Initial Analysis of and Predictive Model Development for Weather Reroute Advisory Use

Heather Arneson

Aviation Systems Division
NASA Ames Research Center
Moffett Field, CA 94035


**Previous work and ongoing**

- Focused on identifying similar weather days
- Analyzing reroutes used on similar days
- Difficult to generate meaningful clusters of days
Approaches

**Previous work and ongoing**
- Focused on identifying similar weather days
- Analyzing reroutes used on similar days
- Difficult to generate meaningful clusters of days

**This work**
- Build models to predict the use of reroutes based on weather data
Objective
Develop a framework and process to analyze the use of reroutes and develop models to predict reroute use.
**Objective**
Develop a framework and process to analyze the use of reroutes and develop models to predict reroute use.

**Challenges**
**Objective**
Develop a framework and process to analyze the use of reroutes and develop models to predict reroute use.

**Challenges**
- Large amount of weather data available ⇒ difficult to extract relevant features
- Flexibility in route selection and descriptions ⇒ spatially similar routes with different descriptions
- Routes used infrequently ⇒ difficult to find similarities
**Objective**
Develop a framework and process to analyze the use of reroutes and develop models to predict reroute use.

**Challenges**
- Large amount of weather data available ⇒ difficult to extract relevant features
- Flexibility in route selection and descriptions ⇒ spatially similar routes with different descriptions
**Objective**
Develop a framework and process to analyze the use of reroutes and develop models to predict reroute use.

**Challenges**
- Large amount of weather data available ⇒ difficult to extract relevant features
- Flexibility in route selection and descriptions ⇒ spatially similar routes with different descriptions
- Routes used infrequently ⇒ difficult to find similarities
• Advisory details
• Methodology
  • Identification of routes used by flights
  • Identification of similar routes
  • Weather feature extraction
  • Development of predictive models
• Prediction results
• Concluding remarks
- Advisory details
  - Methodology
    - Identification of routes used by flights
    - Identification of similar routes
    - Weather feature extraction
    - Development of predictive models
  - Prediction results
  - Concluding remarks
Advisories consist of . . .

- Name
- Valid time range
- Text description of several routes
  - From an origin Center or airport
  - To a destination airport
Defining advisories

**Advisories consist of . . .**

- Name
- Valid time range
- Text description of several routes
  - From an origin Center or airport
  - To a destination airport

**June to August 2011**

- 1,669 reroute advisories issued
- 735 unique advisory names
- 34,247 routes
- 2,770 origin-destination pairs
# ATCSCC Advisory

**ATCSCC ADVZY 062 DCC 06/21/2011 ROUTE RQD /FL**

**RAW TEXT:**

ATCSCC ADVZY 062 DCC 06/21/11 ROUTE RQD /FL  
NAME: TX_ZME_2_EWR_LGA  
CONSTRAINED AREA: ZME  
REASON: WEATHER  
INCLUDE TRAFFIC: ZFW/ZHU/ZME DEPARTURES TO EWR/LGA  
FACILITIES INCLUDED: /ZDC/ZFW/ZHU/ZID/ZME/ZNY/ZOB/ZTL  
FLIGHT STATUS: ALL FLIGHTS  
VALID: ETD 211800 TO 220100  
PROBABILITY OF EXTENSION: LOW  
REMARKS: THIS REPLACES ADVZY033.  
ASSOCIATED RESTRICTIONS:  
MODIFICATIONS:  
ROUTES:

<table>
<thead>
<tr>
<th>ORIG</th>
<th>DEST</th>
<th>ROUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZHU</td>
<td>LGA</td>
<td>&gt;HRV J37 MGM AHN J208 HPW J191 PXT KORRY3&lt;</td>
</tr>
<tr>
<td>ZHU</td>
<td>EWR</td>
<td>&gt;HRV J37 SPA J14 CREWE J51 FAK PHLBO2&lt;</td>
</tr>
<tr>
<td>ZME ZFW(--BNA)</td>
<td>LGA</td>
<td>&gt;MEM J29 DBJ CXR J146 ETG MIP3&lt;</td>
</tr>
<tr>
<td>ZME ZFW(--BNA)</td>
<td>EWR</td>
<td>&gt;MEM J29 DORET J584 FQM FQM1&lt;</td>
</tr>
</tbody>
</table>

**TMI ID:** RRDCC062  
211728-220100  
11/06/21 17:28 DCCOPS./nfs/lxstn18
**ATCSCC Advisory**

**ATCSCC ADVZY 062 DCC 06/21/2011 ROUTE RQD /FL**

**RAW TEXT:**

- **NAME:** TX_ZME_2_EWR_LGA
- **CONSTRAINED AREA:** ZME
- **REASON:** WEATHER
- **INCLUDE TRAFFIC:** ZFW/ZHU/ZME DEPARTURES TO EWR/LGA
- **FACILITIES INCLUDED:** /ZDC/ZFW/ZHU/ZID/ZME/ZNY/ZOB/ZTL
- **FLIGHT STATUS:** ALL FLIGHTS
- **VALID:** ETD 211800 TO 220100
- **PROBABILITY OF EXTENSION:** LOW
- **REMARKS:** THIS REPLACES ADVZY033.
- **ASSOCIATED RESTRICTIONS:**
- **MODIFICATIONS:**
- **ROUTES:**

<table>
<thead>
<tr>
<th>ORIG</th>
<th>DEST</th>
<th>ROUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZHU</td>
<td>LGA</td>
<td>&gt;HRV J37 MGM AHN J208 HPW J191 PXT KORRY3&lt;</td>
</tr>
<tr>
<td>ZHU</td>
<td>EWR</td>
<td>&gt;HRV J37 SPA J14 CREWE J51 FAK PHLBO2&lt;</td>
</tr>
<tr>
<td>ZME ZFW(-BNA)</td>
<td>LGA</td>
<td>&gt;MEM J29 DJB CXR J146 ETG MIP3&lt;</td>
</tr>
<tr>
<td>ZME ZFW(-BNA)</td>
<td>EWR</td>
<td>&gt;MEM J29 DORET J584 FQM FQM1&lt;</td>
</tr>
</tbody>
</table>

**TMI ID:** RRDCC062
211728-220100
11/06/21 17:28 DCCOPS./nfs/lxstn18
**ATCSCC Advisory**

**ATCSCC ADVZY 062 DCC 06/21/2011 ROUTE RQD /FL**

**RAW TEXT:**

ATCSCC ADVZY 062 DCC 06/21/11 ROUTE RQD /FL  
NAME: TX_ZME_2_EWR_LGA  
CONSTRAINED AREA: ZME  
REASON: WEATHER  
INCLUDE TRAFFIC: ZFW/ZHU/ZME DEPARTURES TO EWR/LGA  
FACILITIES INCLUDED: /ZDL/ZFW/ZHU/ZID/ZME/ZNY/ZOB/ZTL  
FLIGHT STATUS: ALL FLIGHTS  
VALID: ETD 211800 TO 220100  
PROBABILITY OF EXTENSION: LOW  
REMARKS: THIS REPLACES ADVZY033.  
ASSOCIATED RESTRICTIONS:  
MODIFICATIONS:  

<table>
<thead>
<tr>
<th>ORIG</th>
<th>DEST</th>
<th>ROUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZHU</td>
<td>LGA</td>
<td>HRV J37 MGM AHN J208 HPW J191 PXT KORRY3&lt;</td>
</tr>
<tr>
<td>ZHU</td>
<td>EWR</td>
<td>HRV J37 SPA J14 CREWE J51 FAK PHLBO2&lt;</td>
</tr>
<tr>
<td>ZME ZFW(-BNA)</td>
<td>LGA</td>
<td>MEM J29 DJB CXR J146 ETG MIP3&lt;</td>
</tr>
<tr>
<td>ZME ZFW(-BNA)</td>
<td>EWR</td>
<td>MEM J29 DORET J584 FQM FQM1&lt;</td>
</tr>
</tbody>
</table>

TMI ID: RRDCC062  
211728-220100  
11/06/21 17:28 DCCOPS./nfs/lxstn18
### ATCSCC Advisory

**ATCSCC ADVZY 062 DCC 06/21/2011 ROUTE RQD /FL**

**RAW TEXT:**

```
ATCSCC ADVZY 062 DCC 06/21/11 ROUTE RQD /FL
NAME: TX_ZME_2_EWR_LGA
CONstrained AREA: ZME
REASON: WEATHER
INCLUDE TRAFFIC: ZFW/ZHU/ZME DEPARTURES TO EWR/LGA
FACILITIES INCLUDED: /ZDC/ZFW/ZHU/ZID/ZME/ZNY/ZOB/ZTL
FLIGHT STATUS: ALL_FLIGHTS
VALID: ETD 211800 TO 220100
PROBABILITY OF EXTENSION: LOW
REMARKS: THIS REPLACES ADVZY033.
ASSOCIATED RESTRICTIONS:
MODIFICATIONS:
ROUTES:

<table>
<thead>
<tr>
<th>ORIG</th>
<th>DEST</th>
<th>ROUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZHU</td>
<td>LGA</td>
<td>&gt;HRV J37 MGM AHN J208 HPW J191 PXT KORRY3&lt;</td>
</tr>
<tr>
<td>ZHU</td>
<td>EWR</td>
<td>&gt;HRV J37 SPA J14 CREWE J51 FAK PHLBO2&lt;</td>
</tr>
<tr>
<td>ZME ZFW(−BNA)</td>
<td>LGA</td>
<td>&gt;MEM J29 DJB CXR J146 ETG MIP3&lt;</td>
</tr>
<tr>
<td>ZME ZFW(−BNA)</td>
<td>EWR</td>
<td>&gt;MEM J29 DORET J584 FQM FQM1&lt;</td>
</tr>
</tbody>
</table>
```

**TMI ID:** RRDCC062
211728-220100
11/06/21 17:28 DCCOPS./nfs/lxstn18
Example advisory
Outline

- Advisory details
- **Methodology**
  - Identification of routes used by flights
  - Identification of similar routes
  - Weather feature extraction
  - Development of predictive models
- Prediction results
- Concluding remarks
Methodology

• Identification of routes used by flights
• Identification of similar routes
• Weather feature extraction
• Development of predictive models
Methodology

- Identification of routes used by flights requires distance metric to compare routes and flight tracks
- Identification of similar routes
- Weather feature extraction
- Development of predictive models
Methodology

- Identification of routes used by flights requires distance metric to compare routes and flight tracks
- Identification of similar routes requires distance metric to compare routes
- Weather feature extraction
- Development of predictive models
• Identification of routes used by flights requires distance metric to compare routes and flight tracks
• Identification of similar routes requires distance metric to compare routes
• Weather feature extraction requires domain knowledge
• Development of predictive models
• Advisory details
• Methodology
  • Identification of routes used by flights
  • Identification of similar routes
  • Weather feature extraction
  • Development of predictive models
• Prediction results
• Concluding remarks
Distance metric

\[
distance(path A, path B) = 1 - \frac{\text{length}(\text{grid overlap})}{\min(\text{length}(path A), \text{length}(path B))}
\]
Distance metric

\[ \text{distance}(\text{path A}, \text{path B}) = 1 - \frac{\text{length(overlap)}}{\text{min(length(path A), length(path B))}} \]
Distance metric

\[
\text{distance}(\text{path A}, \text{path B}) = 1 - \frac{\text{length(grid overlap)}}{\text{min(length(path A), length(path B))}}
\]
Distance metric

\[ \text{distance}(\text{path A}, \text{path B}) = 1 - \text{length(\text{grid overlap})} \]

\[ \text{min} (\text{length(\text{path A})}, \text{length(\text{path B})}) \]
Distance metric

\[ \text{distance}(\text{path A, path B}) = 1 - \frac{\text{length(grid overlap)}}{\min(\text{length(path A), length(path B)})} \]
• June through August 2011
• Routes and flights inbound to New York Center (ZNY)
• Define use:
  flight track and reroute overlap for at least 85% of shorter path
• Of 4,476 issued routes, 905 were used by at least one flight
● Advisory details

● Methodology
  ● Identification of routes used by flights
  ● Identification of similar routes
  ● Weather feature extraction
  ● Development of predictive models

● Prediction results

● Concluding remarks
905 used routes grouped into 253 clusters

Example cluster
Cluster routes

905 used routes grouped into 253 clusters

Example cluster

Origin Center
905 used routes grouped into 253 clusters

Example cluster

Cluster routes
Cluster routes

905 used routes grouped into 253 clusters

Example cluster

Origin Center

Cluster member

Destination
905 used routes grouped into 253 clusters

Example cluster

Cluster centroid

Cluster member

Origin Center

Destination
Outline

• Advisory details

• Methodology
  • Identification of routes used by flights
  • Identification of similar routes
  • Weather feature extraction
  • Development of predictive models

• Prediction results

• Concluding remarks
**Echo tops**

- Estimates of tops of clouds based on radar measurements
- Values are discrete altitude levels
  0 ft to 50,000 ft at 5,000 ft intervals
**Echo tops**

- Estimates of tops of clouds based on radar measurements
- Values are discrete altitude levels
  - 0 ft to 50,000 ft at 5,000 ft intervals
- 108,955 data points cover the continental US
**Echo tops**

- Estimates of tops of clouds based on radar measurements
- Values are discrete altitude levels
  - 0 ft to 50,000 ft at 5,000 ft intervals
- 108,955 data points cover the continental US
- 2,614,920 echo top values per hour
**Echo tops**

- Estimates of tops of clouds based on radar measurements
- Values are discrete altitude levels
  - 0 ft to 50,000 ft at 5,000 ft intervals
- 108,955 data points cover the continental US
- 2,614,920 echo top values per hour

**Grid**

- Spatial resolution of 75 nmi by 58 nmi
  - (1.25° lat by 1.25° lon)
- 1,000 grid elements cover the continental US
- Temporal resolution of one hour
- 1,000 averaged echo top values per hour
High resolution weather data
Lower resolution weather data
Lower resolution weather data

Cluster centroid
Lower resolution weather data

Cluster centroid

Direct route
Cluster centroid

Direct route
● Advisory details
● **Methodology**
  ● Identification of routes used by flights
  ● Identification of similar routes
  ● Weather feature extraction
  ● Development of predictive models
● Prediction results
● Concluding remarks
Reduced data

- June to August 2011
  \[\Rightarrow 2,208 \text{ one-hour time windows}\]
- 905 ZNY-bound routes used
  \[\Rightarrow 253 \text{ reroute clusters}\]
  \[\Rightarrow \text{20 most frequently used clusters}\]
  \[\text{(used 50 to 240 times)}\]
- 2,614,920 echo top data points per hour
  \[\Rightarrow 1,000 \text{ echo top points per hour}\]
  \[\Rightarrow \text{34 created features per hour per cluster}\]
Reduced data

- June to August 2011
  ⇒ **2,208 one-hour time windows**
- 905 ZNY-bound routes used
  ⇒ 253 reroute clusters
  ⇒ **20 most frequently used clusters**
  (used 50 to 240 times)
- 2,614,920 echo top data points per hour
  ⇒ 1,000 echo top points per hour
  ⇒ 34 created features per hour per cluster

Data for model development for one cluster

- 2,208 observations
- 34 created features
- class label
  + reroute cluster used
  – reroute cluster not used
Model performance metrics

**Classification error**

\[ \varepsilon = \frac{\text{# incorrectly predicted observations}}{\text{total # observations}} \]
Model performance metrics

Classification error

\[ \varepsilon = \frac{\text{# incorrectly predicted observations}}{\text{total # observations}} \]

True positive rate

\[ \text{TPR} = \frac{\text{# of correctly predicted positive observations}}{\text{total # of positive observations}} \]
Model performance metrics

**Classification error**

\[
\varepsilon = \frac{\text{# incorrectly predicted observations}}{\text{total # observations}}
\]

**True positive rate**

\[
\text{TPR} = \frac{\text{# of correctly predicted positive observations}}{\text{total # of positive observations}}
\]

**True negative rate**

\[
\text{TNR} = \frac{\text{# of correctly predicted negative observations}}{\text{total # of negative observations}}
\]
Shallow trees cannot capture more complex connections

Deep trees tend to overfit
• Shallow trees cannot capture more complex connections
• Deep trees tend to overfit
Random forest

- Consists of many weak learners (shallow decision trees)
- Each decision tree is built with:
  - Randomly selected subset of observations
  - Randomly selected subset of features
- Ensemble prediction: weighted vote of each weak learner
• Consists of many weak learners (shallow decision trees)
• Each decision tree is built with:
  • Randomly selected subset of observations
  • Randomly selected subset of features
• Ensemble prediction: weighted vote of each weak learner
  ⇒ Advantage: reduce sensitivity to noise ⇒ reduce overfitting
Model development

Observations

\[ \varepsilon = \text{sub test prediction error} \]

\[ \alpha = \begin{cases} \uparrow & \text{as } \varepsilon \downarrow, \varepsilon < 0, \\ 0, & \text{otherwise} \end{cases} \]
Model development

Observations

\[ \varepsilon = \text{sub test prediction error} \]

\[ \alpha = \begin{cases} \uparrow & \text{as } \varepsilon \downarrow, \varepsilon < 0 \\ 0, & \text{otherwise} \end{cases} \]
Observations

ε = test prediction error

α = { \uparrow as ε \downarrow, ε > 0.

0, otherwise}
Model development

Observations

Features

Weak learner

\[ \varepsilon = \text{sub test prediction error} \]

\[ \alpha = \begin{cases} \uparrow & \text{as } \varepsilon \downarrow, \varepsilon < 0.5 \\ 0, & \text{otherwise} \end{cases} \]
Observations

ε = sub test prediction error

α = \{ \uparrow \text{as } \varepsilon \downarrow, \varepsilon < 0 \}, \text{0 otherwise}
Model development

Observations

ε = test prediction error
α = \{ \uparrow \text{as } ε \downarrow, ε < 0 \}, \text{ otherwise}
ε = \text{sub test prediction error}

\alpha = \begin{cases} 
\uparrow & \text{as } \varepsilon \downarrow, \varepsilon < 0, \\
0, & \text{otherwise} 
\end{cases}
ε = sub test prediction error

α = { ↗ as ε ↘, ε < 0, otherwise}
Model development

Observations

Features

Weak learner

\[ \epsilon = \text{sub test prediction error} \]

\[ \alpha = \begin{cases} \uparrow \text{as } \epsilon \downarrow, \epsilon < 0, \\ 0, \text{otherwise} \end{cases} \]
\( \varepsilon = \text{sub test prediction error} \)
\[ \varepsilon = \text{sub test prediction error} \]

\[ \alpha = \begin{cases} 
\uparrow & \text{as } \varepsilon < 0.5 \\
0, & \text{otherwise}
\end{cases} \]
Model development

Weak learner 1  Weak learner 2  ···  Weak learner 100

Observations  Features  Observations  Features  Observations  Features
Model development

Weak learner 1  Weak learner 2  \cdots  Weak learner 100

**Ensemble prediction:**
Weighted vote from each weak learner
**Ensemble prediction:**
Weighted vote from each weak learner
Model development

Weak learner 1  Weak learner 2  ⋮  Weak learner 100

Observations  Features  Observations  Features  Observations  Features

*Ensemble prediction*: Weighted vote from each weak learner
• Advisory details
• Methodology
  • Identification of routes used by flights
  • Identification of similar routes
  • Weather feature extraction
  • Development of predictive models
• Prediction results
• Concluding remarks
Prediction results

Cluster Ordered by Use

TPR
TNR

Rate

TPR
TNR

0
0.2
0.4
0.6
0.8
1

0 5 10 15 20
**Synthetic Minority Oversampling Technique (SMOTE)**

Within the training set:

- Select a positive observation
- Select one of its nearest neighbors
- Create a new observation: Convex combination of these two observations

Prediction results with SMOTE
● Advisory details

● Methodology
  ● Identification of routes used by flights
  ● Identification of similar routes
  ● Weather feature extraction
  ● Development of predictive models

● Prediction results

● Concluding remarks
Conclusions

- Developed a framework to
  - analyze the historical use of reroutes
  - develop models to predict reroute use
- With improvements, this approach could provide insight into advisory use

Future work

- Include weather conditions at fixes and along jet routes
- Use Convective Weather Avoidance Model (CWAM)
- Use Collaborative Convective Forecast Product (CCFP)
Conclusions

• Developed a framework to
  • analyze the historical use of reroutes
  • develop models to predict reroute use
• With improvements, this approach could provide insight into advisory use

Future work

• Include weather conditions at fixes and along jet routes
• Use Convective Weather Avoidance Model (CWAM)
• Use Collaborative Convective Forecast Product (CCFP)
Questions?

Heather Arneson
heather.arneson@nasa.gov