Models for Estimating Monthly Delays and Cancellations in the NAS

Avijit Mukherjee
Michael Ball
Bargava Subramanian

University of Maryland
Introduction

• Objectives
  – Develop a metric that indicates the level of congestion faced by each operation (arrivals/departures) at an airport.
  – Estimate various percentiles of the distribution of the congestion metric across all operations in the NAS.
  – Model NAS average flight delay and cancellation probability as a function of the congestion level (i.e. the percentiles of the distribution of the congestion metric).

• Application in NAS Strategy Simulator
  – Estimating flight delays and cancellation under various demand growth or capacity increase scenarios in the NAS
  – Estimating passenger delay
Measure of Congestion

The demand/capacity ratio is a good measure of congestion. Each operation at an airport faces a congestion level depending on the capacity and demand at the airport in a time window at which the operation takes place.

**Concept of Rho**

\[ \rho = \frac{\text{# operations during } I \text{ at O's airport}}{\text{airport capacity during } I \text{ at O's airport}} \]

**Percentiles of Rho**

- \( \rho_{50} \): 50th percentile
- \( \rho_{95} \): 95th percentile
- \( \rho_{99} \): 99th percentile

\[ Y = \% \text{ of operations with } \rho \leq X \]
Estimating Rho at an Airport

- Input
  - Hourly scheduled demand
  - GA demand
  - VMC / IMC capacities

- Associate demand/capacity in an hour as the Rho value for each operation during that hour

- Estimate the monthly distribution (or percentiles) of Rho at an airport

![Graph showing Rho distribution at ORD and MSP airports with average delays and connection probabilities.](image-url)
• Estimate Rho50 (50th percentile) and Rho95 (95th percentile) for all airports

• NAS Rho50 and Rho95 are weighted average of corresponding airport Rho’s

• Weights are proportional to the fraction of NAS operations at each airport
Impact of Rho95 on Delays and Cancellation

![Graphs showing the impact of Rho95 on average delay and probability of cancellation. The graphs are color-coded by year: 2000 (red), 2001 (blue), 2002 (yellow), 2003 (black), and 2004 (green).]
Estimating Flight Delays and Cancellation Probability

- Hourly scheduled demand and capacity at 35 major airports obtained from ASPM database

- Monthly load factor data obtained from Database Products Inc.

- Average flight delay and proportion of flights cancelled obtained from ASPM
Monthly Cancellation Model

\[ y = -3.3432x - 3.746 \]

\[ R^2 = 0.6132 \]

Log (Load Factor * (1 - Rho50))

Log (P_Canc)

September 2001 considered outlier, hence excluded
Monthly Delay Model

\[ y = 38.62x - 23.84 \]

\[ R^2 = 0.6862 \]
Functional Relationships

Canc. Probability = $e^{-3.75} \times [\text{Loadfactor} \times (1 - \text{Rho50})]^{-3.34}$

Average Delay = 38.62 * [ Rho95 ( 1 – Canc. Prob.) ] – 23.84
Predicted vs. Observed Prob. Of Cancellation in 2005
Predicted vs. Observed Delay in 2005

Graph showing the comparison between predicted and actual delays in minutes from January to December 2005.
Applications
Airport Categorization Based on Rho

**Objective**: To cluster airports using monthly Rho 50, Rho 95, Average Delays, Probability of Cancellation.

**Method used**: Clustering is carried out using K-Means clustering procedure. This is an iterative procedure wherein airports are assigned to various clusters so as to minimize the distance from the clusters’ centroids.

**Result**: The airports can be divided into four categories:

**Category 1**: Low Rho50, Low Delay, Low Cancellation probability

**Category 2**: Rho50, Delay, Cancellation probability similar to NAS average

**Category 3**: “Constrained” airport-set – delay similar to NAS average but cancellations increase beyond a particular Rho50.

**Category 4**: High Rho, High Delay, High Cancellation Probability
Airport Categories

![Graph showing the relationship between Rho 50 and % of Cancellations for different categories of airports. The graph includes lines for Category 1, Category 2, Category 3, and Category 4, each represented by different markers and colors. The x-axis represents Rho 50 ranging from 0 to 1, and the y-axis represents % of Cancellations ranging from 0 to 7.]
Airport Categories continued

Category 1
- TPA
- MCO
- SLC
- BWI
- PDX

Category 2
- PIT
- MIA
- CLT
- FLL
- LAS
- MDW
- CVG
- SAN
- PHX
- SEA
- IAH
- MSP
- STL

Category 3
- DEN
- IAD
- DFW
- DFW
- DCA
- SFO
- JFK
- LAX

Category 4
- PHL
- BOS
- EWR
- ATL
- ORD
- LGA
Application: Estimating Passenger Delay

\[ P_{\text{miss}} = \text{Prob}\{\text{delay} > \text{layover time} | \text{delay} > 15 \text{ minutes}\} \]
Flight Delay Distribution of Delayed Flights

From Individual flight data, construct histograms of delayed flights in each time interval for each month. Need to determine distribution of flight delays.

Two approaches to achieve the objective:

1) Regress each time interval to have a functional form in terms of average delay and cancellation probability.
2) Fit a smoothing spline to the time interval.
3) Normalize it to obtain distribution of flight delays.

1) Fit one family of distribution for each month
2) Regress shape parameters to obtain distribution of flight delays as a functional form of average delay and cancellation probability
Comparison of two methods

• Data calibrated over 48 months ranging from Jan’00 to Dec ’04

• 10 months used for validation – 5/00, 10/00, 7/01, 4/02, 10/02, 12/03, 7/03, 3/04, 8/04, 11/04

• Both the curves fit well.

• But Bi-weibull distribution is marginally better than bezier curve because the rate of descent initially (15-30 and 30-45 min) in bi-weibull matches more closely to the actual than bezier curve.
Work in Progress

- Introducing a metric (similar to Rho) for enroute congestion

- Evaluate the impact of demand growth at various airport categories on NAS Rho ⇒ Delays and Cancellation