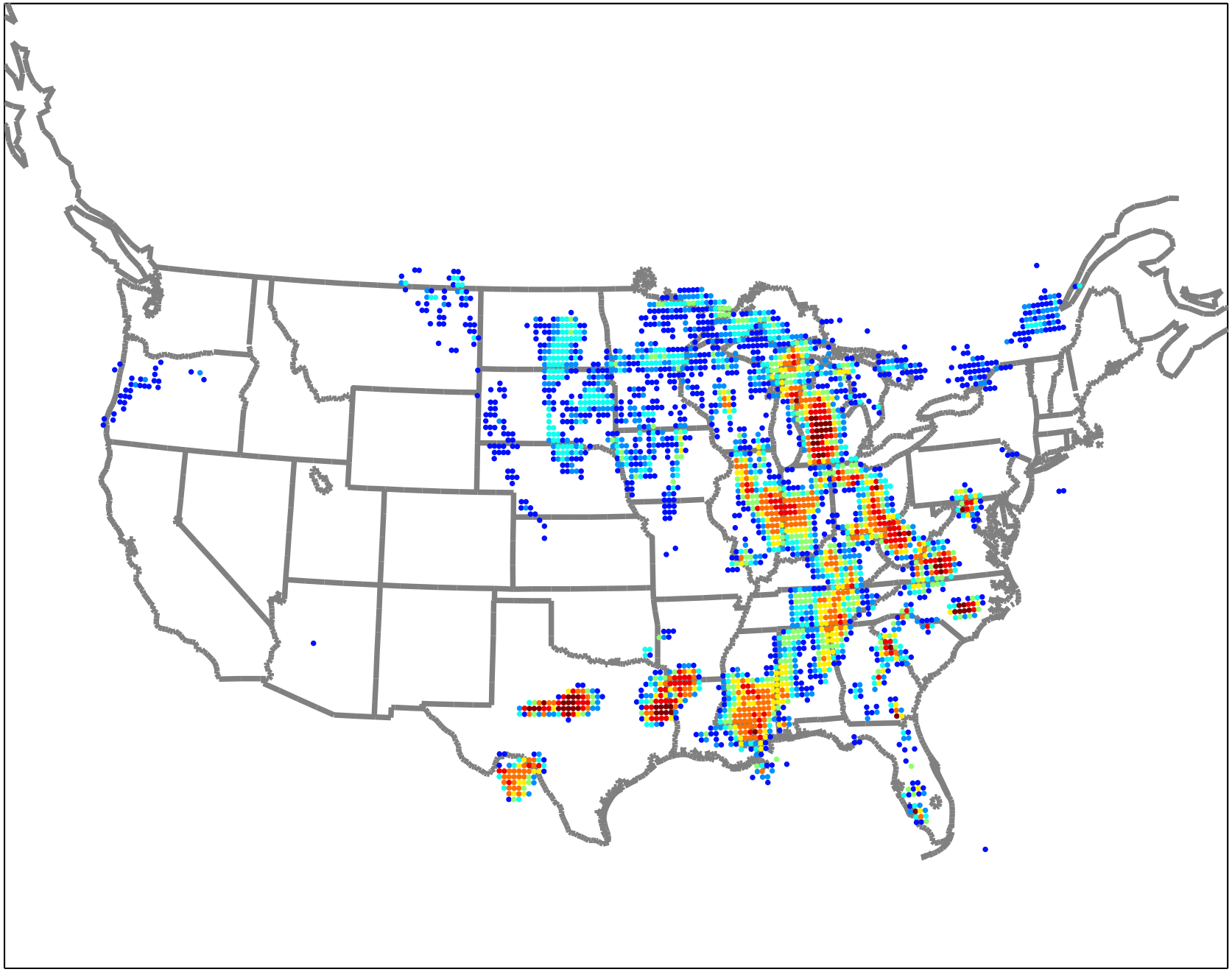
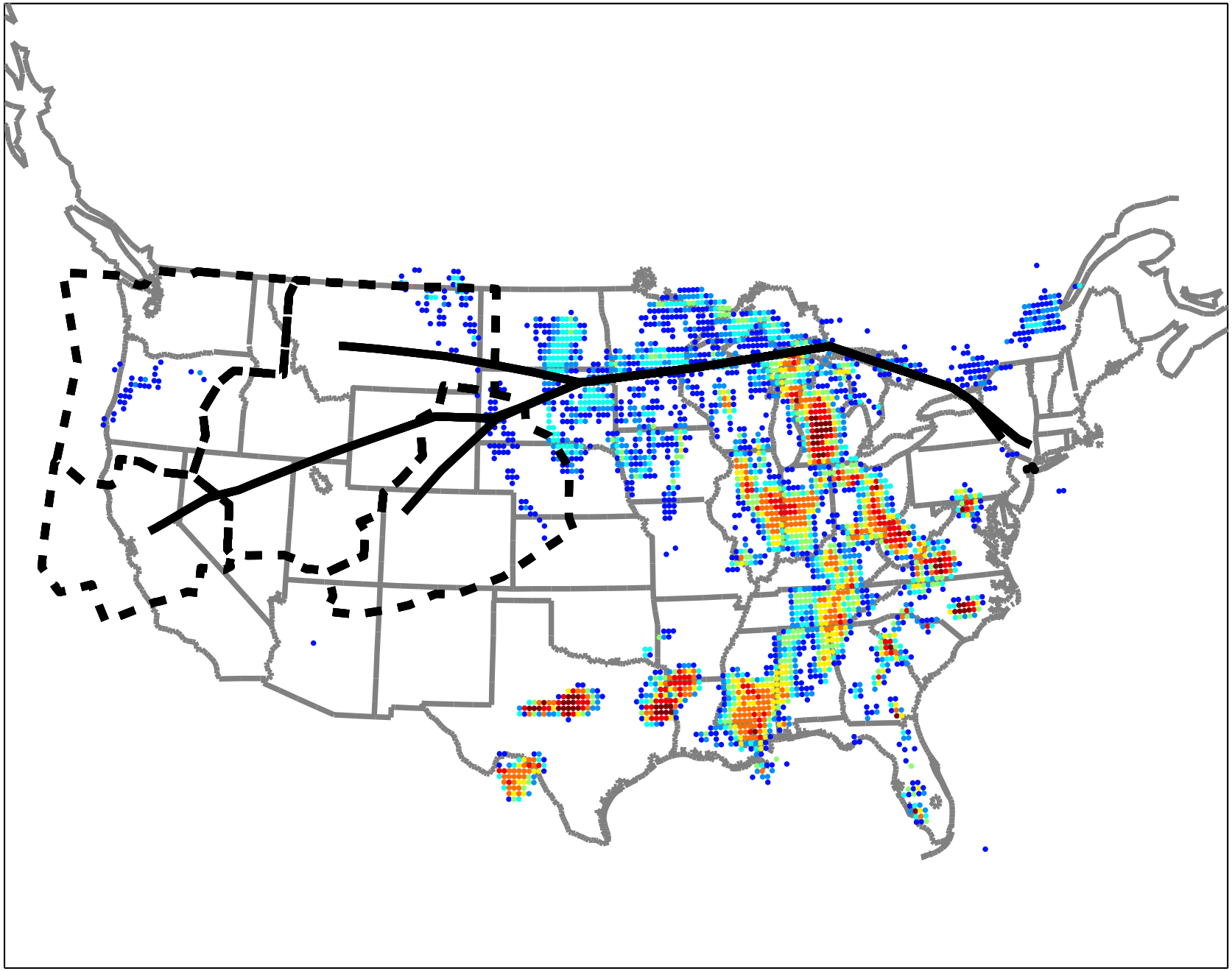
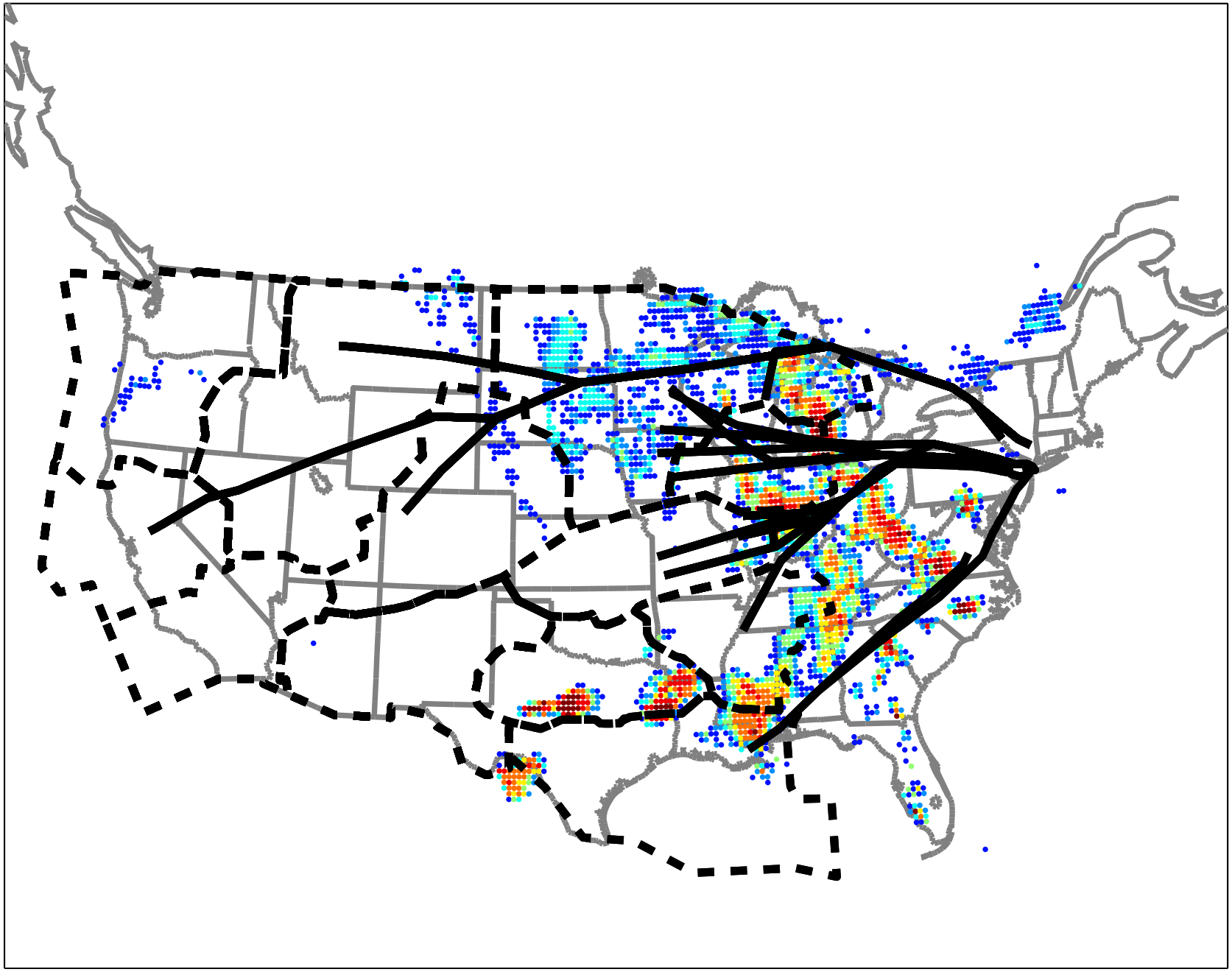
Initial Analysis of and Predictive Model Development for Weather Reroute Advisory Use

Heather Arneson

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







Approaches

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
  $\Rightarrow$ difficult to extract relevant features

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
  $\Rightarrow$ difficult to extract relevant features
- Flexibility in route selection and descriptions
  $\Rightarrow$ 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
  $\Rightarrow$ difficult to extract relevant features
- Flexibility in route selection and descriptions
  $\Rightarrow$ spatially similar routes with different descriptions
- Routes used infrequently
  $\Rightarrow$ difficult to find similarities

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

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

Defining advisories

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

Example advisory

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:

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


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

Example advisory

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:

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


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

Example advisory

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:

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


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

Example advisory

ATCSCC Advisory

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

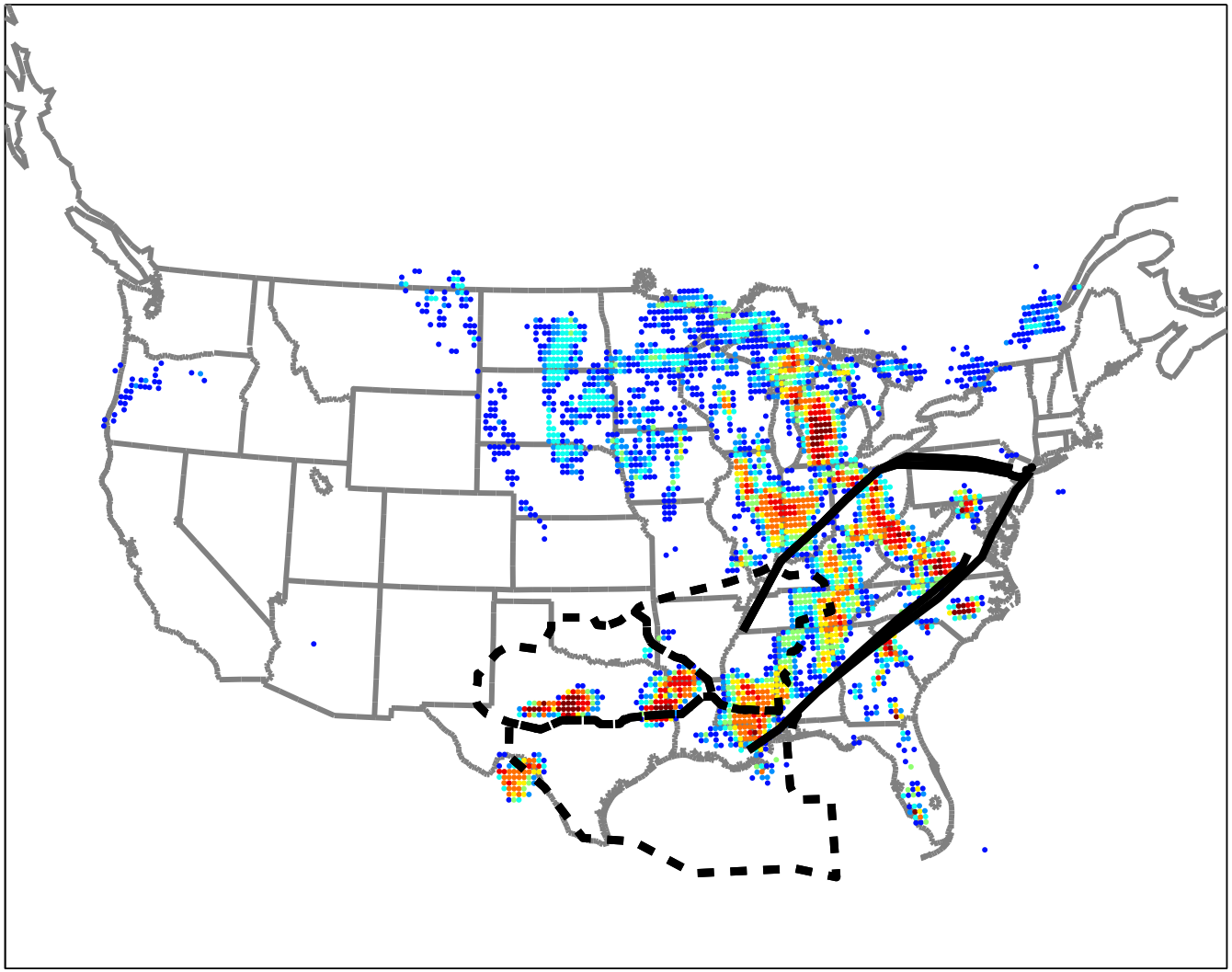
RAW TEXT:

```
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:

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


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

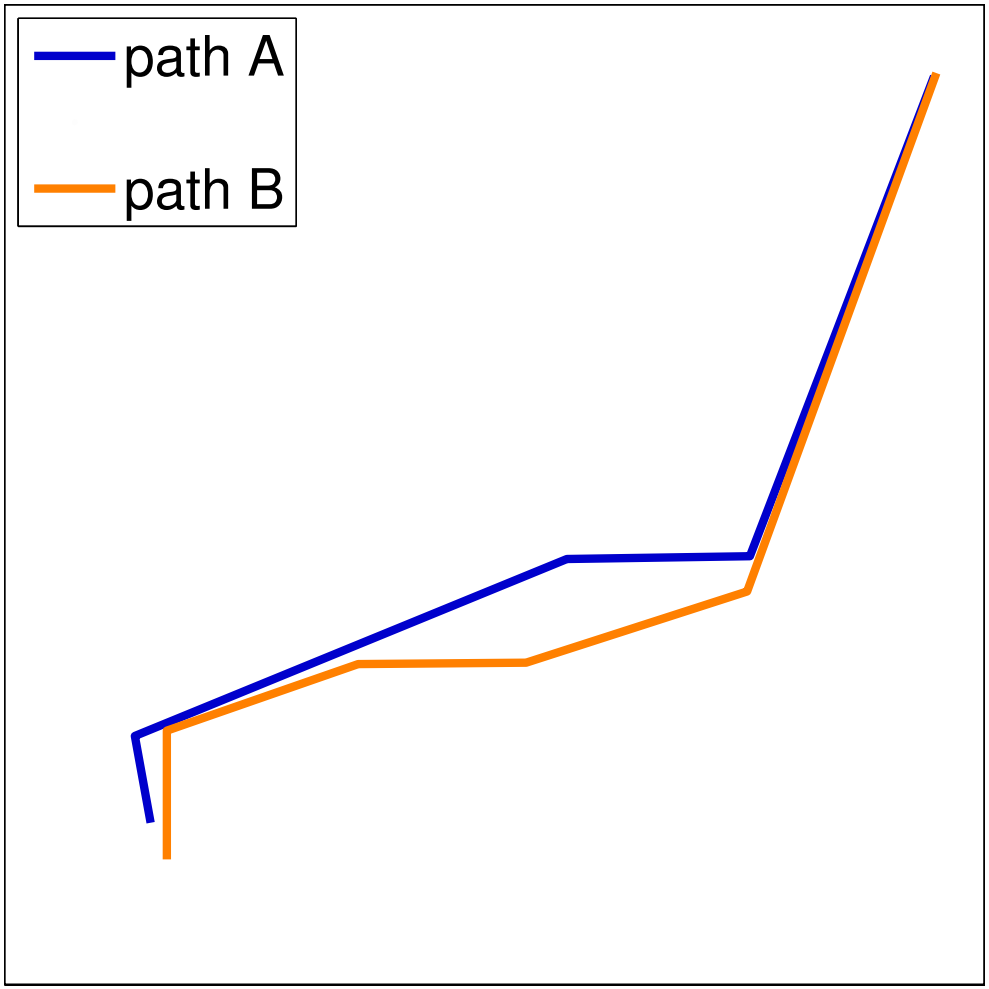
Methodology

- 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
  requires domain knowledge
- Development of predictive models

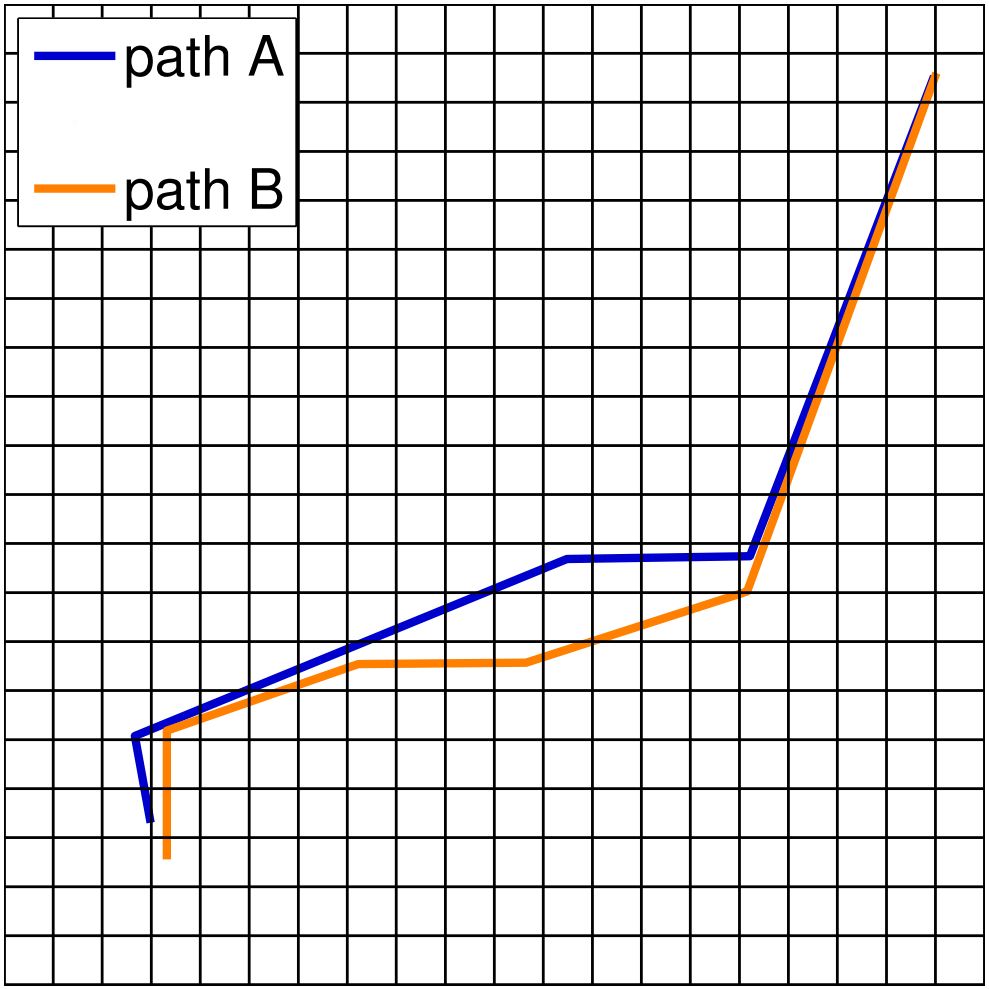
Outline

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

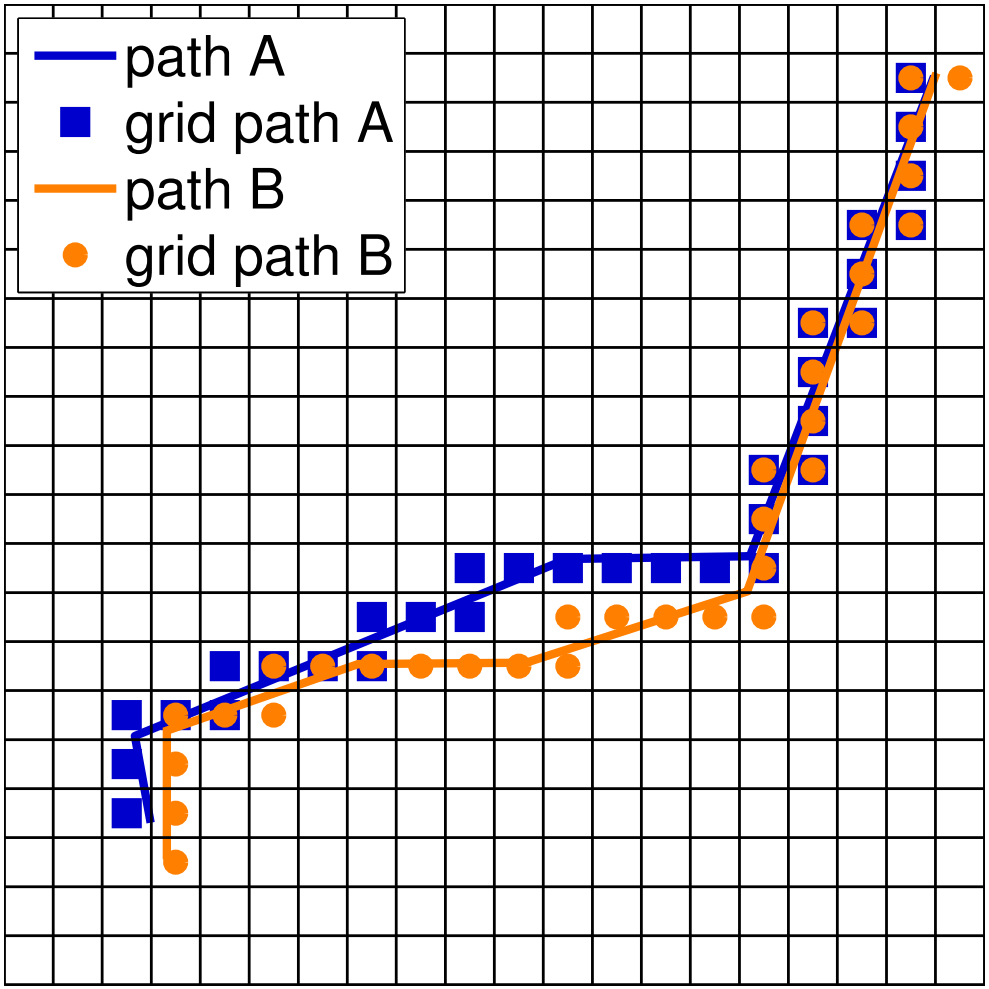
Distance metric



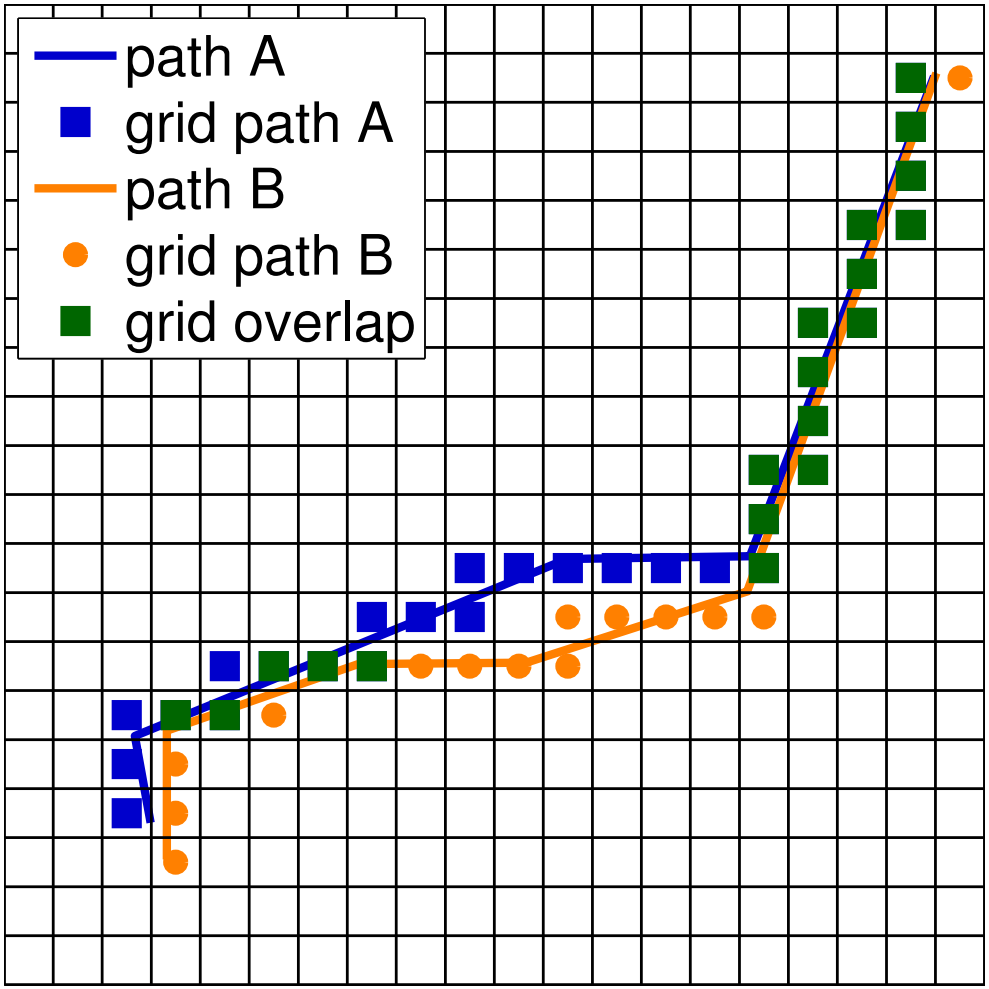
Distance metric



Distance metric

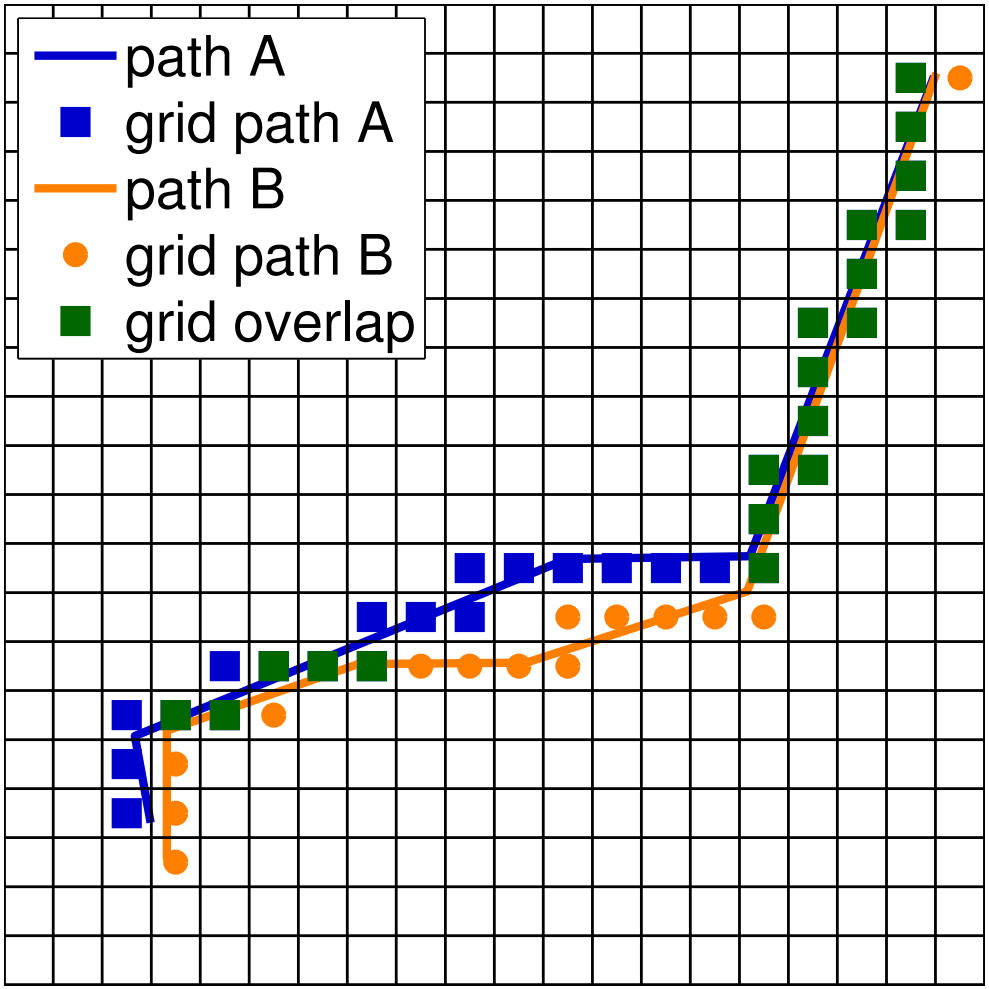


Distance metric



Distance metric

$$\text{distance}(\text{path A}, \text{path B}) = 1 - \frac{\text{length}(\text{grid overlap})}{\min(\text{length}(\text{path A}), \text{length}(\text{path B}))}$$



Route usage

- 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

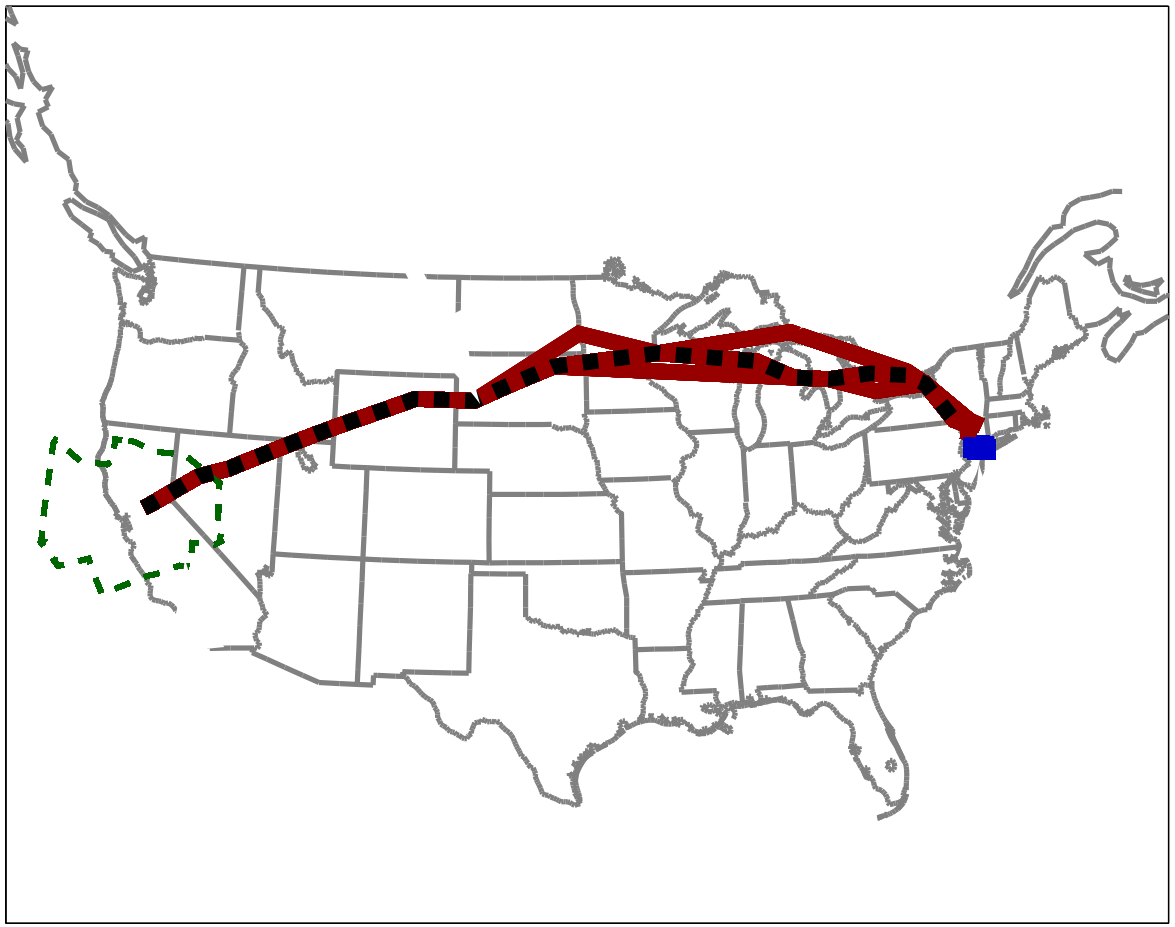
Outline

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

Cluster routes

905 used routes grouped into 253 clusters

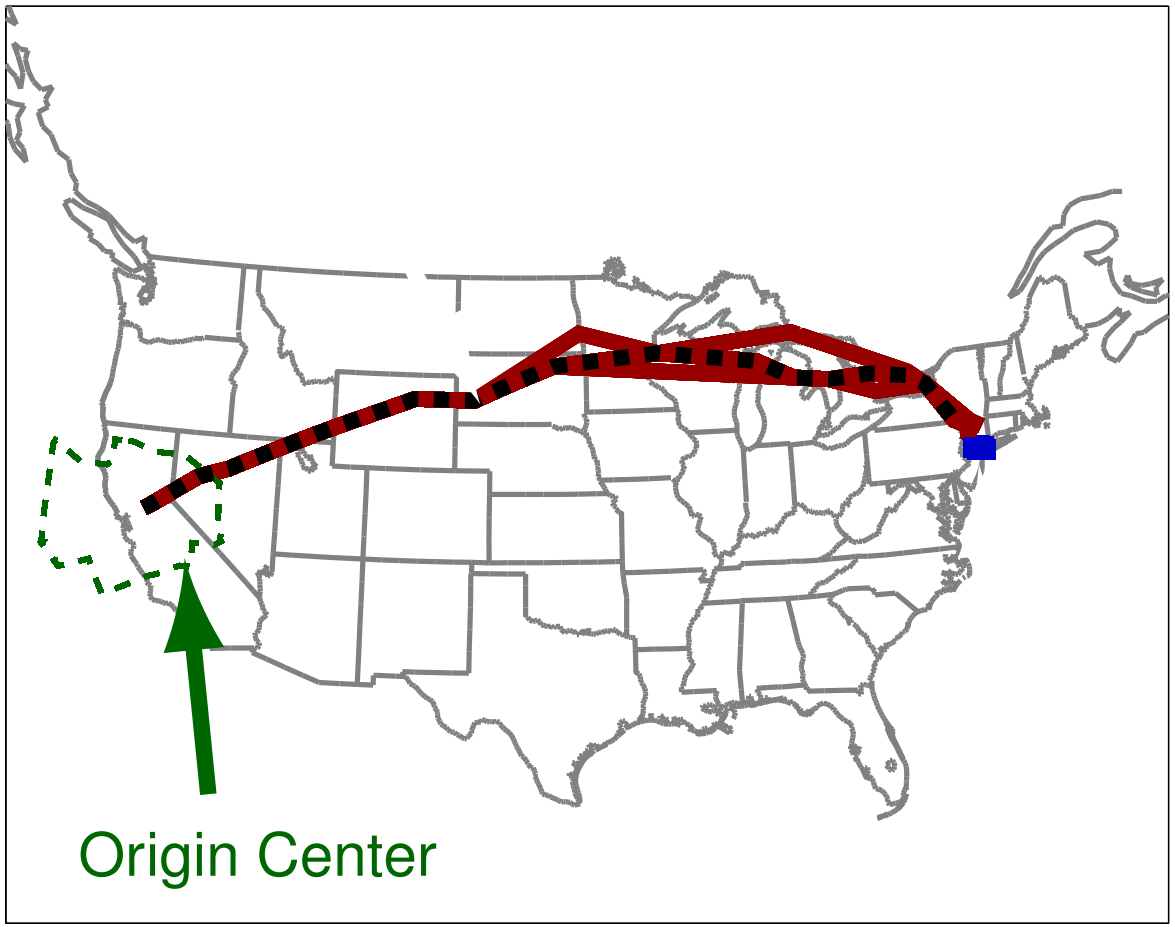
Example cluster



Cluster routes

905 used routes grouped into 253 clusters

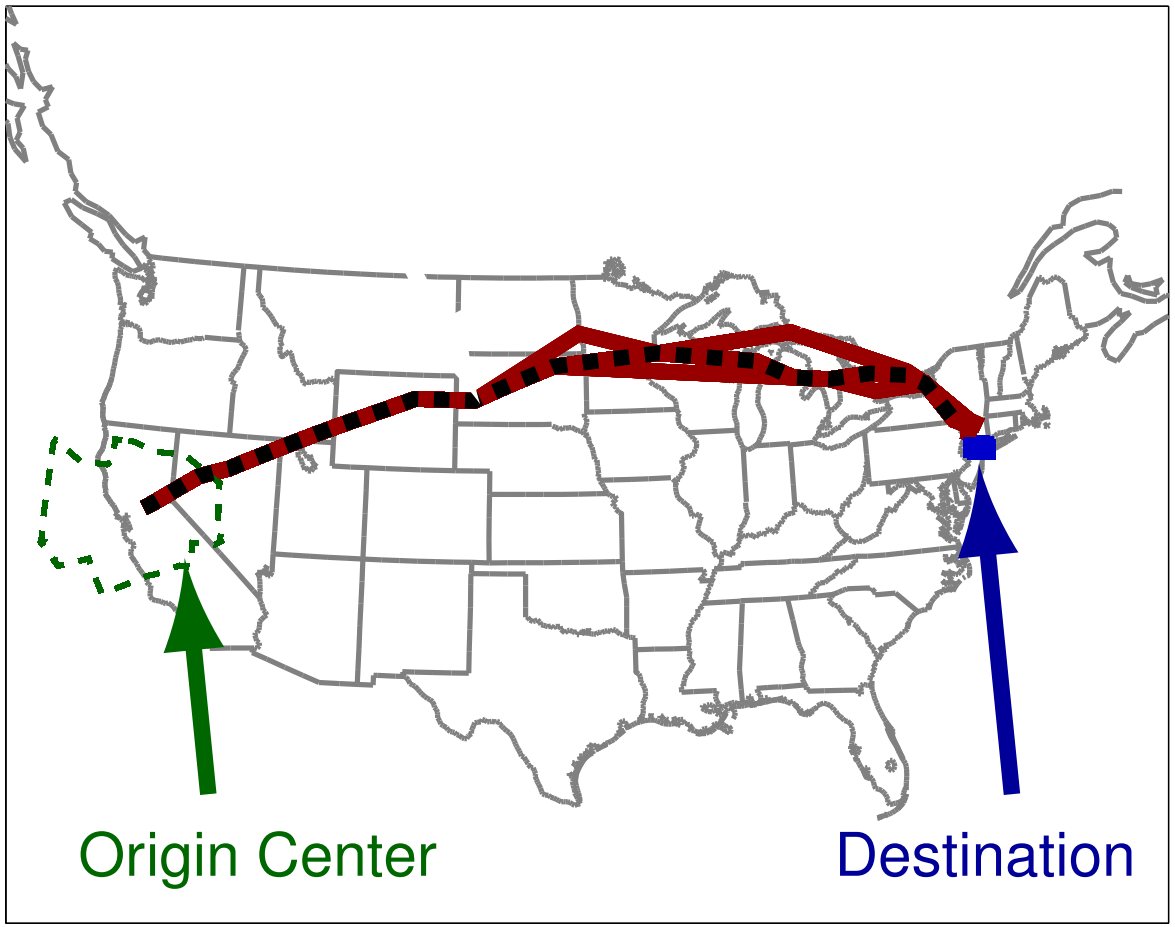
Example cluster



Cluster routes

905 used routes grouped into 253 clusters

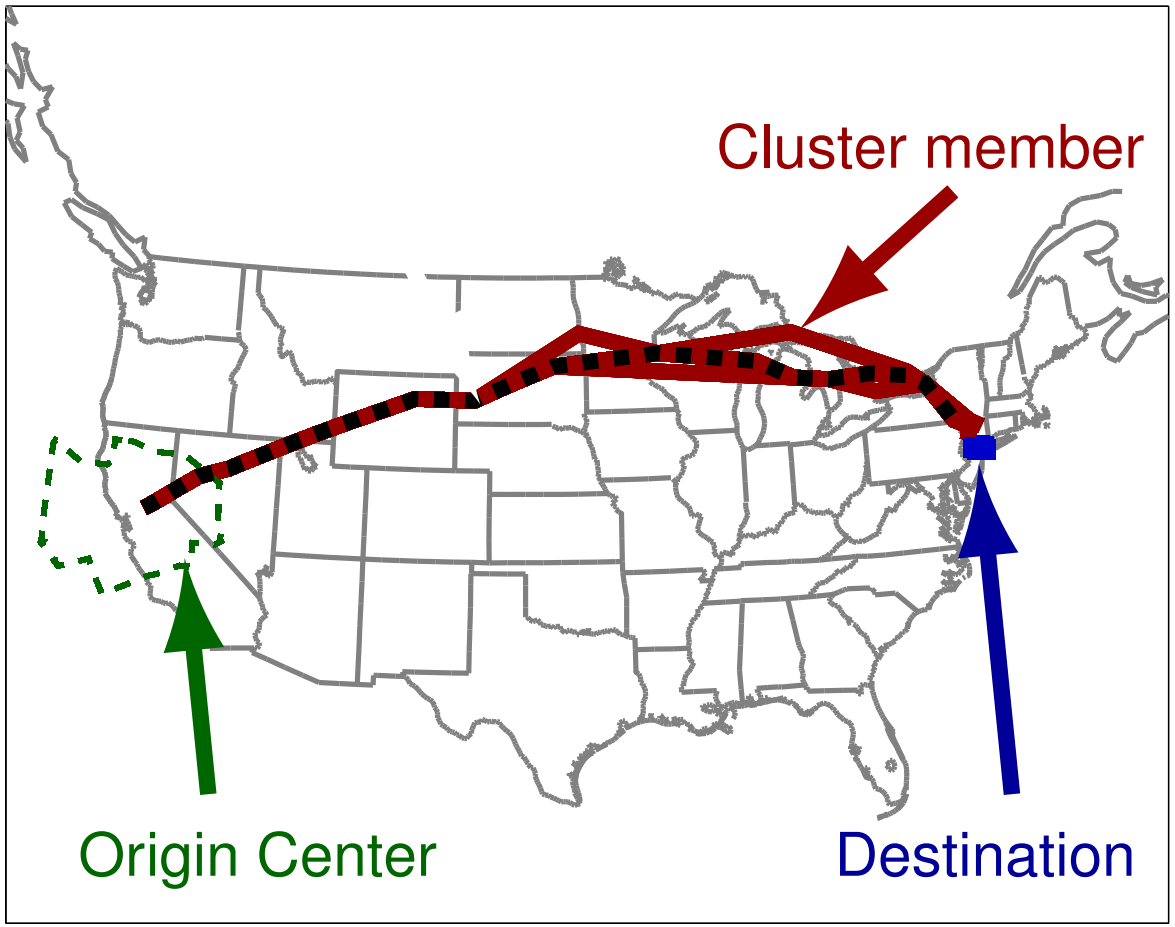
Example cluster



Cluster routes

905 used routes grouped into 253 clusters

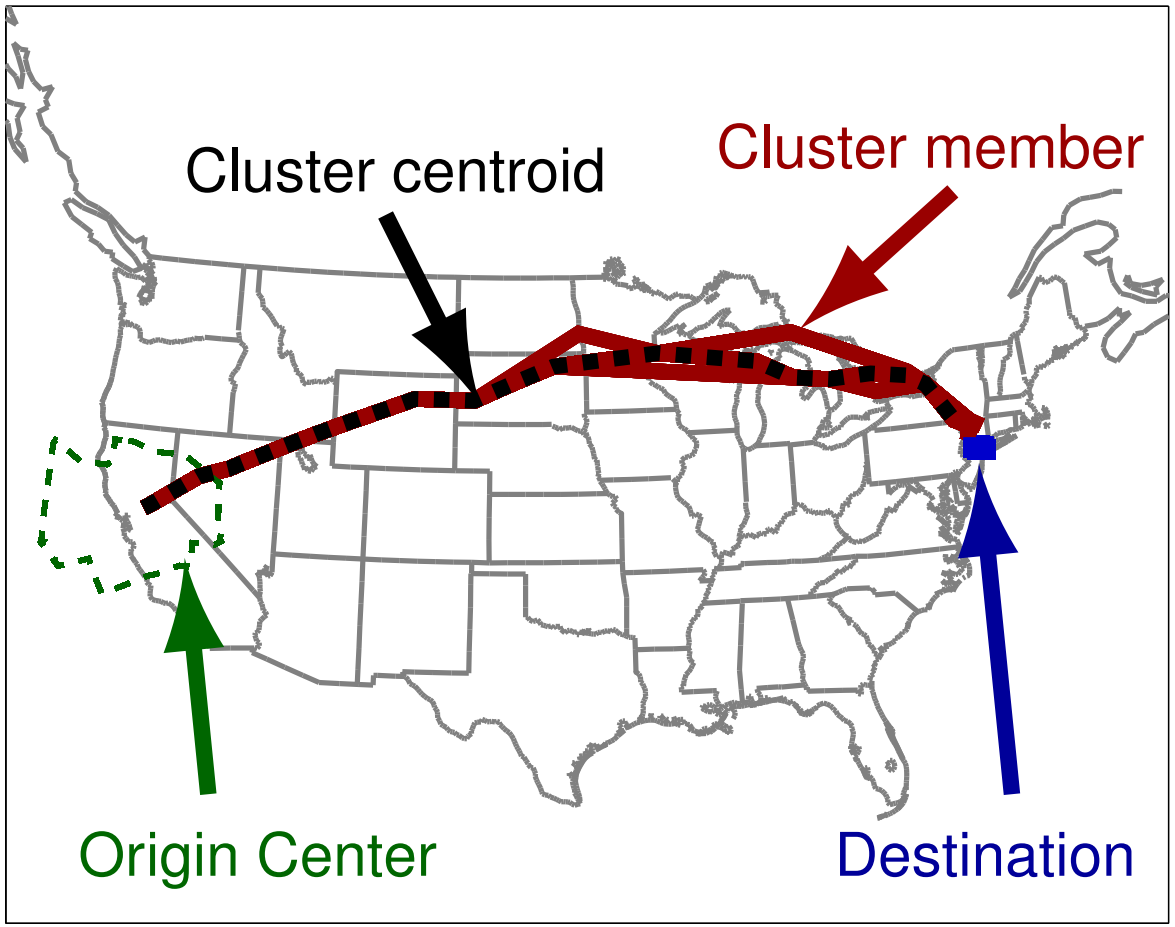
Example cluster



Cluster routes

905 used routes grouped into 253 clusters

Example cluster



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

Convective weather data

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

Convective weather data

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

Convective weather data

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

Convective weather data

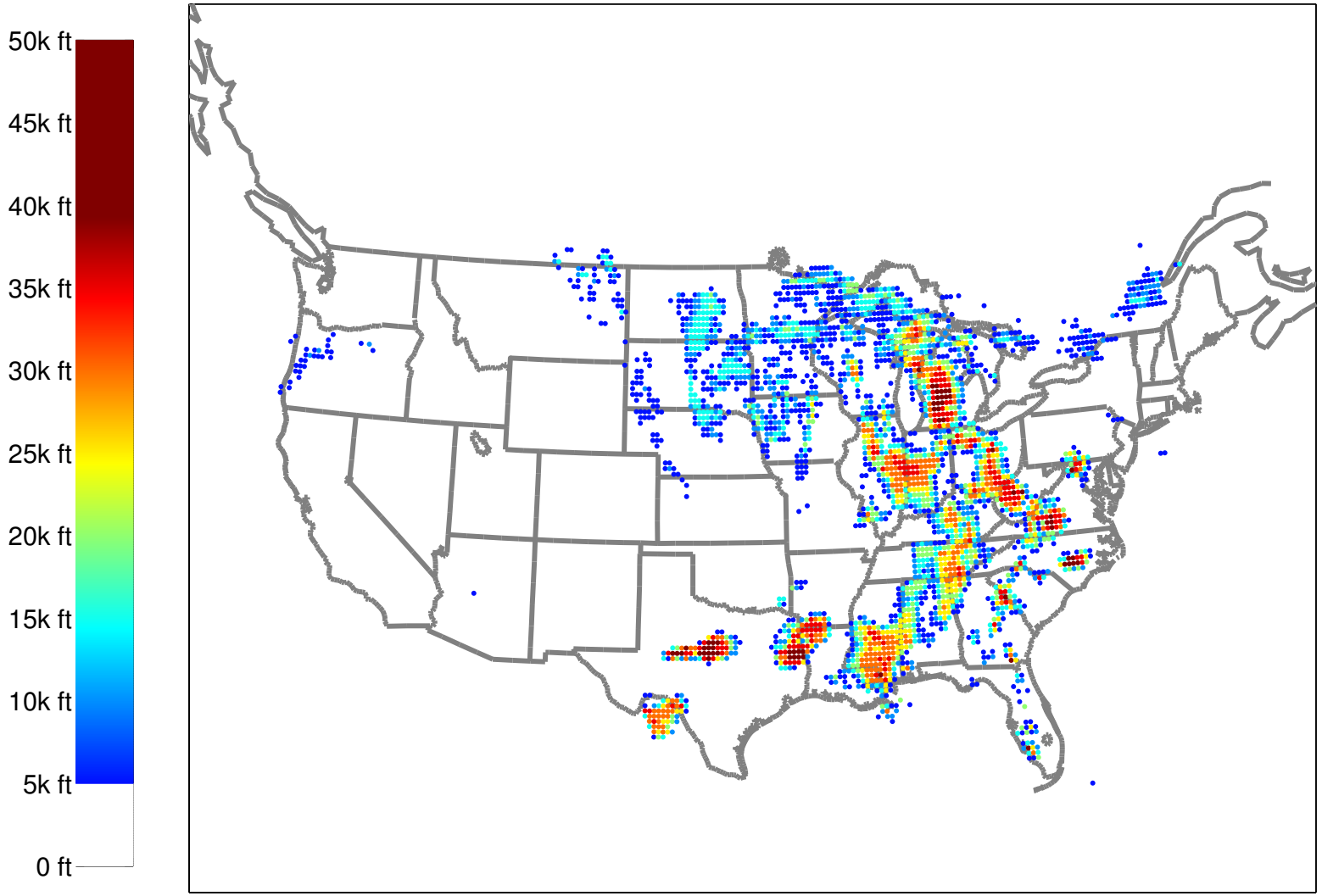
Echo tops

- 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

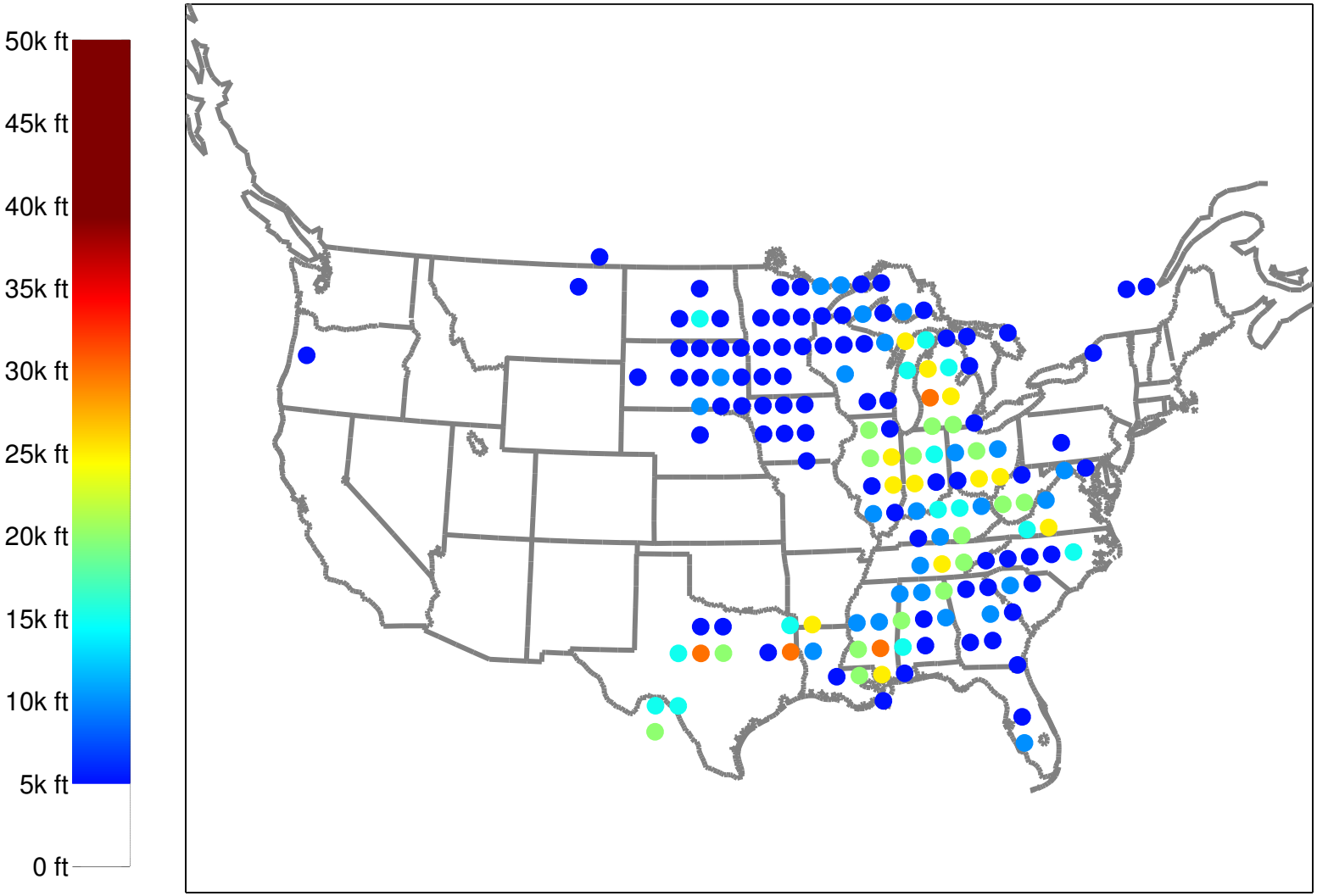
Grid

- 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

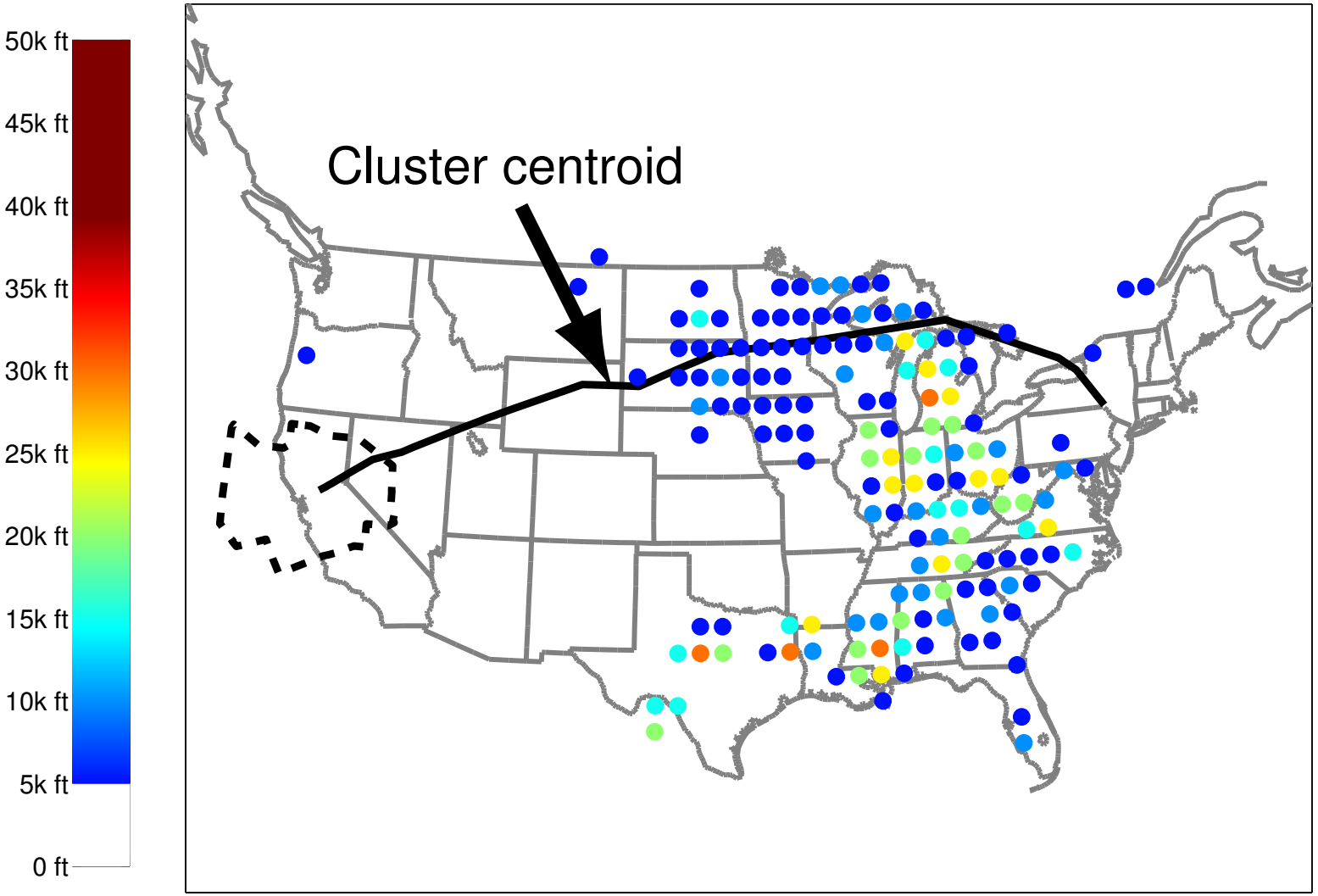
High resolution weather data



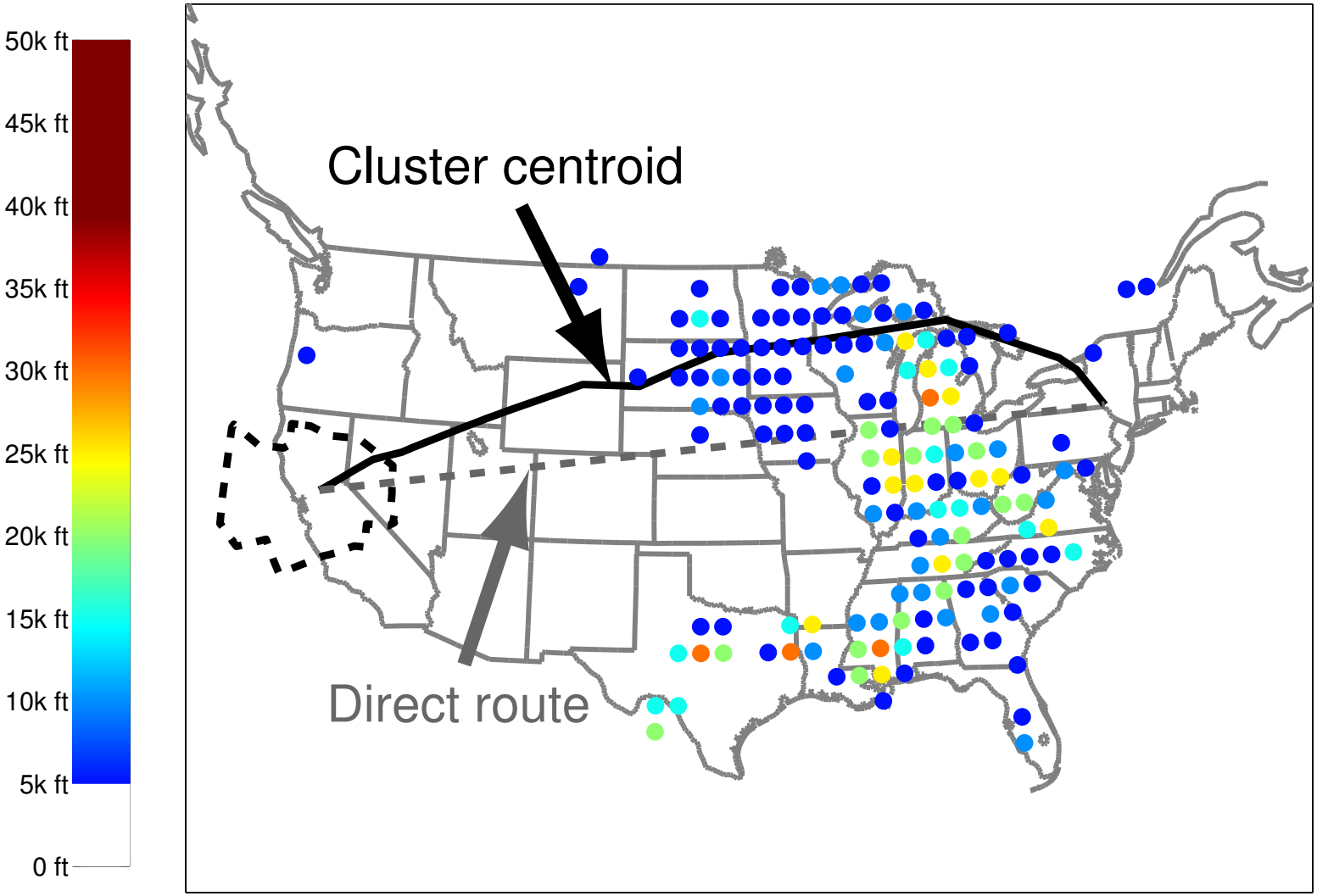
Lower resolution weather data



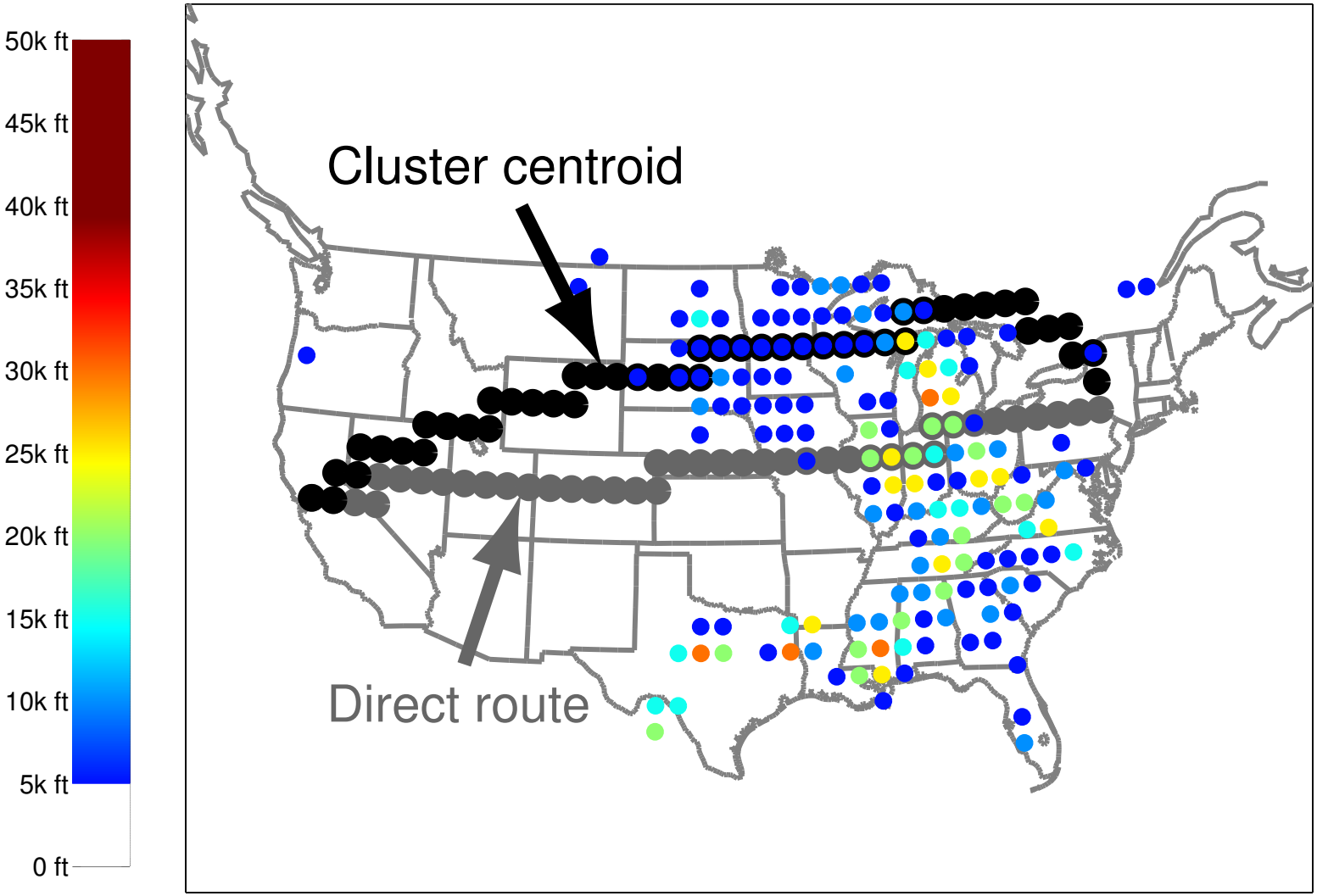
Lower resolution weather data



Lower resolution weather data



Lower resolution weather data



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

Data summary

Reduced data

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

Data summary

Reduced data

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

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

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

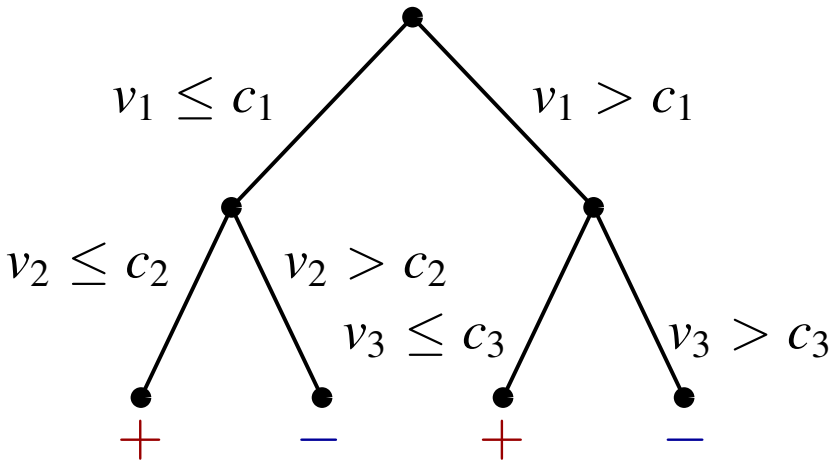
True positive rate

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

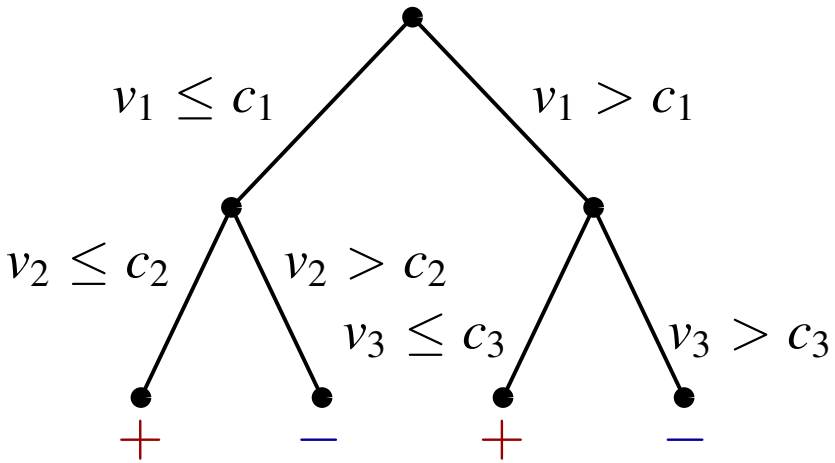
True negative rate

$$\text{TNR} = \frac{\text{\# of correctly predicted negative observations}}{\text{total \# of negative observations}}$$

Decision tree



Decision tree



- Shallow trees cannot capture more complex connections
- Deep trees tend to overfit

Random forest

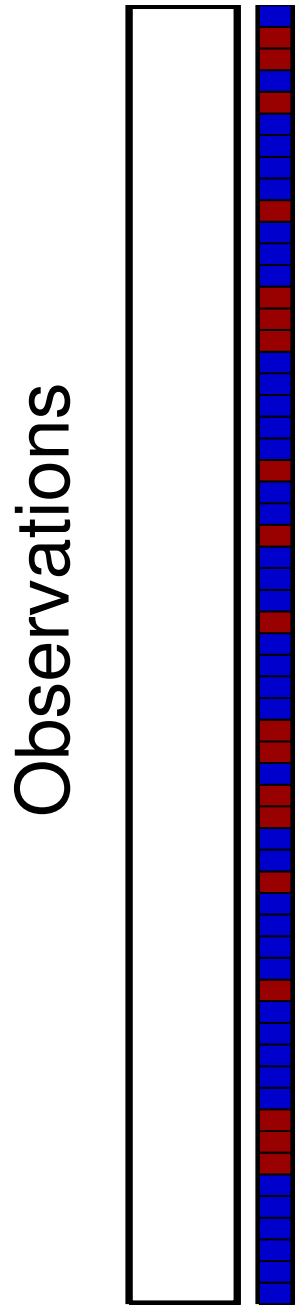
- Consists of many weak learners (shallow decision trees)
- Each decision tree is built with:
  - Randomly selected subset of observations
  - Randomly selected subset of features
- Ensemble prediction: weighted vote of each weak learner

Random forest

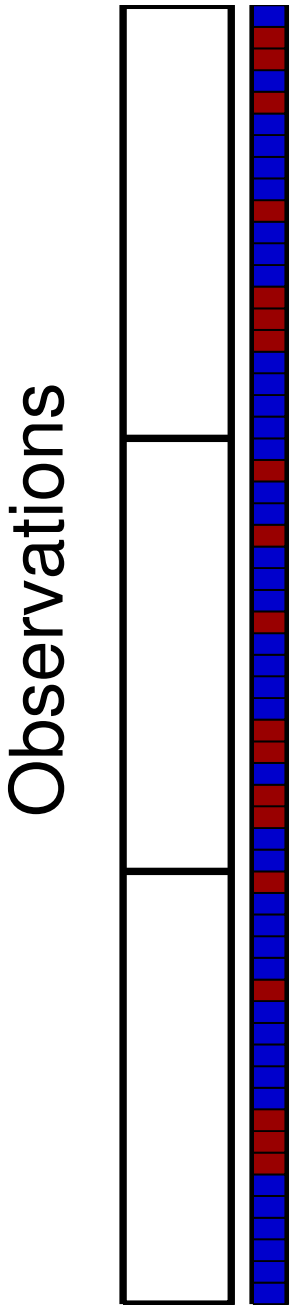
- Consists of many weak learners (shallow decision trees)
- Each decision tree is built with:
  - Randomly selected subset of observations
  - Randomly selected subset of features
- Ensemble prediction: weighted vote of each weak learner

$\Rightarrow$ Advantage: reduce sensitivity to noise $\Rightarrow$ reduce overfitting

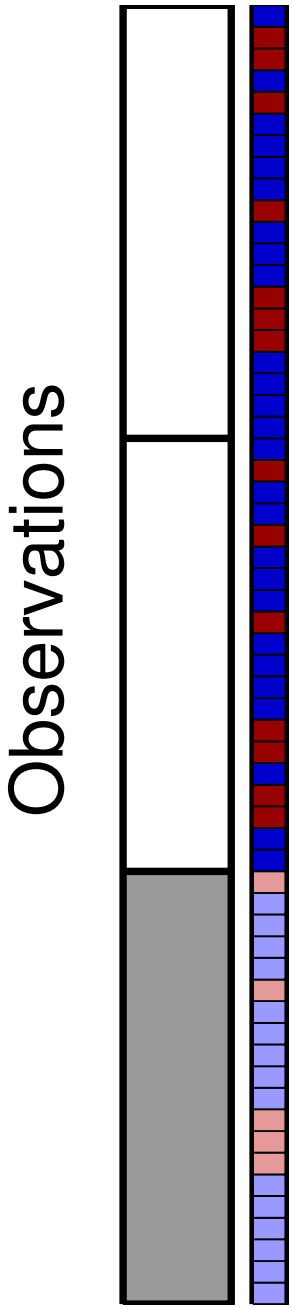
Model development



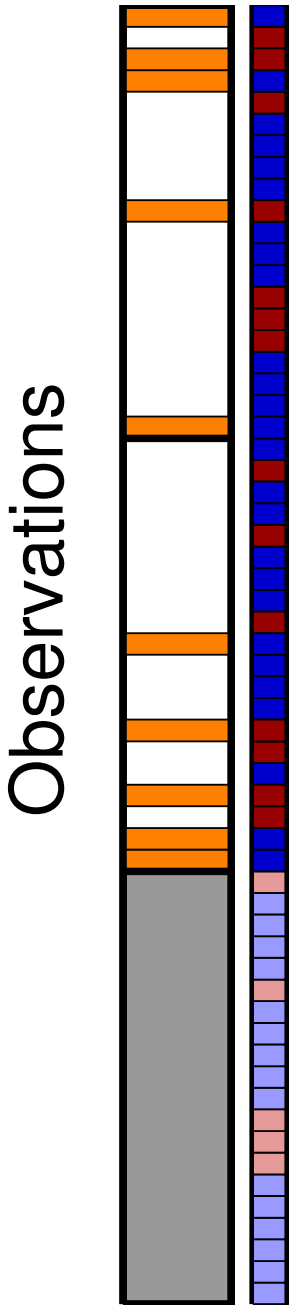
Model development



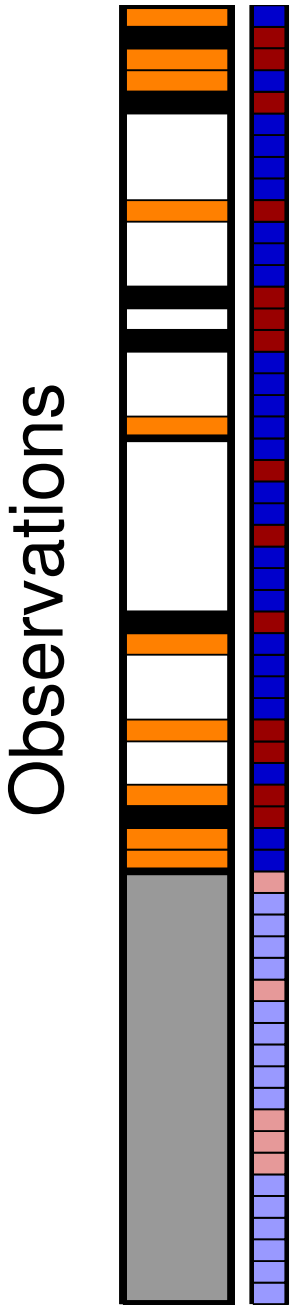
Model development



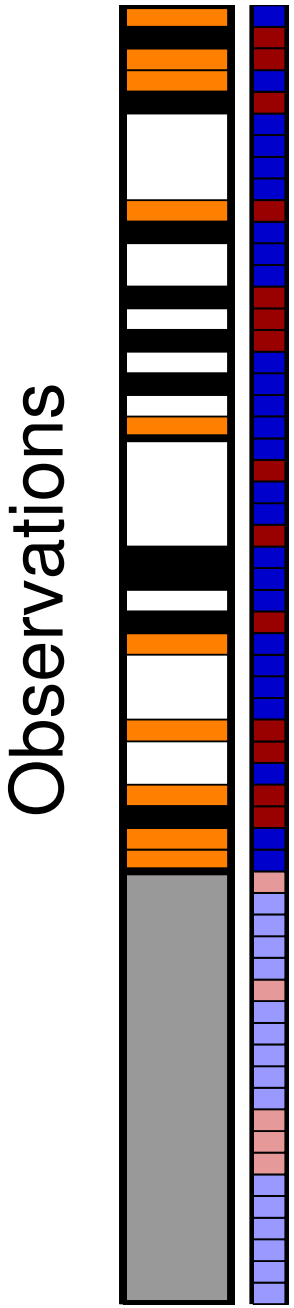
Model development



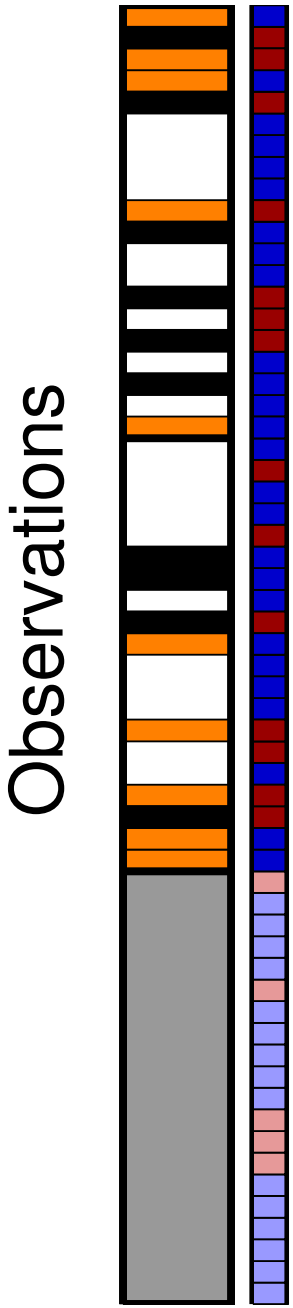
Model development



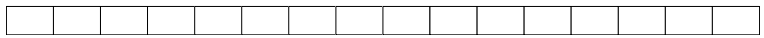
Model development



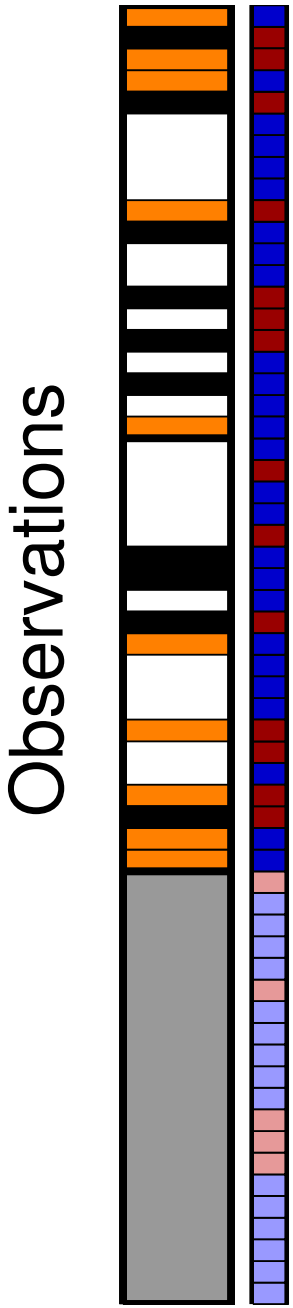
Model development



Features

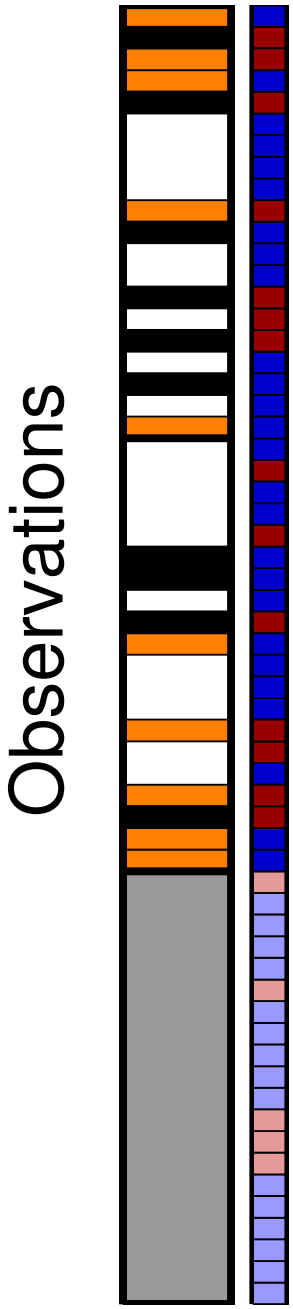


Model development



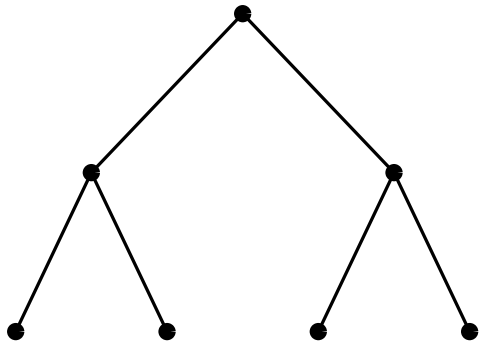
Features

Model development

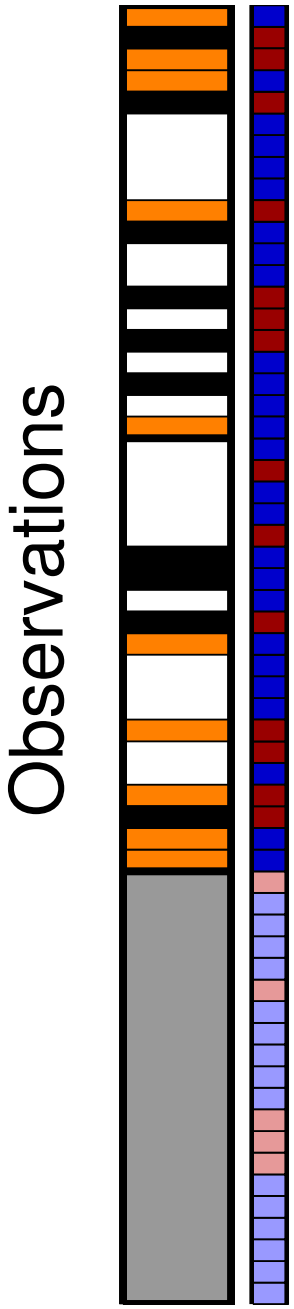


Features

Weak learner

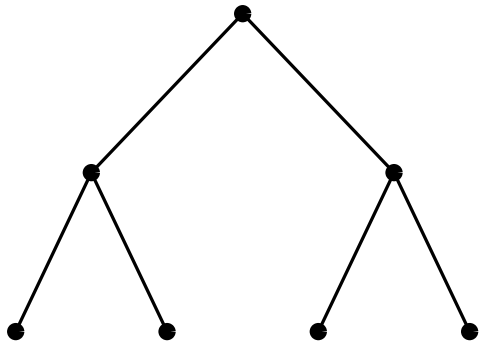


Model development



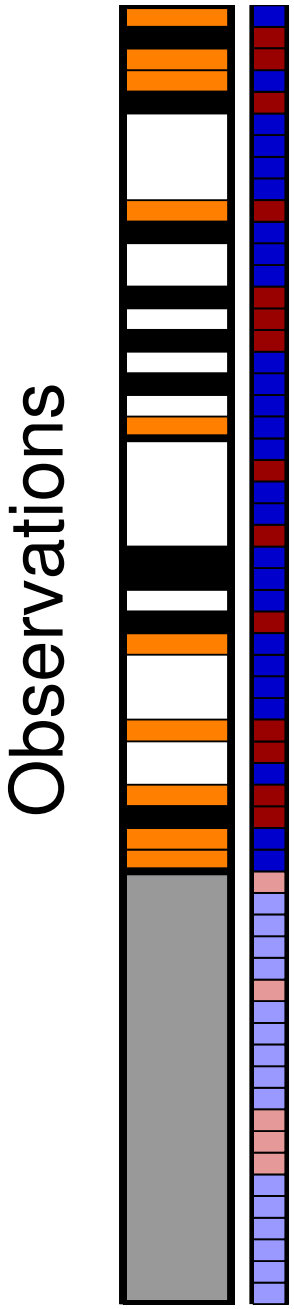
Features

Weak learner



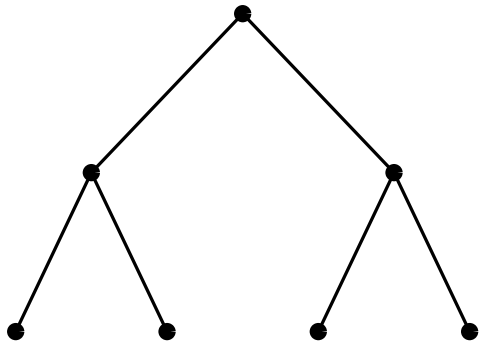
$\varepsilon =$ sub test prediction error

Model development



Features

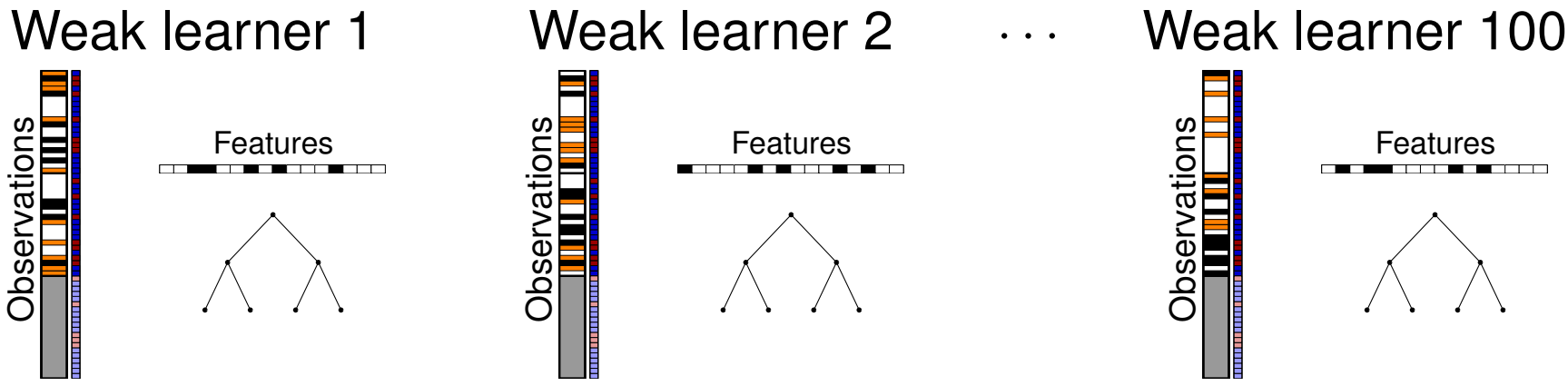
Weak learner



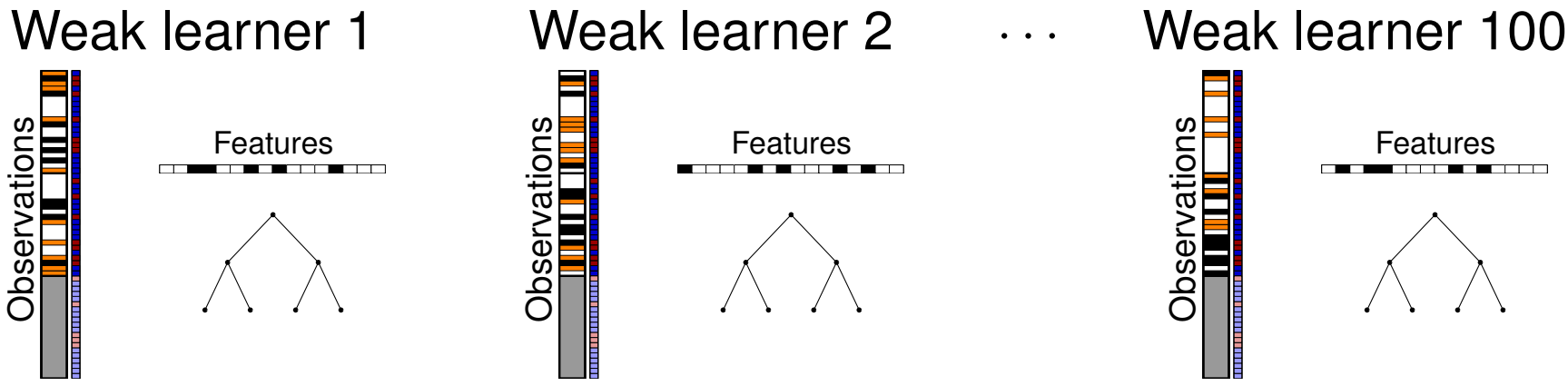
$\varepsilon =$ sub test prediction error

$$\alpha = \begin{cases} \nearrow \text{ as } \varepsilon \searrow, \varepsilon < 0.5 \\ 0, \text{ otherwise} \end{cases}$$

Model development



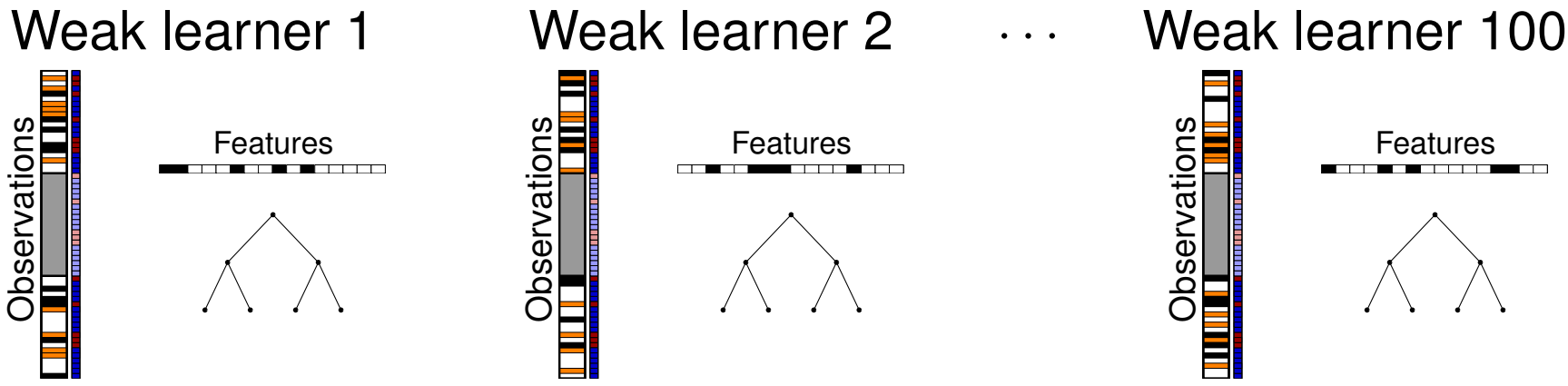
Model development



Ensemble prediction:

Weighted vote from each weak learner

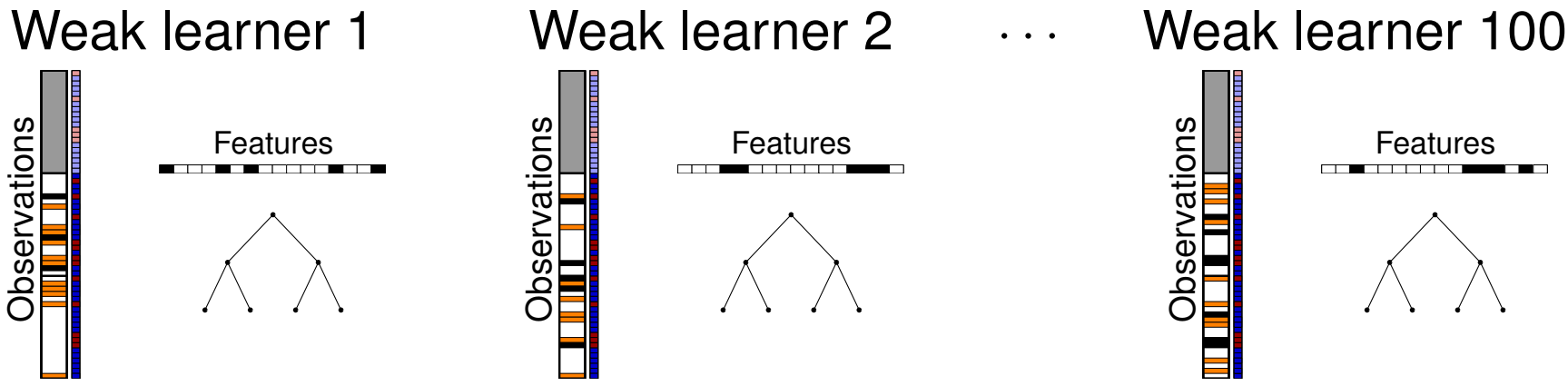
Model development



Ensemble prediction:

Weighted vote from each weak learner

Model development



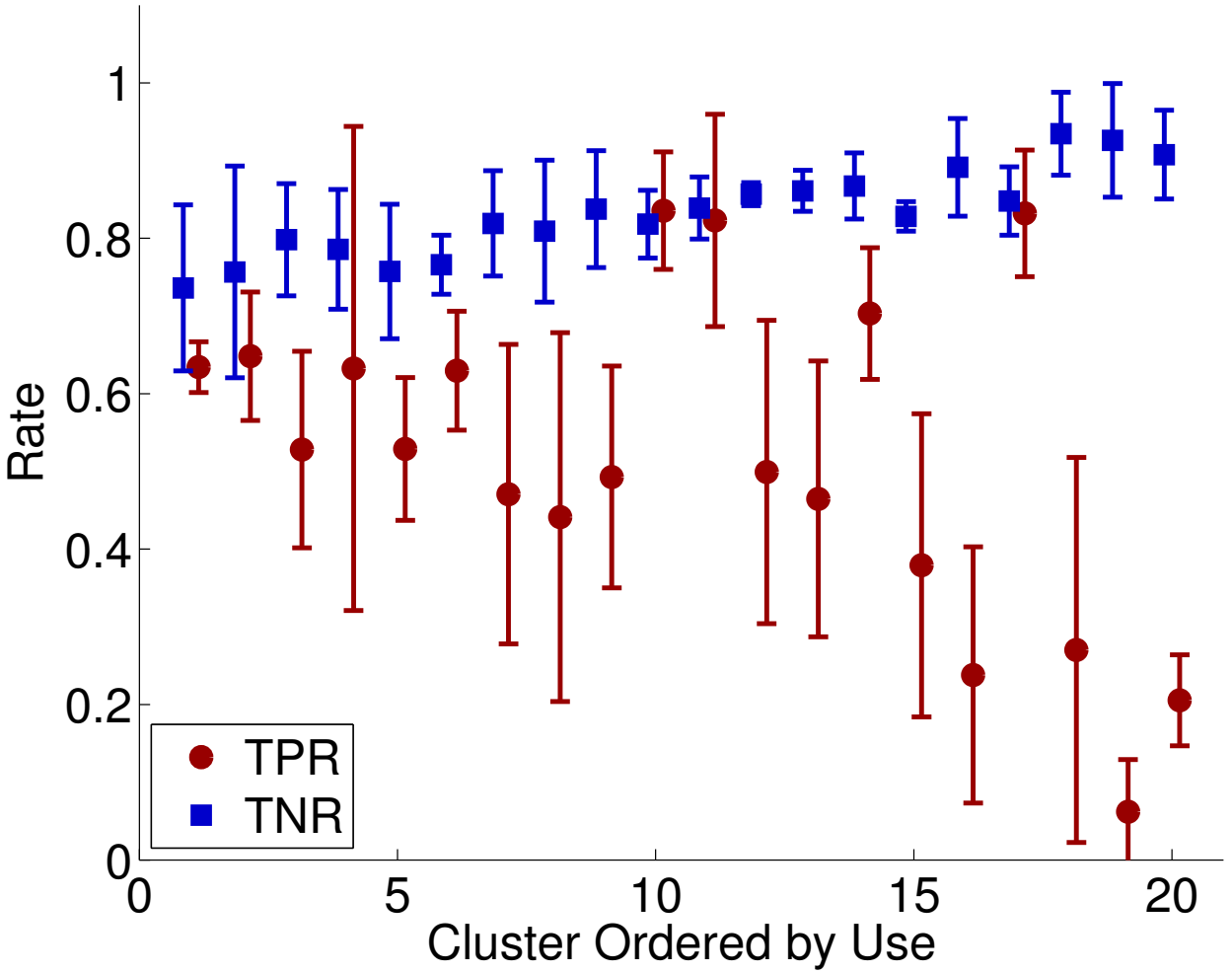
Ensemble prediction:

Weighted vote from each weak learner

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

Prediction results



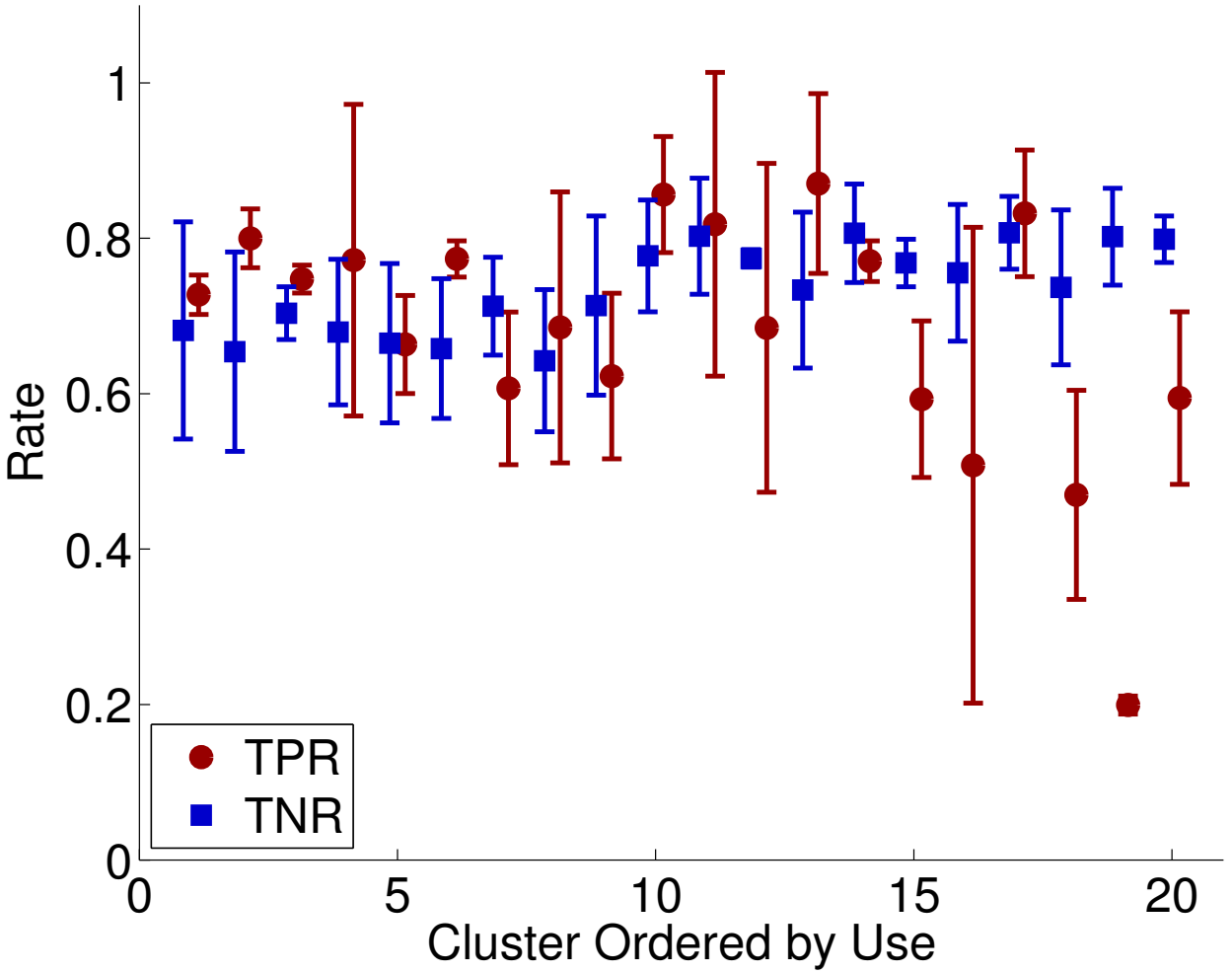
Synthetic Minority Oversampling Technique (SMOTE)

Within the training set:

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

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., SMOTE: Synthetic Minority Over-sampling Technique, *Journal Of Artificial Intelligence Research*, Vol. 16, 2002, pp. 321 357.

Prediction results with SMOTE



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

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

Conclusions and future work

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