Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Introduction

• Growing air traffic demand and the implementation of autonomous NextGen technologies create risks to NAS safety and efficiency

• Go-around, is initiated by either the pilot or the controller, to abort landing of an aircraft that is on final approach under unsafe conditions. E.g., wind shear, obstructions on the runway
Introduction

• Go-around is an operational anomaly that degrades the system efficiency
  – Go-around itself is a challenging maneuver
  – The outcome of go-around can be hazardous (e.g. fuel emergency)
  – Extra fuel, extra crew workload, extra delay and decreased airport throughput

• Current literature have investigated go-around decision-making policy and the performance of go-around maneuvers

• Our analysis is unique in its attention to the causal factors behind the go-around occurrence
Project Goals

• Given factor of interest, quantify how these factors contribute to the go-around occurrence
  – In-trail relationship
  – Airport and weather condition
  – Go-around clustering effect
  – Aircraft localization
  – Flight specific characteristics

• Eventually predict the probability of go-around for each specific flight at a certain time prior to the go-around occurs
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Data Sources

• Sherlock Data Warehouse
  – Integrated Flight Format (IFF) Track Data (6~30 second update)
  – Reduced Data Summary

• Aviation System Performance Metrics (ASPM)
  – Airport information for each quarter hour

• After data cleaning and matching, there are on average 525 arrivals per day in JFK airport within the analysis period, April 1\textsuperscript{st} to December 24\textsuperscript{th}, 2018
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Overview

• Apply trajectory-based anomaly detection algorithm to raw trajectory data for each flight to identify go-around occurrence
• Derive a set of contributing factors and map them with the flight trajectory using extrapolation technique
• Estimate and apply statistical model to quantify the contributions of different factors
  – Principal Component Logistic Regression (PCLR) Model
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Go-around Detection Algorithms

**Flight Profile During Final Approach (Normal)**

- **Altitude** (in 100ft)
- **Distance to the airport** (in nautical mile)

**Flight Profile During Final Approach (Go-around)**

- **Altitude** (in 100ft)
- **Distance to the airport** (in nautical mile)

A: the altitude at the start of ascent (≤ 5500 feet)
the distance to the airport (< 10 nm)

B: the total altitude gain during the ascent (≥ 400 feet)
Go-around Detection Algorithms

- **Step 0: Trajectory Cleaning**
  - Filter out JFK arrival trajectories
  - Exclude discontinuity trajectories

- **Step 1: Pairing Landing Runway**
  - Project the 5-min final approach trajectory segment on runways
  - Calculate the projection distance of each track point, which votes for the closest runway line segment

- **Step 2: Altitude/Distance Test**
  - Apply piecewise linear regression to identify points at which the slope of the altitude/distance to runway profile is changed
  - Thresholding

- **Step 3: Subject-Matter Expert (SME) Consultations**
  - Visualization inspection
Go-around Detection Algorithms

- 691 detected go-arounds from April 1st to Dec 24th, 2018 are validated by visualization inspection (accounting for 0.42% of JFK arrivals)
- 90% of the go-arounds occur within [0, 5) nm away from its landing runway end
- Assume "5nm" as information cutoff point
  - Set 1 if flight is detected as a go-around occurring within [0, 5) nm to its landing runway end, 0 otherwise
  - Any Information (features) after flight passing 5nm arc is unavailable
  - Metrics are comparable if all the flights are at 5nm to its landing runway

1. From 2012 to 2017, the average percent of go-arounds reported by FAA across 30 Core airports in US, is 0.3%.
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Feature Engineering

• **Feature Extrapolation**
  – Extrapolation technique is applied to make sure that no information from future is included at a certain time prior to the go-around
  
  – 1D Fourier Transform: Period $x[n]$ and its N-point Discrete Fourier Transform (DFT) sequence $X[k]$ are **periodic** by N, the inverse DFT:
    
    $$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \exp\left(j \cdot \frac{2\pi}{N} \cdot kn\right)$$
    
  – Extrapolate samples $x[n + N]$
  – RMSE varies between 0.012 – 0.110
Feature Engineering

• **Localization Performance:** metrics which are derived from the track data to measure how anomalous the trajectory is
  
  – Flight altitude at 5-nautical-mile away from the landing runway $h_i^d$
  – Ground speed at 5-nautical-mile away from the landing runway $v_i^d$
  – Kinetic energy height$^1$

$$H_{energy} = h_i^d + \frac{(v_i^d)^2}{2g} \text{ (feet)}$$

where $g$ is the gravitational acceleration. This metric only uses surveillance data, no need of aircraft mass and fuel consumption profile

Feature Engineering

• Localization Performance
  – Projection distance to Extended Runway Centerline (ERC)
  – Angle with ERC
  – Glideslope (decent) angle
Feature Engineering

• **In-trail Relationship:** pairing leading-trailing aircrafts when the analyzed (trailing) aircraft is at 5nm to its landing runway end
  – Loss of separation: \( S_{\text{los}} = \max(0, S_{\text{FAA}} - S_{\text{actual}}) \)
  – Ground speed difference
  – "Closing” (vs. opening) dummy var.
  – No leading aircraft dummy var.
Feature Engineering

• **Airport & Weather Condition:** quarter-hourly features are matched when the analyzed aircraft is at 5nm to its landing runway end
  – Arrival demand: # aircrafts intending to arrive
  – Arrival rate: Airport supplied Arrival Rate (AAR) for capacity
  – The change of runway configuration: 1 if the used runway configuration during the observed time period is different from the preceding/succeeding 15-min period, and 0 otherwise
  – Wind speed: headwind component is subtracted
  – Visibility: discretized into [0, 3), [3, 5), [5, 10) miles
  – Ceiling: discretized at 500, 1000, 3000 feet
  – Meteorological condition: 1 represents VMC
Feature Engineering

• **Clustering Effect:** we observed that a go-around was more likely to occur when leading aircrafts initiated go-arounds
  – Closest go-around time: the minimal time interval between the initiation time of all (other) go-around flights (in minutes; $T$)
  – Go-around count: # go-around flights within a one-hour time period of the analyzed flight (counts; $3$)
Feature Engineering

• **Flight-specific Characteristics**
  – Airline type: 1 if the analyzed flight is operated by domestic carriers
  – Aircraft type: 1 if the analyzed flight is wide body aircraft
  – Landing runway: 1 if the runway (04L, 04R, 22L, 22R, 13L, 13R, 31L, 31R) is the landing runway of the analyzed aircraft
Methodology

• **Vanilla Logit Model**
  \[ V = \sum_{i} \beta_i x_i ; \quad \Pr(y_i = 1 \mid X) = \frac{1}{1 + \exp(-V)} \]
  - \( x_i \)'s are original derived features
  - Multicollinearity: the majority of coefficients are not significant at 0.05 level, and most of which have unexpected signs

• **Principal Component Logit Regression (PCLR) Model**
  - Apply PCA to de-correlate and reduce the dimensionality of the original features
  - Data point “\( x_i \)” in the original coordinate system are transformed as “\( x_i v_j \)” in the PC coordinate system, where \( v_j \) is the eigenvector of \( X^T X \)
  - Regressing dependent variable on the principal components formed by original features
Methodology

• **Principal Component Logit Regression**
  
  – Normalize the numerical data matrix and center the categorical data matrix. Perform **PCA** for a mixture of categorical and numerical variables using Generalized Singular Value Decomposition (GSVD)
  
  – **Select** the first 19 principal components which explain 90.4% of the total variance based on the factor eigenvalues > 0.6
  
  \[ F = W^T \gamma \]

  – **Regress** the go-around indicator \((y)\) on the selected PCs \((F)\), using logistic regression model, and remove insignificant PCs
  
  \[ y = F \gamma = X (W^T \gamma) = X \beta \]

  – **Transform** the factor coefficients \((\gamma)\) back to the scale of the actual covariates \((\beta)\), using eigenvectors corresponding to the selected principal components \((W)\)
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Results

• In-trail relationship features
  – The loss of separation has positive effect on go-around occurrence
  – Too high or too low flight ground speed and closing scenario during final approach are positively associated with go-around occurrence

• Airport and weather condition features
  – Increased ceiling and visibility reduces the likelihood of go-around occurrence, and this effect is diminishing when the weather condition is good in itself
  – Higher wind speed increases the likelihood of go-around occurrence
  – The change of runway configuration has significant positive impact, which could be caused by the change of approach pattern, interrupted landing procedure, and increased crew workload
  – Go-around is more likely to occur when the airport has high arrival demand, low airport capacity, under IMC conditions
Results

• **Clustering effect features**
  – The time duration to the closest go-around has negative effect, while the count of go-arounds within 1-hr time window has positive effect
  – This implies go-around tend to occur in clusters and this behavior cannot be fully explained by the other variables in the model

• **Localization performance features**
  – Flights with high altitude, high ground speed, and high kinetic energy height at 5nm to the runway touchdown zone would more likely to have go-arounds

• **Flight-specific characteristics**
  – Flights operated by non-domestic carriers are more likely to have go-arounds, which could be caused by language issue or the flight crew are not familiar with the airport
  – Wide body aircrafts are more likely to have go-arounds
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Counterfactual Analysis

• To further quantify the contributions for different factors on go-around occurrence

• Selected factor is set to the best scenario value to minimize the probability of go-around occurrence at one time

• Calculate the percentage reduction between the baseline go-around rate $P_{baseline}$ (0.347%) and the expected go-around rate given the selected factor is set to the best scenario value $X'$

$$\% \text{ reduction} = \frac{P_{baseline} - E[P(Y = 1 \mid X')]}{P_{baseline}}$$
Counterfactual Analysis

- The baseline go-around rate is 0.347%.
- The airport ceiling (20%) and aircraft speed control (22%) are the most important factors of go-around occurrence.
- Go-around would decline about 9.5% if the aircraft is properly aligned with ERC at 5nm.
- No change of runway configuration and high airport capacity could reduce go-arounds about 7.8% and 3.75%, respectively.
Outline

• Introduction
• Data Sources
• Methodology
  – Dependent Variable: Go-Around Detection
  – Metrics: Feature Engineering
• Results
• Causal Analysis of Go-Arounds
• Conclusions
Conclusions

• Effects of aircraft localization, airport and weather condition, go-around clustering effect, in-trail relationship and flight-specific characteristics have been estimated for JFK arrivals.

• Estimates are statistically significant and have the expected signs, and are in line with conclusions drawn from:
  – researches using full-flight simulator trails that speed and localizer deviation have strong influences on go-around decision.
  – interviews with controllers and pilots that go-around often occur when the aircraft is not properly aligned or wind shears.

• We quantify the contribution of factors in ATM system in causing go-arounds - the airport ceiling, aircraft speed control, aircraft localization performance are the three most salient factors on go-around occurrence.
Applications

• Identify countermeasures to reduce go-arounds, and more generally the anomalous states that are inherently undesirable
• Develop a real-time tool that can remediate, situations in which there is a substantial risk of go-around
• Augment the limited individual experience of air traffic controllers by summarizing historical patterns of go-around occurrence
• This study makes the foundation for the work of real-time prediction
Ongoing Research

• Incorporate a broader range of features
  – Airport surface operation data are available recently
• Improve statistical modeling
• Extend to analyze other types of flight anomalies for a larger number of airports
• Use machine learning methods to predict go-around occurrence based on causal factors information
• Use Hidden Markov Model to generate the sequential probabilities of go-arounds during the approach process
Thank you!

Q&A

Lu Dai
dailu@berkeley.edu
## Estimation Results

<table>
<thead>
<tr>
<th>Variable Code</th>
<th>Model I: Logistic model</th>
<th>Model II: Logistic principal component regression model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>(Standard Error)</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.857***</td>
<td>(0.370)</td>
</tr>
<tr>
<td>Arr_rate</td>
<td>-0.121</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Visibility_1</td>
<td>-0.046</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Visibility_2</td>
<td>0.029</td>
<td>(0.074)</td>
</tr>
<tr>
<td>Visibility_3</td>
<td>-0.204**</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Ceiling_1</td>
<td>0.048</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Ceiling_2</td>
<td>0.154</td>
<td>(0.150)</td>
</tr>
<tr>
<td>Ceiling_3</td>
<td>-0.126</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Ceiling_4</td>
<td>0.048</td>
<td>(0.060)</td>
</tr>
<tr>
<td>MC=V</td>
<td>-0.767</td>
<td>(0.343)</td>
</tr>
<tr>
<td>Speed</td>
<td>-0.120</td>
<td>(0.219)</td>
</tr>
<tr>
<td>Energy</td>
<td>0.287</td>
<td>(0.200)</td>
</tr>
<tr>
<td>Speed_diff</td>
<td>0.033</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Chasing</td>
<td>-0.020</td>
<td>(0.155)</td>
</tr>
<tr>
<td>No_leading</td>
<td>0.269</td>
<td>(0.184)</td>
</tr>
<tr>
<td>PC1</td>
<td>0.250***</td>
<td>(0.0013)</td>
</tr>
</tbody>
</table>

### Additional Tables

#### Observations:

<table>
<thead>
<tr>
<th>Variable Code</th>
<th>Model I: Logistic model</th>
<th>Model II: Logistic principal component regression model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>(Standard Error)</td>
</tr>
<tr>
<td>Horiz_dist</td>
<td>0.040</td>
<td>(0.295)</td>
</tr>
<tr>
<td>Rwy_13L</td>
<td>-1.629</td>
<td>(0.706)</td>
</tr>
<tr>
<td>Alt</td>
<td>0.391</td>
<td>(0.134)</td>
</tr>
<tr>
<td>AC_airline</td>
<td>0.308</td>
<td>(0.131)</td>
</tr>
<tr>
<td>AC_body</td>
<td>0.342</td>
<td>(0.156)</td>
</tr>
<tr>
<td>Rwy_31R</td>
<td>-1.168</td>
<td>(0.239)</td>
</tr>
<tr>
<td>Rwy_22L</td>
<td>-0.943</td>
<td>(0.180)</td>
</tr>
<tr>
<td>Arr_demand</td>
<td>0.096</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Sub_wind</td>
<td>0.203</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Rwy_22R</td>
<td>-0.781</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Rwy_change</td>
<td>0.099</td>
<td>(0.166)</td>
</tr>
<tr>
<td>Loss_of_sep</td>
<td>0.166</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Rwy_04L</td>
<td>0.746</td>
<td>(0.168)</td>
</tr>
<tr>
<td>Rwy_31L</td>
<td>-1.198</td>
<td>(0.305)</td>
</tr>
<tr>
<td>Closest_time</td>
<td>-0.238</td>
<td>(0.071)</td>
</tr>
<tr>
<td>GA_occur_ct</td>
<td>0.215</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>
Causal Models

• Retrospective causal inference model
  – We look to the past in an attempt to discover any significant correlation between differences in the level of potential causal factors and go-around occurrence
  – Relatively weak causal effect, but we are good to use it for investigating unknown causal factors, especially in the early stages of research

• Randomized-control causal experiments
  – We cannot divide a large number of aircrafts at random into two groups, with one group potentially training to land with high speed, low visibility and generally unsafe operation environment of the sort that we suspect may be associated with go-arounds

• Prospective causal effect model
  – We cannot track flights with preexisting go-around operations and wait to see any emerging level of difference of the effect in the treatment group and control group