value are higher, equation 13 is a better predictor of non-towered airport operations because it includes the non-towered airports data set

The predictability that the model provides could be useful in estimating general aviation airport operations at non-towered airports for a large data set contained in the Terminal Area Forecast. This model could also be used to examine the plausibility of the 5010 data and provide APO staff the means to assess claims of airport operations at poorly documented non-towered airports. The regression equations that we computed were more accurate than those found in the text. The differences in variable coefficients and $R^2$ values occur solely in equations that use the non-towered airports data. Therefore, these errors can be attributed to the manipulation of non-towered airport data set. The errors with non-towered airport data were corrected and the regression equations along with their $R^2$ values that we determined are a more accurate representation of the data. The most accurate regression model out of the 16 models that were produced is equation 13. This model includes the data from all towered and non-towered airports. Equation 13 also includes the dummy variable, which contributes significantly to the model. The $R^2_{adj}$ value for this model is 0.7292.

Equation seventeen estimates the number of operations per based aircraft at an airport. This model was developed from variables in Hoekstra's model and new local variables that were added. This model had much less explanatory power than that of the equations used to estimate annual general aviation airport operations. Furthermore, more research and analysis was not performed on models that estimated operation per based aircraft for many reasons. First and foremost the $R^2$ values for these equations were far lower than those found with equations estimating annual general aviation airport operations. Another reason is because the regression equation found for estimating annual airport operations per based aircraft contained mainly categorical and regional variables. These models did not include numeric characteristics describing local factors surrounding the airport.

**Step 2: Create New Variables**

To further improve the accuracy of the model, we recreated their model using several new adjusted variables. Instead of including a ratio of single engine aircraft to total based aircraft (sebaba), a new variable, single engine based aircraft (seba) was created. This variable was created using the data values from the total based aircrafts and the ratio of single engine aircrafts to total based aircraft. A new regional variable was also created to better assess the location of GA airports in the model. In the FAA's model they used five separate regional variables that would take on the value 1 if the condition was met and 0 otherwise. These variables included an Alaska variable, a pacific coast variable, a FAA west regional variable, and a FAA east regional variable. To simplify the regional variables, only one variable was used to distinguish between the five regional conditions. This new categorical variable takes on the value of 1 if the GA airport is located in Alaska, 2 if the GA airport is located along the Pacific coast, but not in Alaska, 3 if the GA airport is in the FAA West region, but not in Alaska or bordering the Pacific coast, 4 if the GA airport is located in the FAA East region, and 0 if none of these conditions were met. In other words, if the regional variable does not take on a value of 1, 2, 3, or 4 then the airport must be located in the central region of the United States.

Next, we evaluated the new variables that we created by running a regress and main effects model of the data. The $R^2_{adj}$ value that we found for this new version of the FAA's model was 0.7208, which is very similar to the FAA's best model in equation 13. The $R^2$ value for this model that we created was 0.7426, which is a higher $R^2$ value than that of any of the FAA's model. Therefore, with future adjustments we should be able to adjust and make major improvements to the FAA's model for estimated annual operations at a general aviation airport.

To understand the contribution that the regional variable provides for the model, the main effects model was pivoted on the categorical regional variable. To do this, each category of the regional variable was compared to all other categories. In doing this we found that there was no significant difference between the different categories for the regional variable. However, when Alaska was used as the baseline for the regional variable, the p-values were much lower than those of any other regional categories. These values were below 0.10, but not below 0.05. To examine the p-values more closely we made a p-value chart for the regional variables. This chart is shown below.

Table 1: P-Value Chart for Pivoted Regional Variables

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | X | **0.065** | 0.278 | 0.779 | 0.687 |
| 1 | X | X | 0.149 | **0.063** | **0.055** |
| 2 | X | X | X | 0.346 | 0.194 |
| 3 | X | X | X | X | 0.982 |
| 4 | X | X | X | X | X |

When using a level of significance of 0.05, none of the regional variables seem to be significant; however, when we removed the regional variable the $R^2_{adj}$ value went down. Therefore, we increased the level of significance to 0.10. When analyzing the p-value chart with a level of significance of 0.10 we noticed that the regional variable of category 1 (Alaska) was significant for three out of the four other categories. Therefore, we changed the regional variable to isolate Alaska. This categorical variable will take on the value 1 if the GA airport is located in Alaska and 0 otherwise. Next, this new variable was assessed by running a main effects regression model of the data with Alaska as the single regional variable. The $R^2_{adj}$ value for this model was 0.7220.

**Step 3: Compare GLM to choose distribution and link**

Although the model has potential to offer crucial advice for estimating the annual number of GA operations at an airport, a problem exists with the assumption of normal distribution for the model. In equation sixteen, fifteen non-towered airports were randomly removed from the data set to test as examples of the regression model. The data values for these fifteen airports were plugged into the variables of equation sixteen to estimate their annual number of GA operations. These values could then be compared to the values of the state estimates to assess the equations validity. These responses were calculated using the corrected equations that we had determined earlier. The responses for these fifteen airports using the regression model and the state estimates are shown in the table below.

Table 2: Regression Model Estimates and State Model Estimates Using Equation 16

| **Airport** | **Regression Model** | **State Estimate** |
|---|---|---|
| 0W3 | 31957 | 35509 |
| RIF | 2191 | 5922 |
| BUM | 7288 | 7978 |
| AJG | 8752 | 10964 |
| GLY | 13962 | 11277 |
| M58 | 8035 | 12349 |
| FWC | 5028 | 13292 |
| OKV | 32881 | 17887 |
| O61 | 51537 | 20000 |
| RBG | 22127 | 20899 |
| GGW | **-1493** | 21908 |
| 7S5 | 39585 | 27862 |
| CBE | 21359 | 32118 |
| JYO | 66272 | 68448 |
| ESN | 49928 | 75949 |

Attention should be directed to the negative value -1,493 that the regression model calculated for the airport labeled GGW. Since a value of the -1,493 annual GA operations for an airport is physically impossible, the normal distribution assumption is theoretically unrealistic. To have a truer understanding of the data and the estimation of the annual number of GA operations at an airport, other distributions were examined. Generalized likelihood models were created to choose a distribution and the appropriate link. A table including the log likelihood values for each distribution with both dependent variables is displayed in the table below.

Table 3: Distribution Analysis of the Responses OPS and OPSBA

| Distribution | OPS | OPSBA |
|---|---|---|
| Gaussian | -2690.9630 | -1613.3117 |
| Inverse Gaussian | -3765.0426 | -2286.4808 |
| Bernoulli | N/A | N/A |
| Poisson | -2171228.5220 | -15588.2945 |
| Negative Binomial | -2675.8382 | -1649.0778 |
| Gamma | -2675.8351 | -1648.8196 |

To assess the validity of the FAA's model for estimating annual operations at a General Aviation airport, all possible distributions were examined. Furthermore, generalized likelihood models were created for the gamma and negative binomial distributions. The gamma distribution had the lowest likelihood value; therefore, the different links for this model were assessed. Ultimately, the identity link of the gamma distribution contained the lowest log likelihood value at -2621.697. However, when comparing this log likelihood value to that of the normal distribution we noticed that the default link for the normal distribution had a lower log likelihood value. To prevent future mistakes, a table was created to assess the log likelihood values for all appropriate links for each distribution. This table is displayed below. After examining all possible combinations, we concluded the default link (identity) of the Gaussian distribution most accurately represents the data because it contained the lowest log likelihood value.

Table 4: Log Likelihood Values for the Links of Each Distribution

| | Distribution | | | | |
|---|---|---|---|---|---|
| **Link** | Gaussian | Inverse Gaussian | Poisson | Negative Binomial | Gamma |
| Identity | -2533.542 | -3765.04 | -538321.884 | -2621.702 | -2621.697 |
| Log | -2561.679 | -3765.04 | -699513.79 | -2627.935 | -2627.93 |
| Power -1 | -2648.512 | -1.87E+21 | -8845014.933 | -5162745.725 | -2.80E+13 |
| Power -2 | X | X | X | X | X |
| Nbinomial | X | X | X | -5162748.9 | X |

**Step 4: Create 2nd order terms associated with select continuous independent variables**

Many of the p-values for this model were above 0.10; however discarding their interaction with other variables would not be a valid analysis method. A full second order model would also be unrealistic because this would leave the observation to variable ratio at less than two. In order to consider interaction, but not a full second order model, only continuous independent variables with a higher p-value than 0.10 were analyzed. The variables that satisfied this rule were vitfsnum, vitfsemp, in50mi, in100mi, pop50, and pop25. An explanation of these variables can be found in figure 2 of the appendix. Fifteen new variables were created by taking the products between each of these variables. The variable far139 was also removed because it had a high p-value in the previous model and was not continuous. Also, the variables vitfsnum and vitfsemp are better indicators of what the variable far129 describes.

Next, a regression model was created, which included the 15 new variables that were created in order to assess interaction. The $R^2_{adj}$ value jumped substantially from 0.7220 to 0.7753.

**Step 5: Eliminate noise (statistically insignificant variables)**

The next step in this analysis process was to determine which variables contributed to the model and which variables were bringing down the $R^2_{adj}$ value. When looking over the p-values for each variable from the regression model, we noticed that the p-value was very high for each variable that included vitfsemp. Furthermore, the variable vitfsemp and the second order variables that included vitfsemp were removed from the model. When the regression model was ran without these variables, the $R^2_{adj}$ value surprisingly dropped from 0.7753 to 0.7734. This told us that some of these variables did contribute to the model, therefore; they were added to the model once again. To prevent mistakes like this, a step by step analysis of the variables would need to be performed.

While examining the p-values, we noticed that the variable in100mi, which represents the percentage of based aircraft among based aircraft at GA airports within 100 miles, had the highest p-value at 0.914. The variable in50mi is very similar to the variable in100mi, but instead it measures the percentage of based aircraft among based aircraft at GA airports within 50 miles. The p-value for in50mi is 0.001. Knowing that the variable in100mi does not contribute to the model and that the variable in50mi conveys very similar information, the variable in100mi was removed from the model. When the regression model was ran without this variable the $R^2_{adj}$ value increased from 0.7753 to 0.7764. Although, the variable in100mi did not contribute by its self, its second order variables do deem significance in accordance to their p-values, so they were kept in the model.

While examining the p-values of this new model, one particular variable popped out to be very insignificant. This was the interaction variable vitfsnumvitfsemp. This variable had a p-value of 1.00. This variable was removed from the model due to its extremely high p-value. After doing this, the $R^2_{adj}$ value increased from 0.7764 to 0.7775. The variable vitfsemppop50 also seemed to be lowering the $R^2_{adj}$ value because it had a p-value of 0.945. This variable was removed from the model, furthermore, causing the $R^2_{adj}$ value to jump from 0.7775 to 0.7786. Many other variables were removed do to their high p-value. After each variable was removed, the regression model was re-run to make sure that the $R^2_{adj}$ value increased. If the $R^2_{adj}$ value decreased, that variable was put back into the model. The final regression model was found to have an $R^2_{adj}$ value of .7831. The adjusted r-squared value ($R^2_{adj}$) is the proportion of variation in GA operations explained by the model, discounted for amount of information required to predict. The p-values of the variables used in the final model are displayed below.

Table 5: Final Regression Variable's Coefficients and P-Values

| Variable | Coefficient | P>t |
|---|---|---|
| towdum | 13646.7 | 0 |
| ba | 177.4382 | 0 |
| pop | -17.84235 | 0.023 |
| pci | 0.2596261 | 0.118 |
| emp | 42.20493 | 0.015 |
| aal | -16693.47 | 0.031 |
| in50mi | 31313.69 | 0.002 |
| pop100 | 0.0020082 | 0 |

| | | |
|---|---|---|
| pop50 | -0.0027599 | 0.11 |
| pop25 | 0.0074031 | 0.073 |
| vitfsnu~50mi | 32902.27 | 0.002 |
| vitfsnu~00mi | -58174.82 | 0.009 |
| vitfsnump~50 | -0.0002633 | 0 |
| vitfsem~00mi | -1639.915 | 0.057 |
| vitfsempp~25 | 0.0001215 | 0.278 |
| in50miin10~i | -59167.93 | 0.01 |
| in50mipop50 | -0.0540262 | 0.005 |
| in100mipop50 | 0.2930406 | 0 |
| in100mipop25 | -0.127762 | 0.156 |
| pop50pop25 | -6.83E-10 | 0.015 |
| ba2 | -0.2267402 | 0.096 |
| vitfsemp | 276.9878 | 0.304 |
| _cons | -7985.697 | 0.036 |

## 4. Conclusions and Future Considerations

The research conducted in this report has produced a more accurate model for estimating the annual number of operations at a general aviation airport. The equations recreated in this report from GRA, Inc.'s original models are a more exact representation of the data. The $R^2_{adj}$ values are a vital tool in developing a reliable statistical quality control model. The final regression model produced in this report can be used to accurately estimate the annual number of operations at a general aviation airport. The adjusted R-squared value found for the final model was 0.7831. The best $R^2_{adj}$ that GRA, Inc. found using the same data sets was .7292. This gives our final model an overall improvement of 7.4% from GRA, Inc.'s best model. The $R^2_{adj}$ value tells us the proportion of variation in GA operations that are explained by the model, discounted for amount of information required to predict. This information may be used to create terminal area forecast summaries for GA airports. This model may also be used to detect unusual behavior based on the annual number of operations at an airport. For example, a small GA airport in New Mexico may only be expected to have 5,000 annual operations at their airport; however, if the airport reports a much larger number of annual operations, it may signal unusual behavior such as illicit drug trafficking. Drug trafficking is a common and overlooked problem in the general aviation industry that can be examined and detected with the model provided in this report.

One future objective of this project is to provide recommendations for multiple data stream integration applied to transportation security. Methods must be created to improve monitoring across collaborative data sources. Improved information technology in the general aviation industry will lead to recommendations for early detection decision aids for GA security. Another future goal is to create controls that will be used to manipulate data and gain a better understanding of what acceptable general aviation characteristics and operations are. If good models of usual activity fail to predict, then unusual activity may indicate a threat. This model-based control of GA security displayed in this report may also be extended to other contexts such as highway, maritime transportation systems, mass transit, pipeline systems, and rail.

Appendix

| GRA | Data Set | # Airports | Dummy Var.? | # Ind. Var. | $R^2$ | GRA $R^2$ | $R^2_{adj}$ |
|---|---|---|---|---|---|---|---|
| 1 | Towered | 127 | NO | 1 | .5564 | .5564 | .5529 |
| 2 | Towered | 127 | NO | 2 | .6402 | .6402 | .6344 |
| 3 | Towered | 127 | NO | 3 | .6664 | .664 | .6583 |
| 4 | Towered | 127 | NO | 4 | .7031 | .7031 | .6934 |
| 5 | Towered | 127 | NO | 5 | .7231 | .7231 | .7117 |
| 6 | Towered | 127 | NO | 6 | .7351 | .7351 | .7218 |
| 7 | Towered | 127 | NO | 7 | .7435 | .7435 | .7284 |
| 8 | Towered | 127 | NO | 6 | .7420 | .7420 | .7291 |
| 9 | Towered | 127 | NO | 7 | .7482 | .7482 | .7334 |
| 10 | All | 232 | NO | 8 | .7107 | .7170 | .7003 |
| 11 | Towered | 127 | NO | 8 | .7274 | .7270 | .7089 |
| 12 | Non-Towered | 105 | NO | 8 | .6448 | .6480 | .6152 |
| 13 | All | 232 | Yes | 8 | .7386 | .7430 | .7292 |
| 14 | Towered | 127 | NO | 7 | .7476 | .748 | .7327 |
| 15 | Non-Towered | 105 | NO | 7 | .5627 | .569 | .5311 |
| 16 | All | 217 | Yes | 8 | .7418 | .745 | .7318 |

Figure 1:  Recreated Data Analysis and Equation Summary for the FAA's Model of Annual GA Operations

| | |
|---|---|
| **Equation 1-9** | • **Same data for equations 1-9**<br>• **One variable is added each time** |
| **Equations 10-12** | • **Same equation is used**<br>• **Different data sets are applied** |
| **Equations 13-15** | • **Same equation is used**<br>• **Different data sets are used**<br>• **Dummy variable used in equation 15** |
| **Equation 16** | • **Same as equation 13, but excludes 15 non-towered airports** |

Figure 2: Description of Equation Sets

| Variable | Description |
|---|---|
| TOWDUM | Categorical variable, 1 if airport is towered airport, 0 otherwise |
| OPS | Annual GA Operations at an airport |
| OPSBA | Annual GA Operations per Based Aircraft(BA) at an airport |
| BAE100 | Categorical variable, 1 if airport based aircraft is 100 or greater, 0 otherwise |
| BA | Total Based Aircraft at an airport |
| BA2 | Based Aircraft squared |
| POP | County population where airport is located |
| PCI | Per Capita Income in the county in which the airport is located |
| EMP | Non-agricultural Employment in the airport's county |
| FAR139 | Categorical variable, 1 if airport is certificated for commercial air carrier service, 0 |
| WSTAK | Categorical variable used in place of WACAORAK in Hoekstra's model |
| WACAORAK | Categorical variable, 1 if state is CA, OR, WA, or AK, 0 otherwise |
| WST | Categorical variable, 1 if airport is located in FAA Western Region, 0 otherwise |
| AAL | Categorical variable, 1 if airport is located in Alaska, 0 otherwise |
| R12 | Categorical variable, 1 if airport is located in FAA New England Region or FAA Eastern |

| | |
|---|---|
| VITFS | Presence of absence of FAR141 certificated pilot school |
| VITFSnum | Number of FAR141 certificated pilot schools at an airport |
| VITFSemp | Employees of FAR141 certificated pilot schools at an airport |
| %in50mi | Percentage of based aircraft among based aircraft at GA airports within 50 miles |
| %in100mi | Percentage of based aircraft among based aircraft at GA airports within 100 miles |
| Se BA/BA | Single engine based aircraft/All based aircraft |
| Pop100 | 1998 Population within 100 miles |
| Pop50 | 1998 Population within 50 miles |
| Pop25 | 1998 Population within 25 miles |
| Pop25/100 | Ratio of Pop25 to Pop100 |

Figure 3: Variable List with Descriptions and Explanations

References

DHS Press Office. (2008). Homeland Security: Fact Sheet General Aviation. Retrieved February 15 ,2009, from http://www.dhs.gov/xnews/releases/pr_1226955842674.shtm.

Iza, Sara. (2003). NASA: Aviation Data Integration System. Retrieved February 6, 2009 from http://ti.arc.nasa.gov/m/pub/569h/0569%20(Kulkarni).pdf.

GRA, Inc. (2001). Federal Aviation Administration: Model for Estimating General Aviation Operations at Non-Towered Airports Using Towered and Non-Towered Airport Data. Retrieved February 28, 2009 from www.faa.gov/data_statistics/**aviation**_data_statistics/**general_aviation**/media/GA**Model**3F.doc

Homeland Security Advisory Council. (2008). DHS: Top Ten Challenges Facing the Next Secretary of Homeland Security. Retrieved October 25, 2009 from http://www.dhs.gov/xlibrary/assets/hsac_dhs_top_10_challenges_report.pdf.

Round Table Discussions. (2007). General Aviation Economic Footprint – Measurement. Retrieved February 6, 2009 from http://www.upac.ca/pdf/TC_Presentation_ECATS_GA_Roundtable.pdf.

Schaufele, Roger. (2007). Federal Aviation Administration: Terminal Area Forecast Summary. Retrieved March 2, 2009 from http://www.faa.gov/data_statistics/aviation/taf_reports/media/TAF2007-2025Summary.pdf.

(2007). FAA: Appendix A: Methodology for the 2007 General Aviation and Part 135 Activity Survey. Retrieved February 28, 2009 from http://www.faa.gov/data_statistics/aviation_data_statistics/general_aviation/CY2006/media/FAA%202006%20Appendix%20A%20Methodology%20102207%20FINAL.pdf.

.