En Route Performance In The National Airspace System

Mark Hansen, Yulin Liu

University of California, Berkeley

Cara Chuang, David Lovell, Michael Ball

University of Maryland, College Park

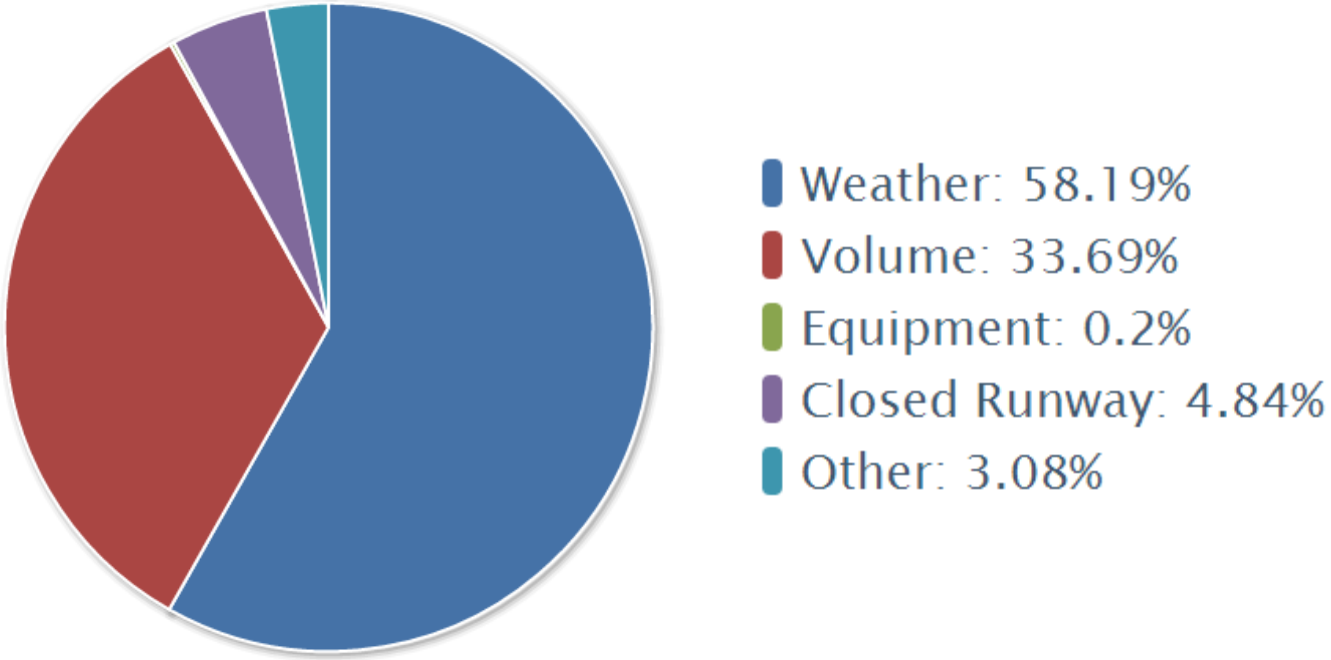
Jan 11st, 2017

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Macroscopic Variation in Flight Inefficiency
- Impact of Route Selection on Flight Inefficiency
- Conclusions

Background

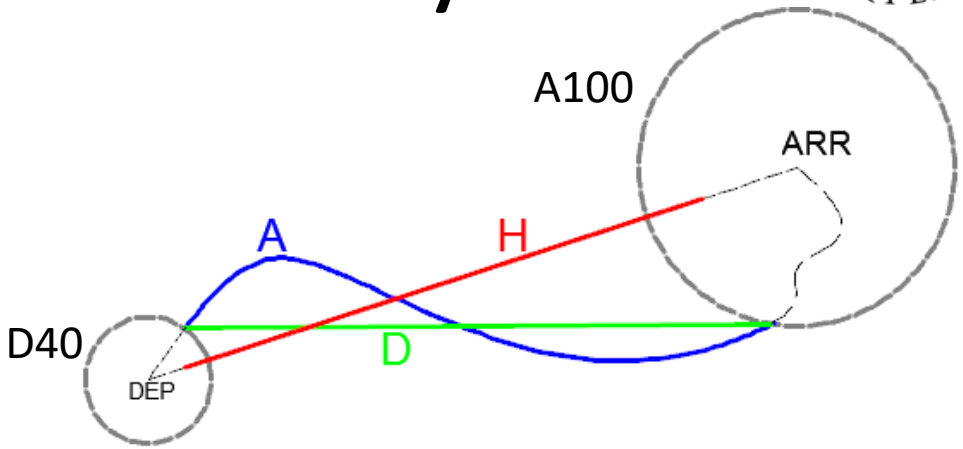
- FAA and Eurocontrol published metrics to evaluate flight en route inefficiency, and understanding the mechanism behind the inefficiency is of great importance;
- For flight delay we have:
- What about **en route inefficiency**?



Sources:
http://www.transtats.bts.gov/ot_delay/ot_delaycause1.asp?type=5&pn=1

Defining En Route Inefficiency

$$Inefficiency = \frac{A - H}{H}$$



- A: Actual flown distance from exit point to entry point;
- D: Great circle distance between local entry and exit point;
- H: Achieved distance (related to great circle distances from exit/entry points to arcs surrounding arrival/departure airports).

Sources:
https://www.faa.gov/air_traffic/publications/media/us_eu_comparison_2013.pdf
https://www.eurocontrol.int/sites/default/files/content/documents/single-sky/pru/news-related/2013-05-08-slides-workshop-achieved-distance.pdf

Project Goals

- Support FAA in developing en route inefficiency performance metrics
- For selected metrics, identify reasons for inefficiency
  - NAS route structure
  - Convective weather
  - Traffic management initiatives (TMIs)
  - Winds
- Eventually allow comparison with other ANSPs such as Eurocontrol

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Macroscopic Variation in Flight Inefficiency
- Impact of Route Selection on Flight Inefficiency
- Conclusions

Data Sources

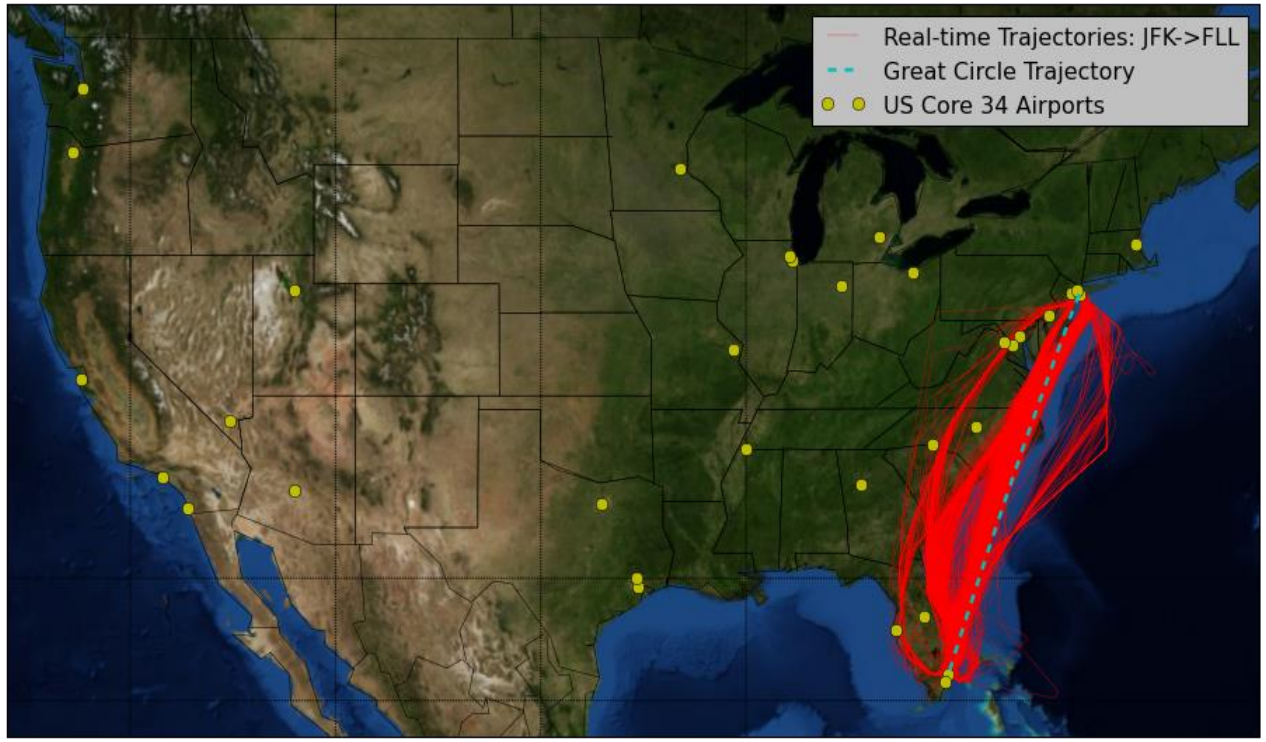
- Flight Event Data
  - From FAA Enhanced Traffic Management System (ETMS)
  - Flight level performance records from 2013 to 2014
  - We only focus on the traffic among the U.S. core 34 airports
- Flight Track Data
  - From FAA Traffic Flow Management System (TFMS)
  - Currently we focus on eight pairs in 2013:

    IAH ⟷ BOS, ORD ⟷ DCA, JFK ⟷ LAX and FLL ⟷ JFK

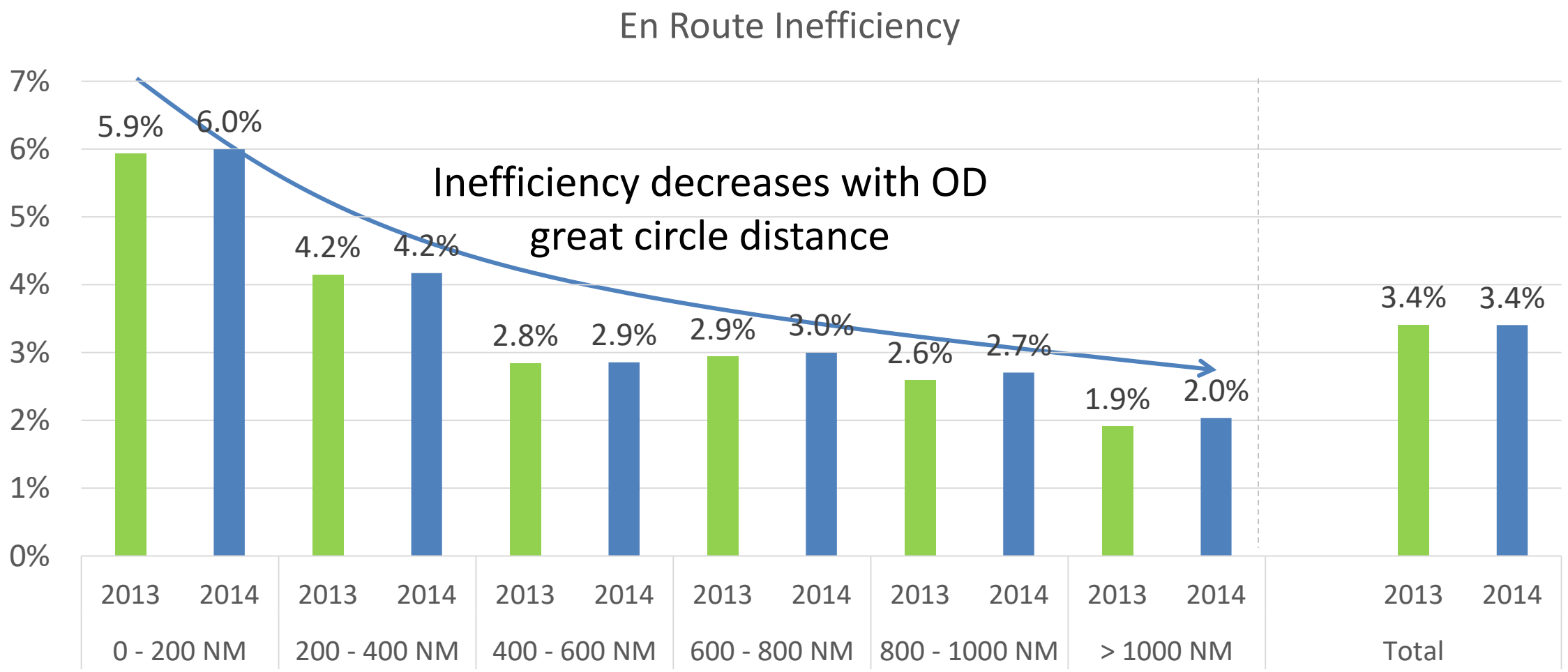
Summary Statistics

- Flight Event Data
  - Record the flight level distance measures, including filed distance, flown distance and achieved (benchmark) distance
  - Around 3 million flights per year in/out of core 34 airports, accounting for about 50% of total flights in/out of the US;
- Flight Track Data
  - Radar track points:

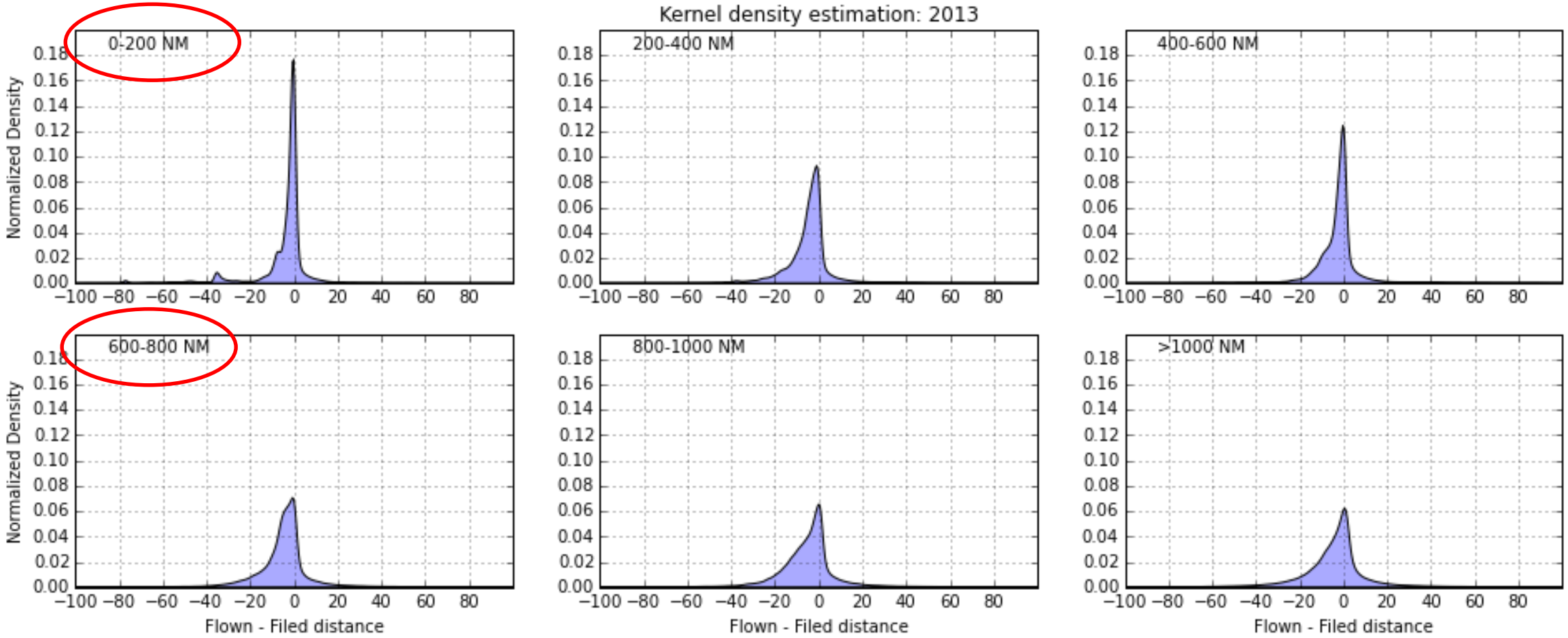
    Latitude, Longitude, Altitude, Time, Ground speed



En Route Inefficiency vs Great Circle Distance

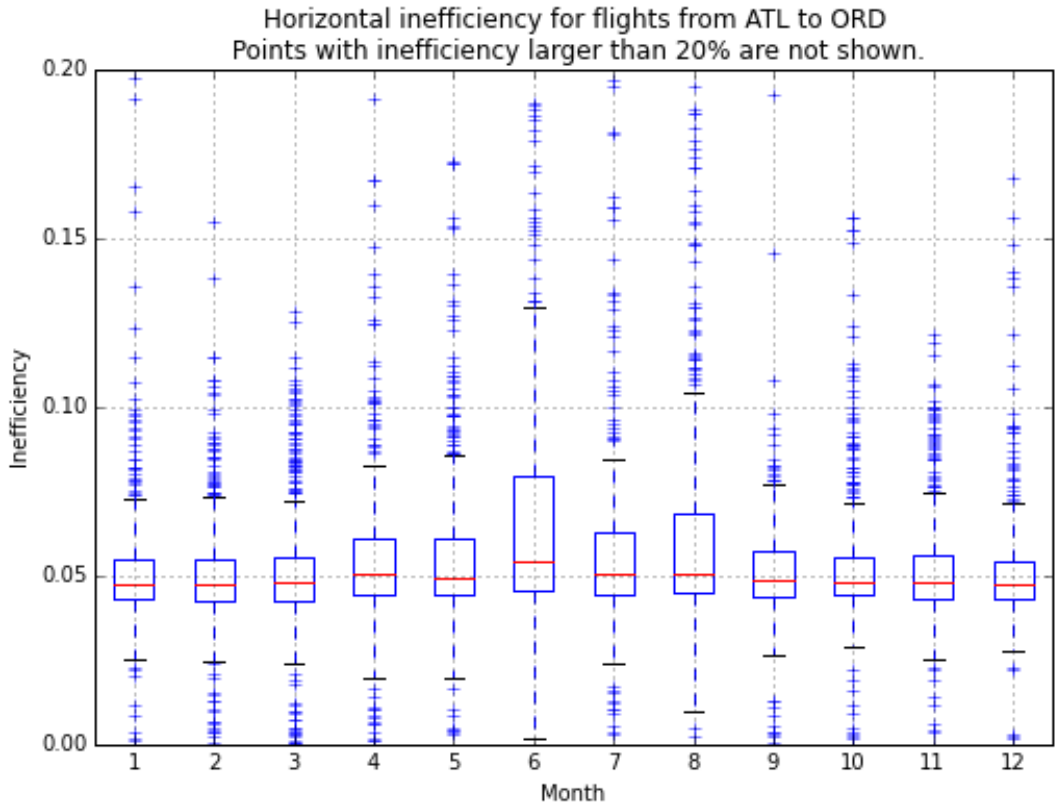


Gap Between Actual and Flight Plan Distance

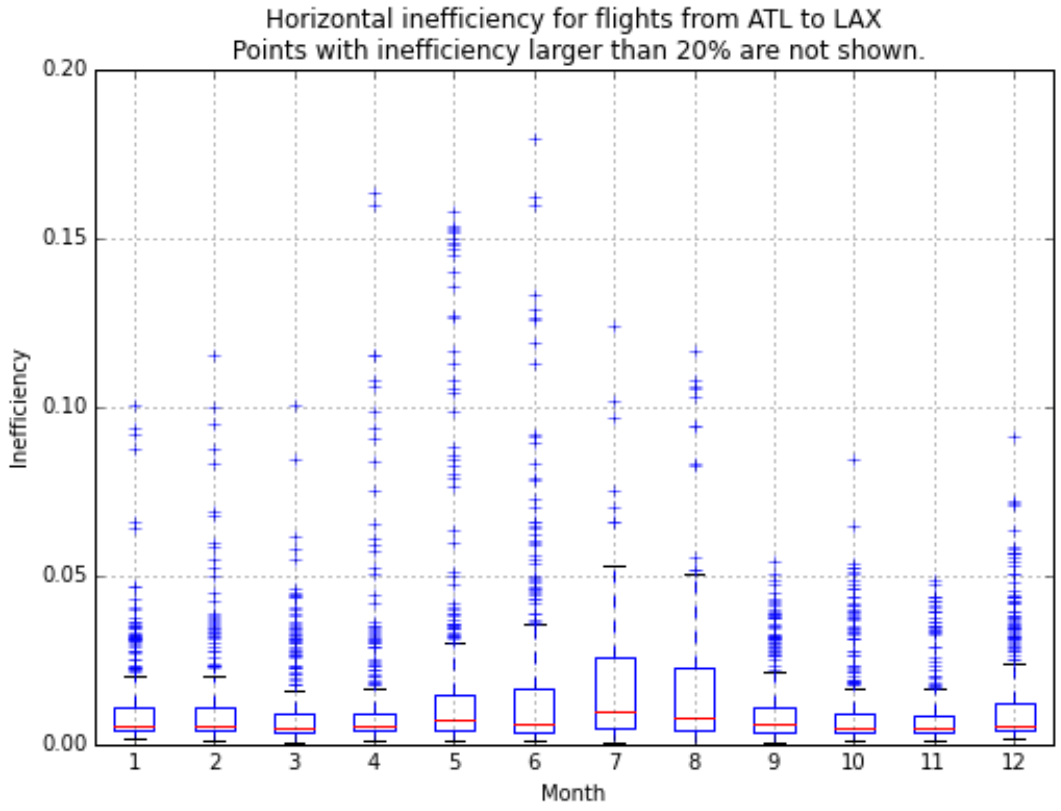


Inefficiencies for Representative Airport Pairs (2013)

ATL to ORD (6.86%)



ATL to LAX (1.28%)



Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Macroscopic Variation in Flight Inefficiency
- Impact of Route Selection on Flight Inefficiency
- Conclusions

Patterns of Variation in Flight En Route Inefficiency

- Quantify how departure/ arrival airports, seasons and flight length affect flights' en route inefficiencies;
- We use linear regression to build two *fixed effect models* to estimate those effects;
- The first model investigates the independent effects of terminals, month, and flight length, while the second model takes a closer look at the monthly variations within each departure/ arrival airport.

Model Specification

- Model I: Include airports, months and flight length categories as explanatory variables, and monthly variation is airport independent. (6M observations, 82 Variables)

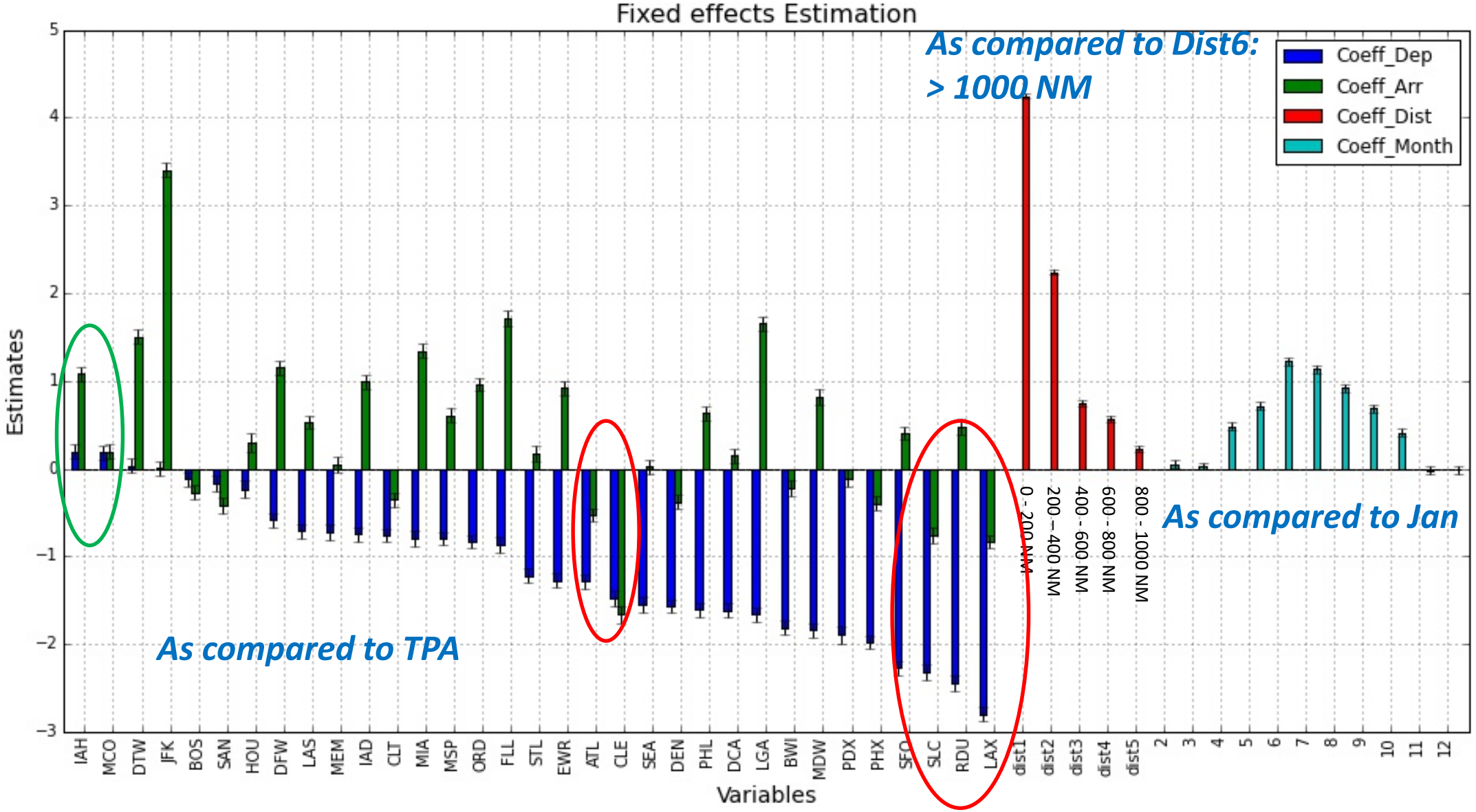
$$Ineffiency = \sum_{dep} \beta_{dep} \cdot X_{dep} + \sum_{arr} \beta_{arr} \cdot X_{arr} + \sum_{mon} \beta_{mon} \cdot \boldsymbol{X_{mon}} + \sum_{i}^{5} \beta_i \cdot Dist_i$$

- Model II: Include (Airport-Month) tuple and flight length categories as explanatory variables, which allows monthly variation to be airport specific. (6M observations, 808 Variables)

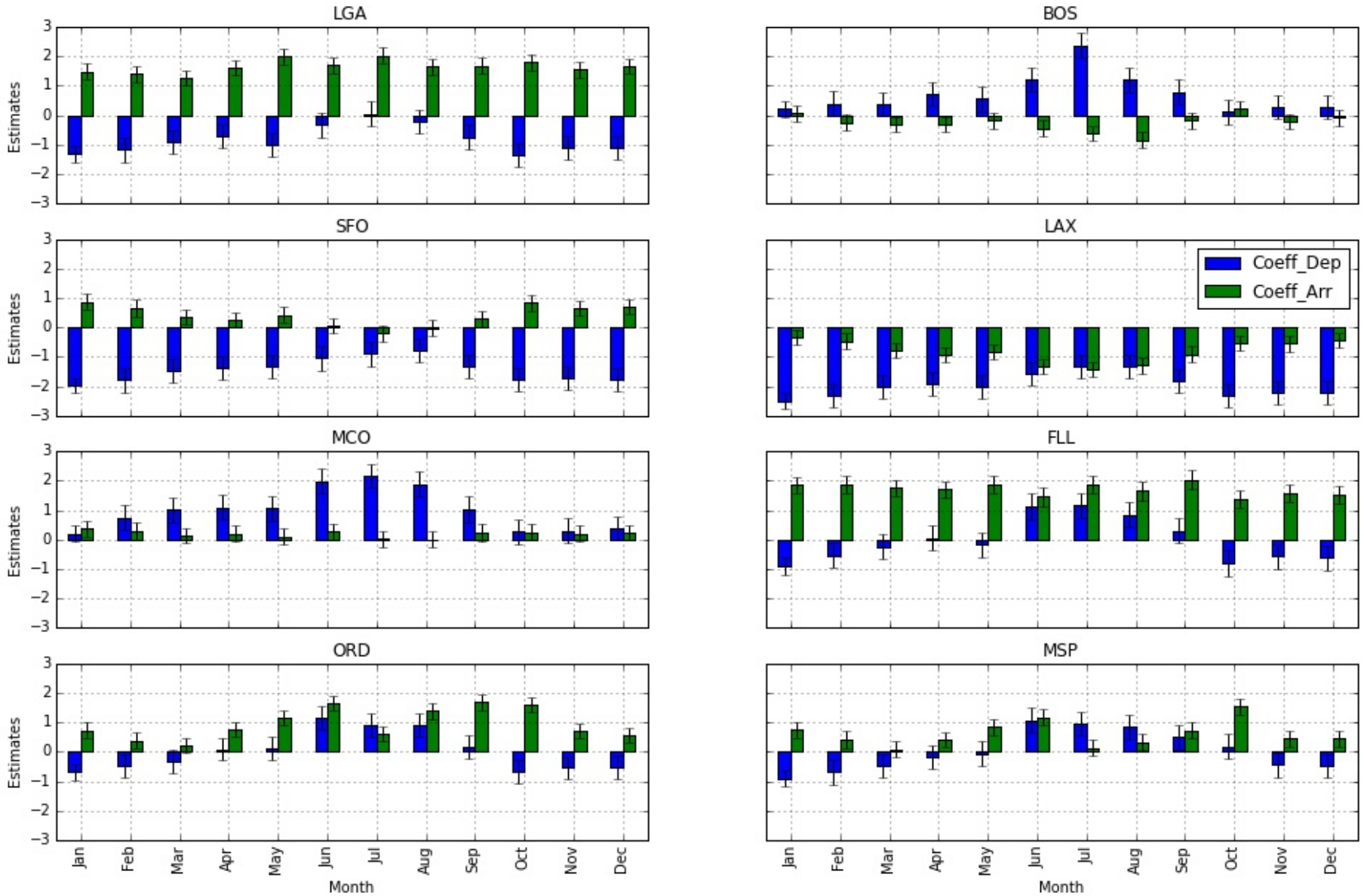
$$Ineffiency = \sum_{dep,mon} \beta_1 \cdot \boldsymbol{X_{dep-mon}} + \sum_{arr,mon} \beta_2 \cdot \boldsymbol{X_{arr-mon}} + \sum_{i}^{5} \beta_i \cdot Dist_i$$

  - $Dist_1: 0 - 200\ NM; Dist_2: 200 - 400\ NM; Dist_3: 400 - 600\ NM;$
  - $Dist_4: 600 - 800\ NM; Dist_5: 800 - 1000\ NM; Dist_6: > 1000\ NM$

Model I - Estimation



Model II – Monthly Variation



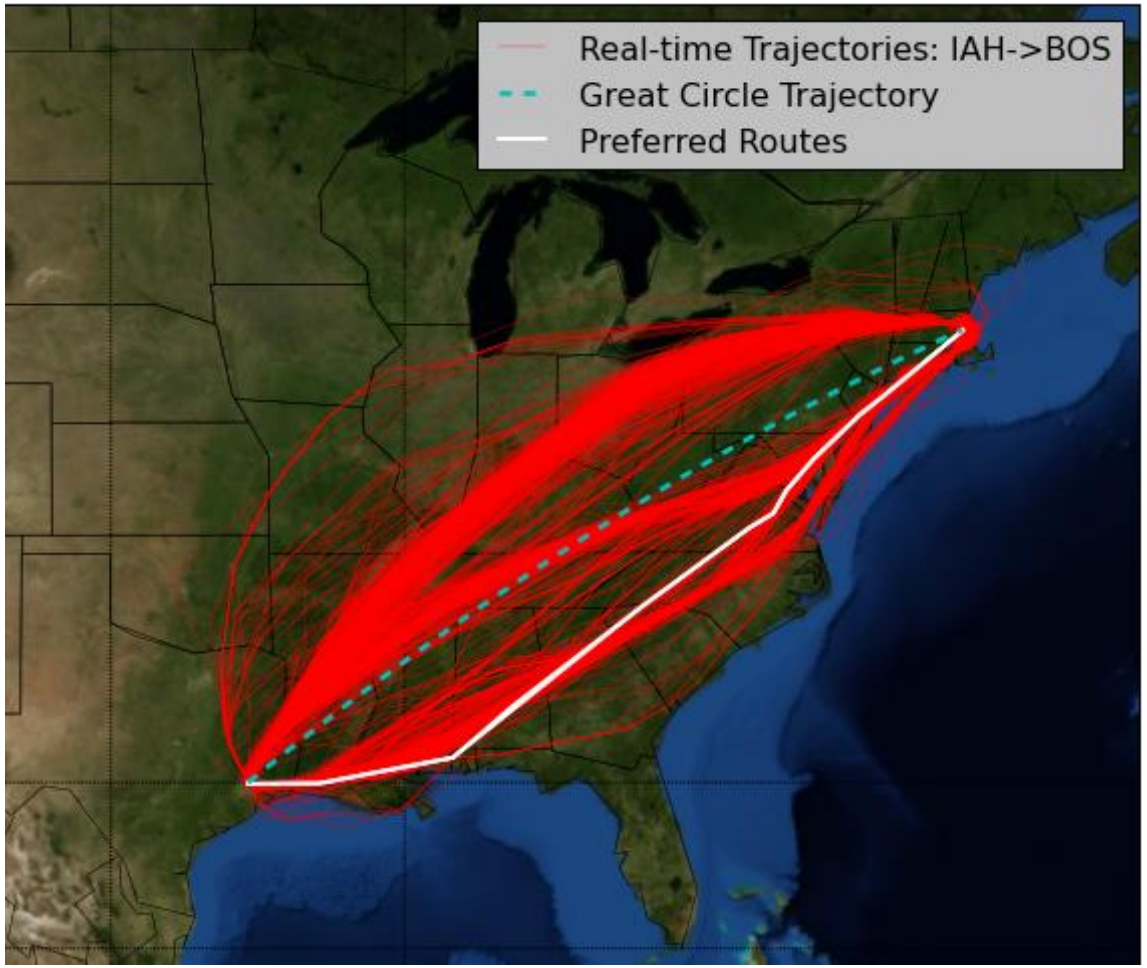
Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Macroscopic Variation in Flight Inefficiency
- Impact of Route Selection on Flight Inefficiency
- Conclusions

Why Route Selection Matters?

- Macroscopic models well explain the variation of en route performance, but have relatively low R squared;
- Trajectories (red curves) show obvious clustering in the airspace;
- Different clusters appear to have different en route performance.

IAH → BOS (2013)



Finding Nominal Routes

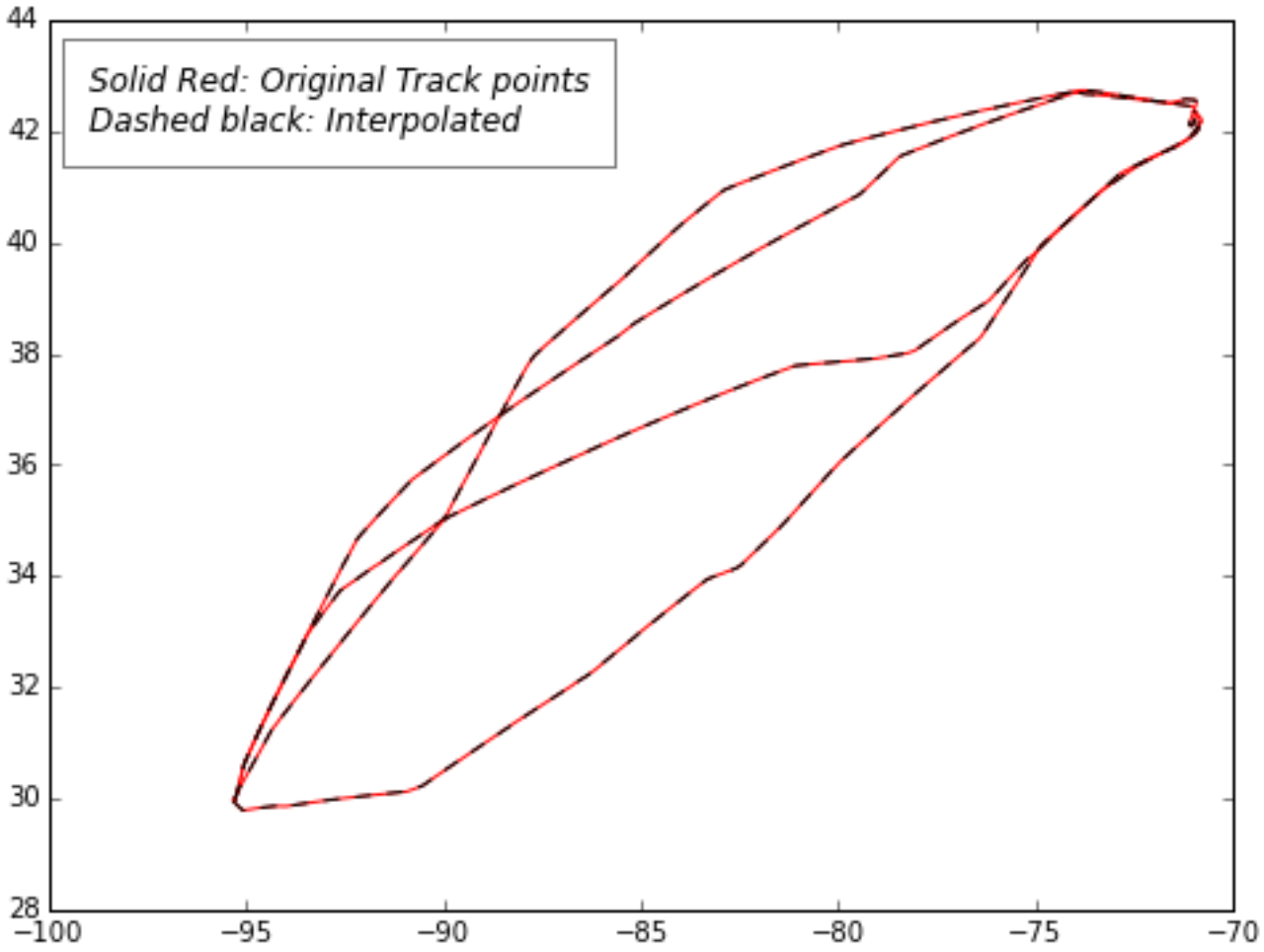
- We define *Nominal Routes* as the set of representative trajectories for a given OD pair;
- Nominal routes help us understand the NAS route structures, and further en route performance;
- Trajectory clustering algorithm helps us achieve such goal.

Clustering Algorithms

- *Step 0: Trajectory Cleaning*
  - Exclude both spatial and temporal discontinuity trajectories;
  - Exclude trajectories starting/ending outside terminal areas.
- *Step 1: Trajectory resampling*
  - Get trajectories with equal numbers of points;
  - Linear Interpolation (with respect to distance flown);
  - Each trajectory is represented by 100 points.
- *Step 2: Principal Component Analysis (PCA)*
  - Dimension reduction & Trajectory smoothing;
  - First five components can capture more than 90% of variations.
- *Step 3: Clustering*
  - Trajectory classifications;
  - DBSCAN algorithm is applied to the PCA components to get representative clusters;
  - solve a 1-median problem to determine nominal route for each cluster

Resampling Example

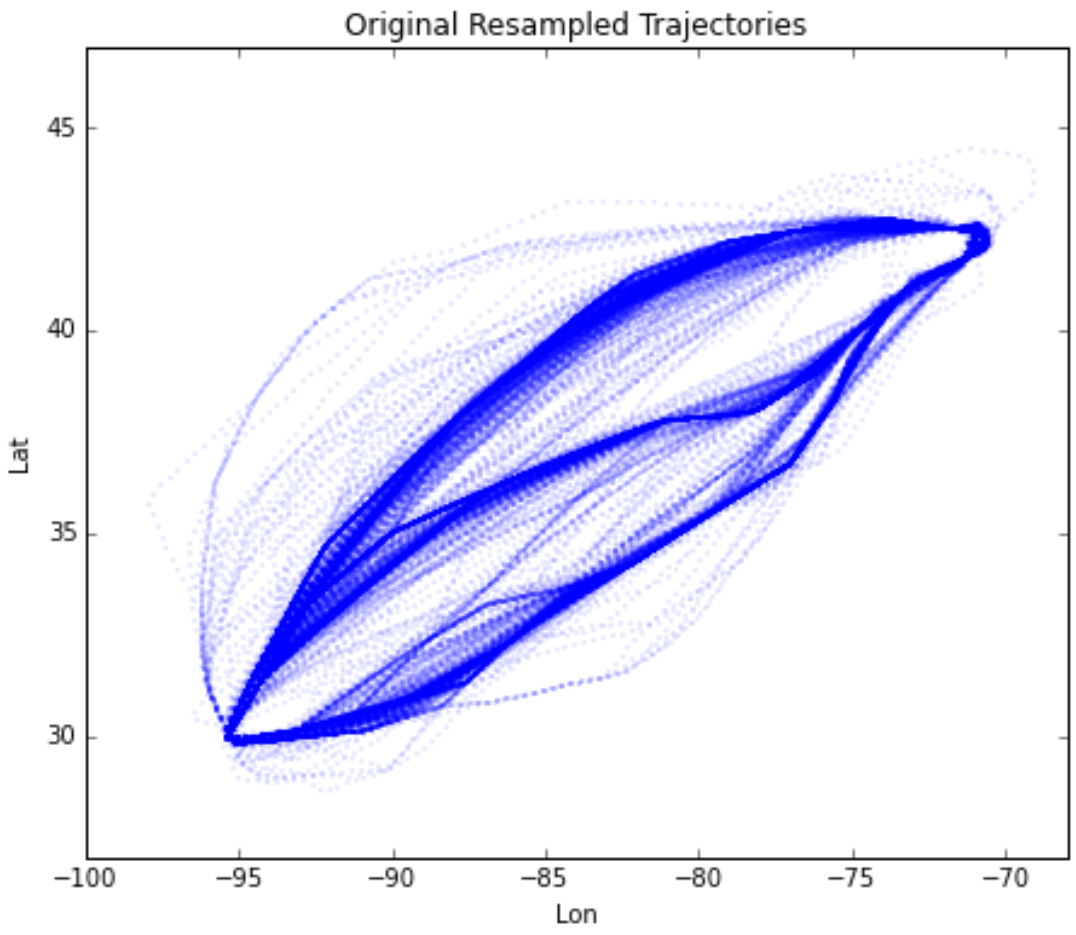
- Linear interpolation between the start and end tracking location for each route
- 100 pseudo points are predicted locations at:
  - Initial location (d0)
  - d0 + trajectory distance/99 (d1)
  - d1 + trajectory distance/99 (d2)
  - …
  - Final trajectory location (d100)

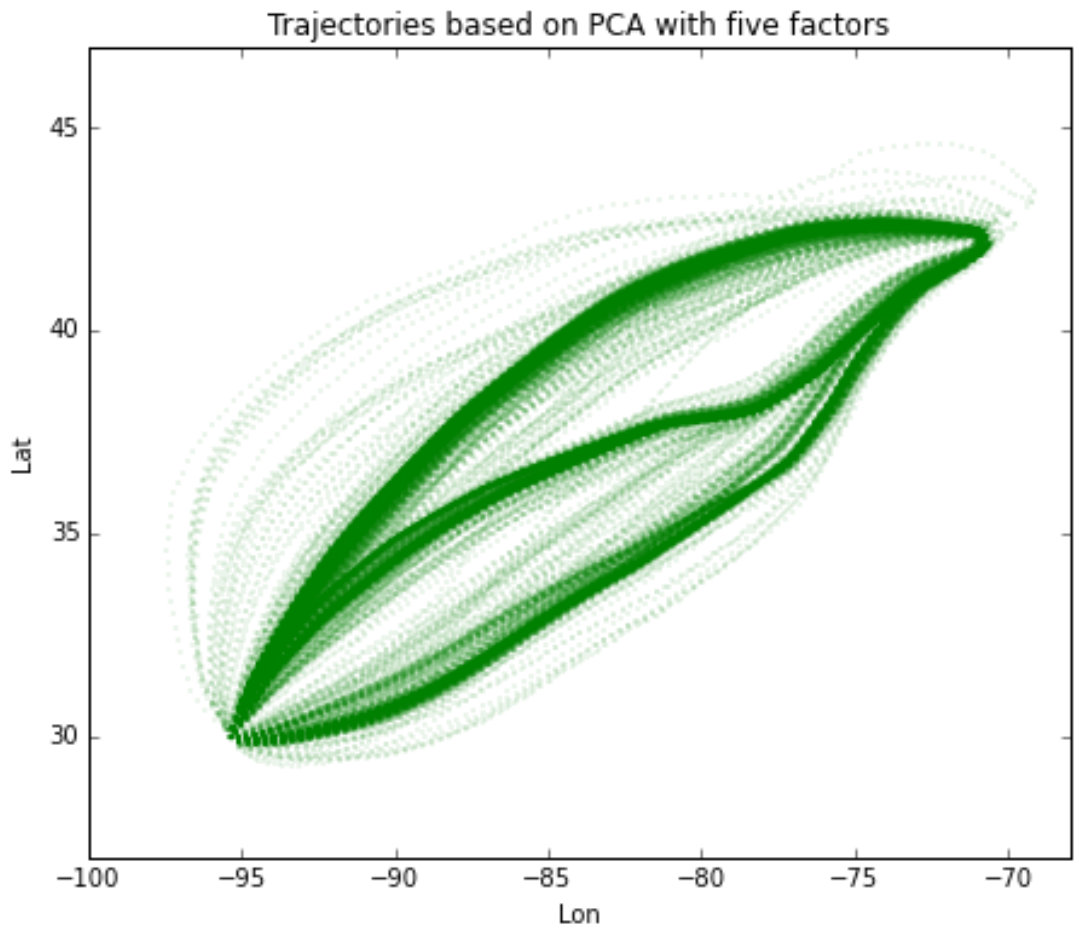


Dimension Reduction

- Reduce the dimension of trajectories – save computational time
- Improve the quality of clustering – Principal Component Analysis (PCA) can help to filter off noise and smooth the data
- Using PCA, we found that the first five components can capture almost all the variation e.g.
  - 99% for IAH → BOS
  - 96% for FLL → JFK
  - 94% for ORD → DCA

Example of Dimension Reduction (IAH→BOS)

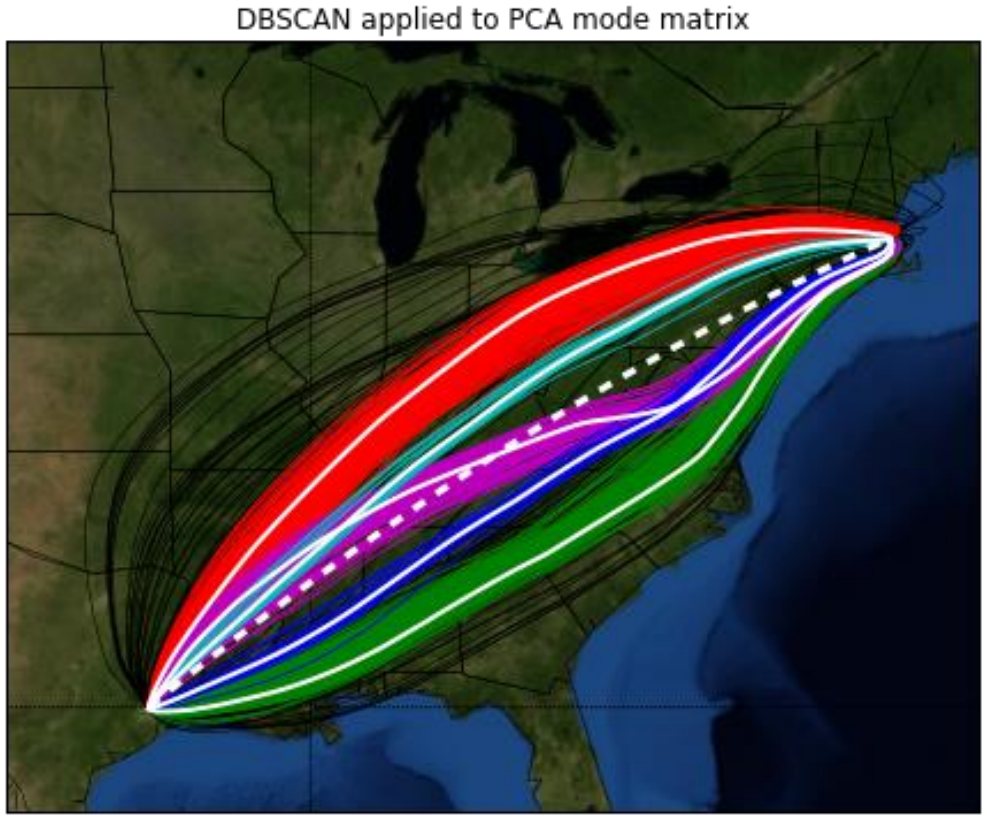


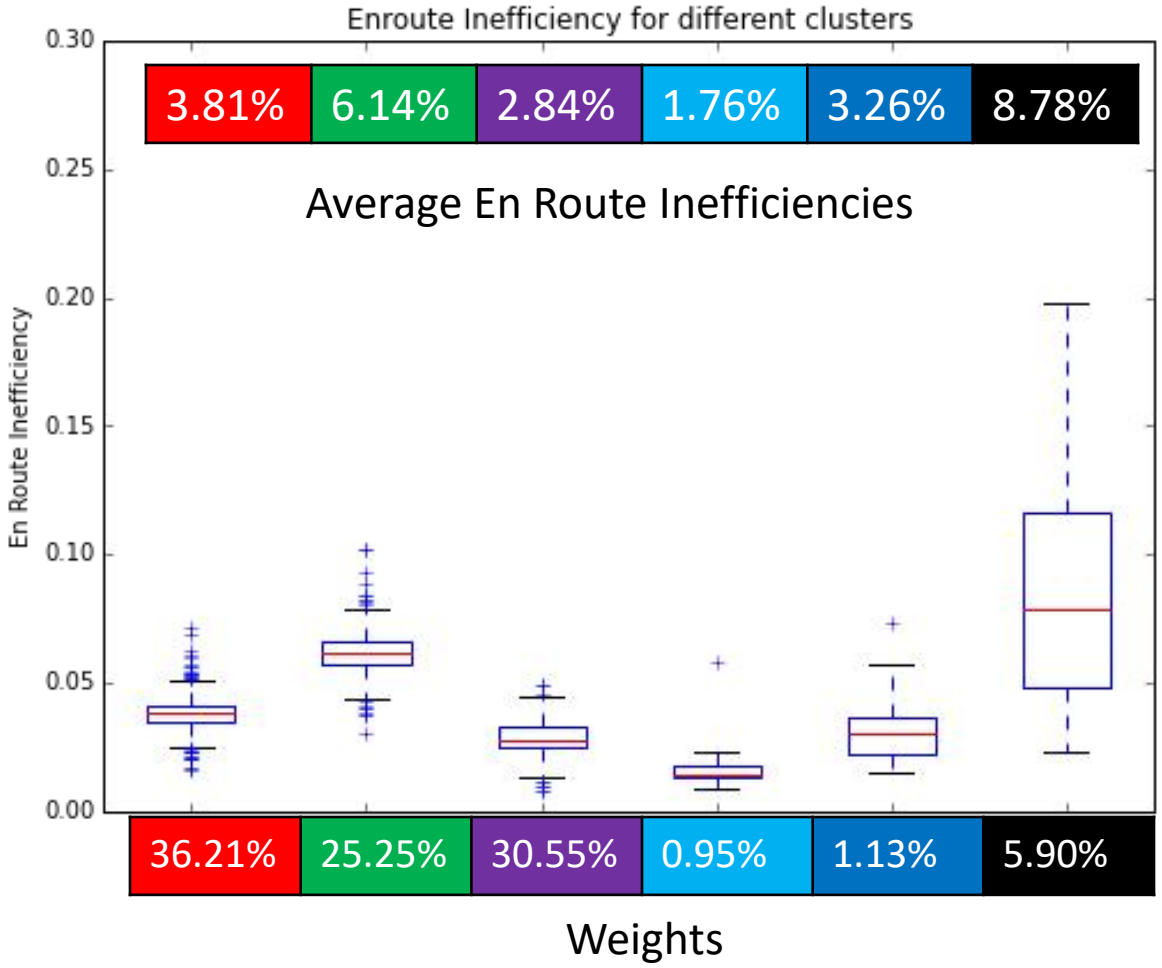


Trajectory Clustering

- Use trajectory PCA components to find sets of trajectories that are similar to each other;
- Apply DBSCAN algorithm because it
  - Does not need to pre-determine number of clusters
  - Allows trajectories to be identified as outliers
  - Can limit variation within each cluster

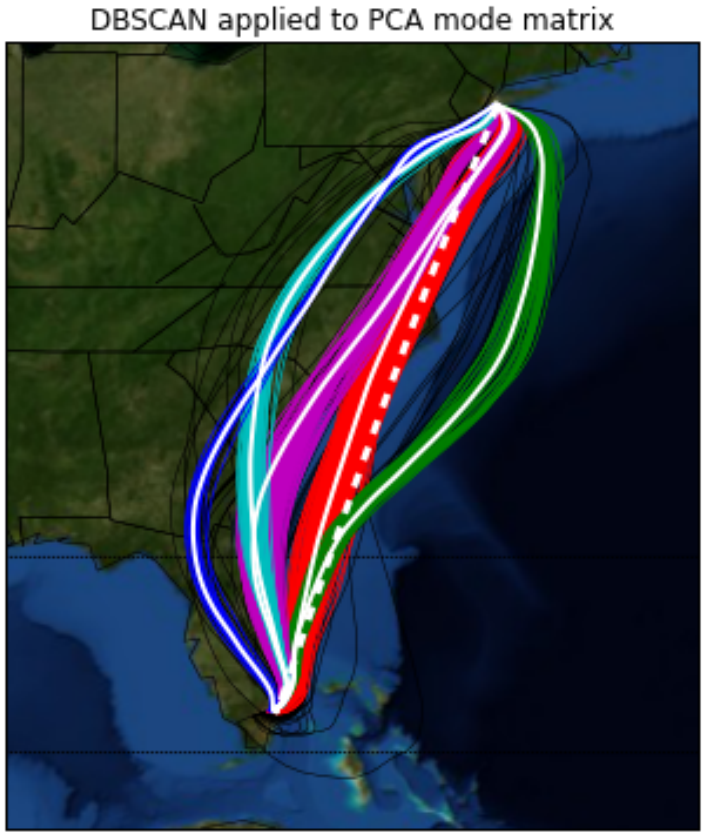
IAH → BOS (1679 of original 1817)

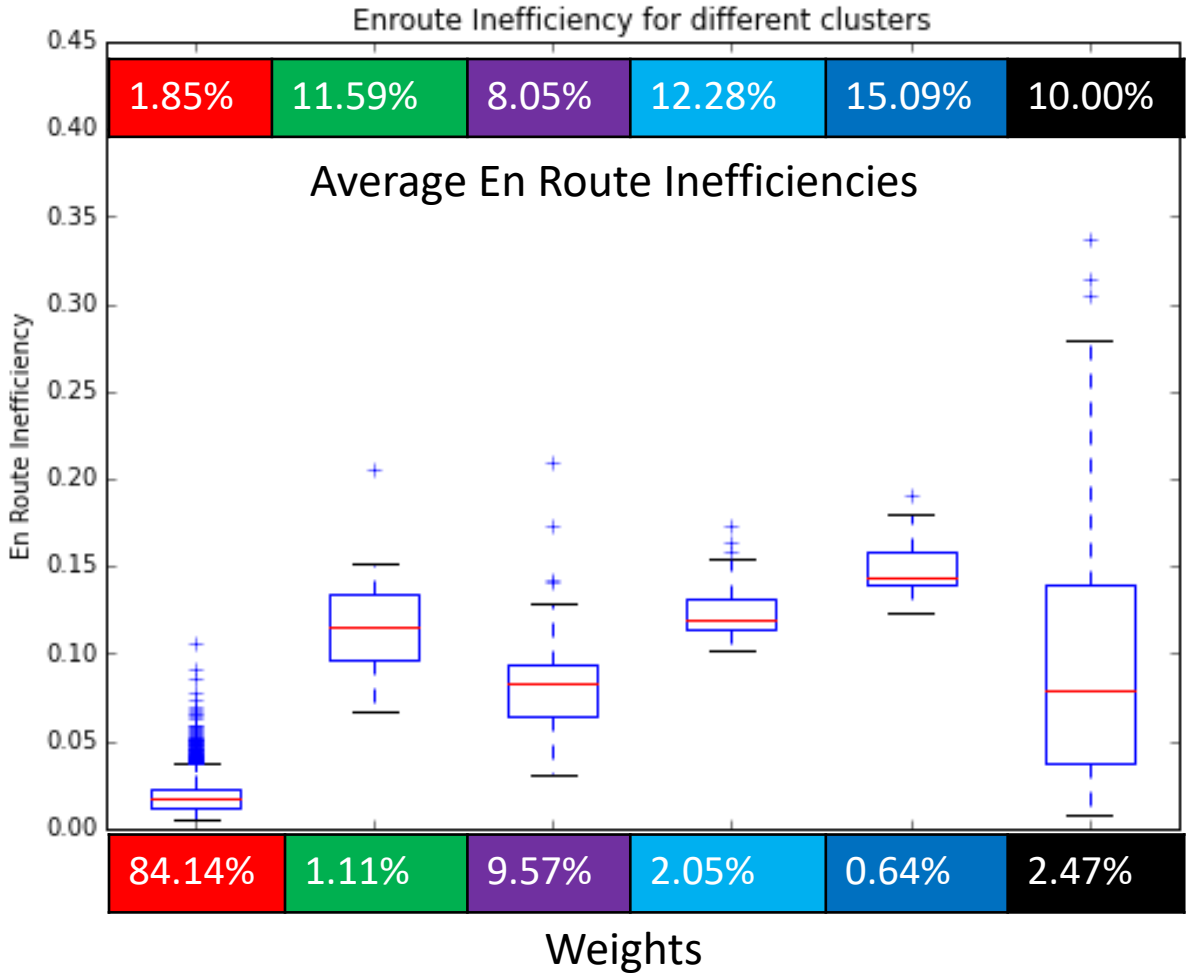




Black curves are classified as outliers
White Solid curves are Nominal Routes
White Dashed curve is great circle trajectory

JFK → FLL (4043 of original 4273)





Black curves are classified as outliers
White Solid curves are Nominal Routes
White Dashed curve is great circle trajectory

Impact of Route Selection

- Build route-specific fixed effect models to capture variations in en route inefficiencies among representative clusters;
- Model specification
  - Separate models for each airport pair
  - $Inefficiency(\%) = \beta_0 + \beta_1' \cdot X_{month} + \beta_2' \cdot X_{ClusterID}$;
  - X_{month} and $X_{ClusterID}$ are categorical variables;
  - Cluster ID can be found on previous slides.

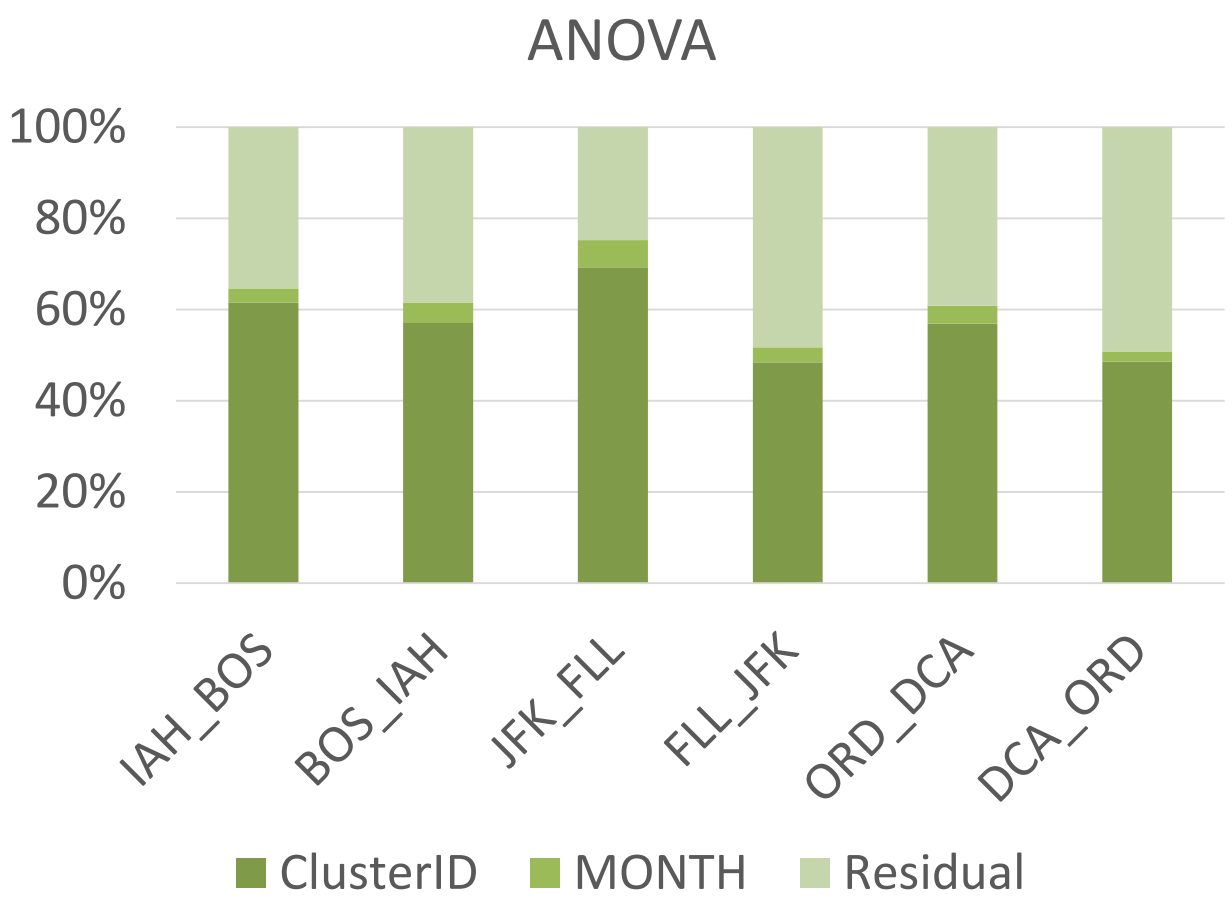
Estimation Results

- Al the cluster fixed effects are compared with the outlier groups;
- While most of them are significant and with plausible sign, the explanatory power greatly enhanced.

	IAH_BOS	BOS_IAH	JKF_FLL	FLL_JFK	ORD_DCA	DCA_ORD
Cluster ID – r	-4.328***	-4.831***	-8.108***	-18.470***	-24.538***	-22.026***
Cluster ID – g	0.525***	-2.463***	1.558***	-12.457***	-7.521***	-21.908***
Cluster ID – m	-3.697***	-5.801***	-1.970***	-12.956***	-15.434***	-11.593***
Cluster ID – c	-4.292***	-7.017***	2.240***	-	-19.705***	-13.695***
Cluster ID - b	-0.409	-5.498***	5.058***	-	-	26.114***
R squared	0.6463	0.6147	0.7523	0.5167	0.6083	0.5076

Notes:
**** p < 0.01; ** p < 0.05; * p<0.1*

Analysis of Variance



- Route Selection explains much of the variation (~60%) in en route inefficiency;
- Identified clusters are helpful in understanding causal reasons for flight en route inefficiency

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Macroscopic Variation in Flight Inefficiency
- Impact of Route Selection on Flight Inefficiency
- Conclusions

Conclusions

- Flight en route inefficiency is on average 3.4%, but varies significantly with airport pairs and seasons;
- Long-haul flights tend to be more efficient than short-haul flights;
- For most airport pairs, individual flight trajectories, while unique, can be divided into natural clusters whose members are very similar to one another;
- “Outlier” trajectories not belonging to a cluster account for from 1-15% of the total, depending on the airport pair;

Conclusions (cont'd)

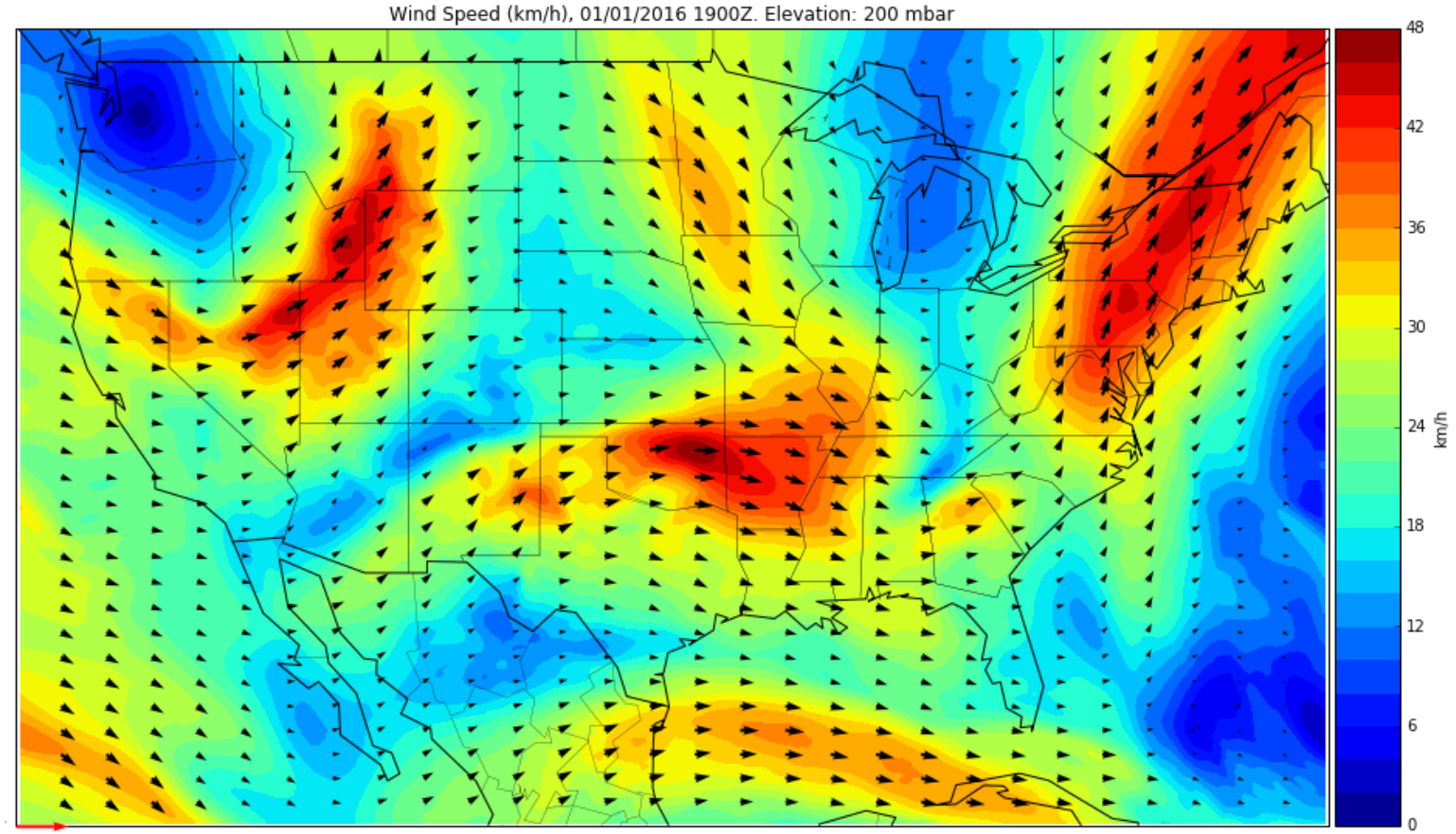
- Cluster membership accounts for about 60% of overall variation in inefficiency;
- Flights in summer seasons (May to August) are in general more inefficient than the others, but seasonal variation accounts for only 2-6% of the variation;

Ongoing Work

Other Causal Factors

- Convective weather
- **Wind**
- Miles-in-trail (MIT)

Wind Field Map

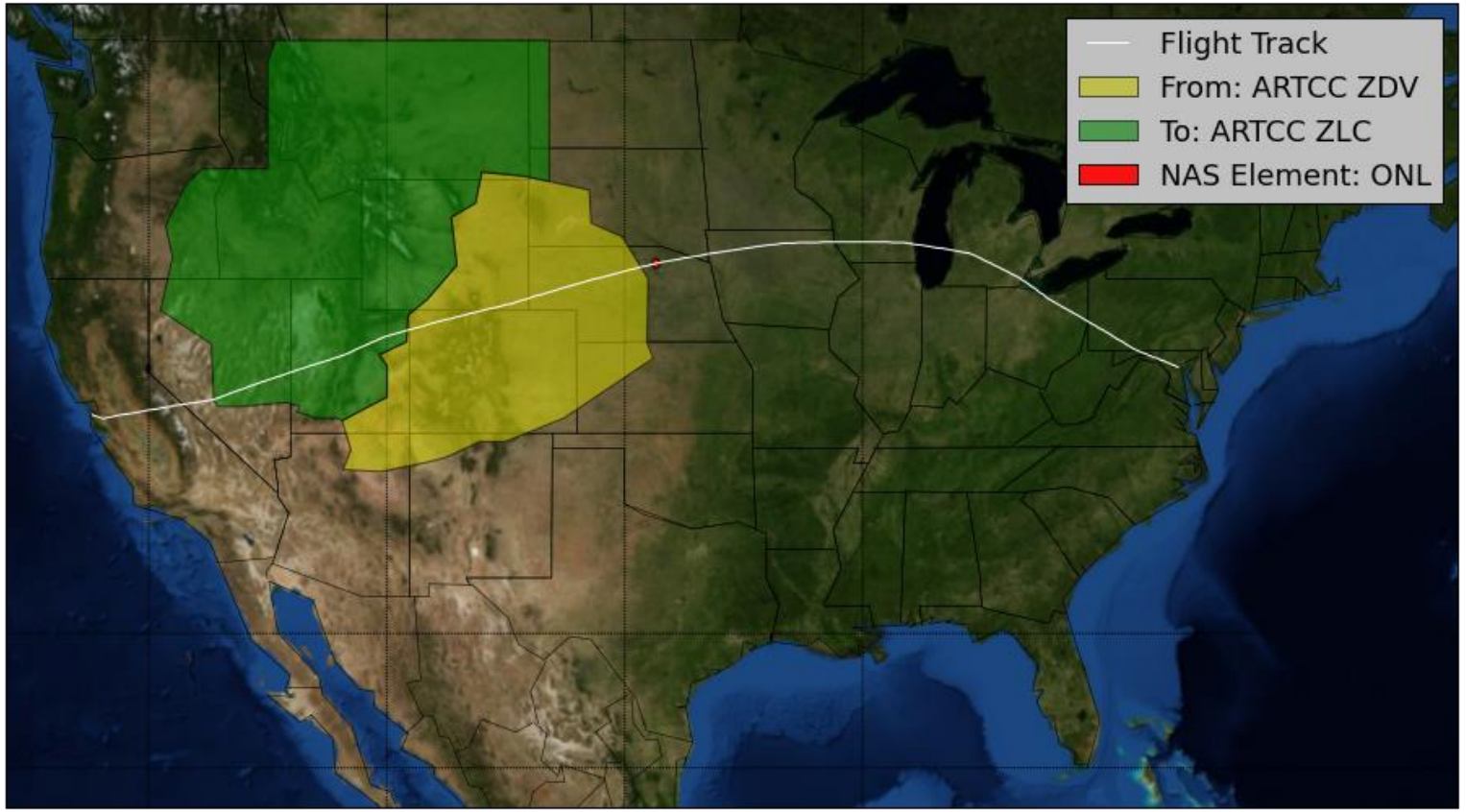


Ongoing Work

Other Causal Factors

- Wind
- Convective weather
- **Miles-in-trail (MIT)**

Example of MIT



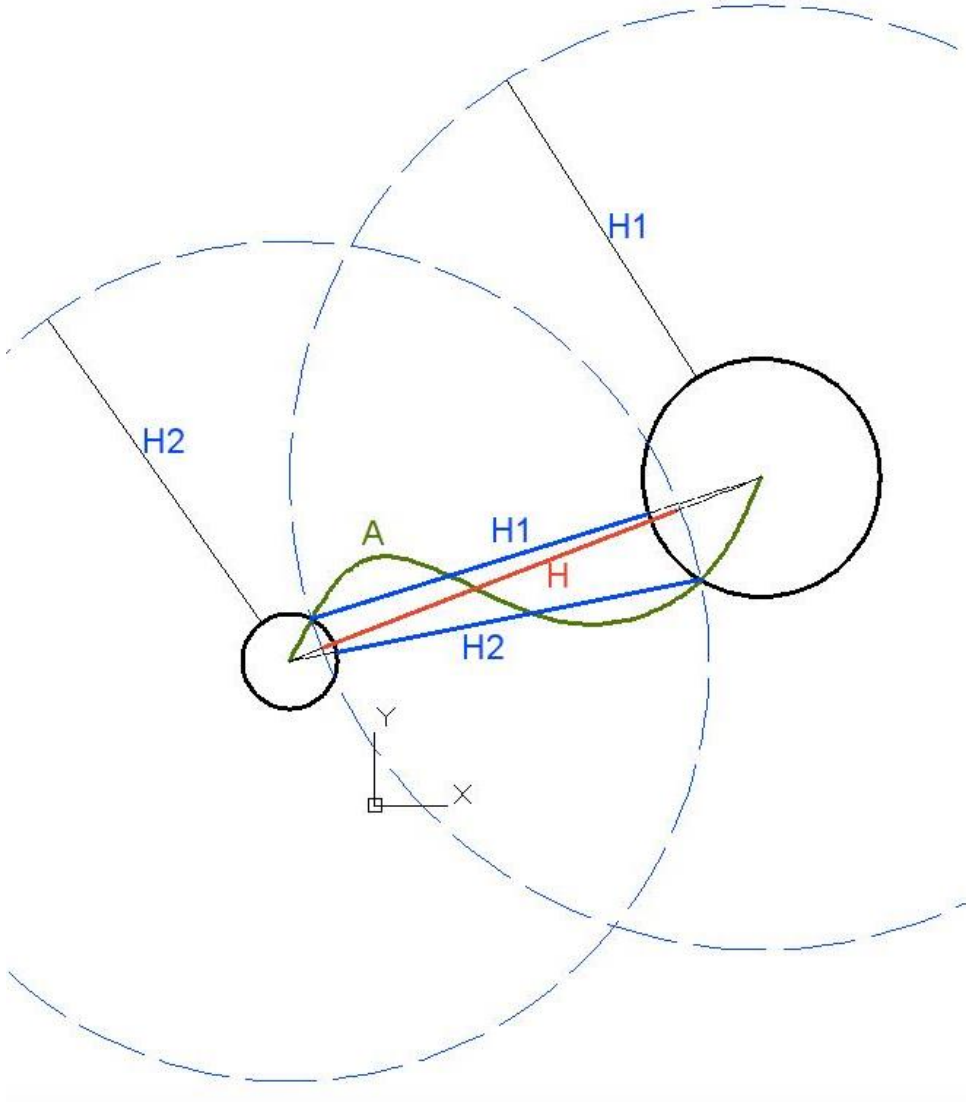
Thanks!

Q&A

liuyulin101@berkeley.edu

Backup Slides

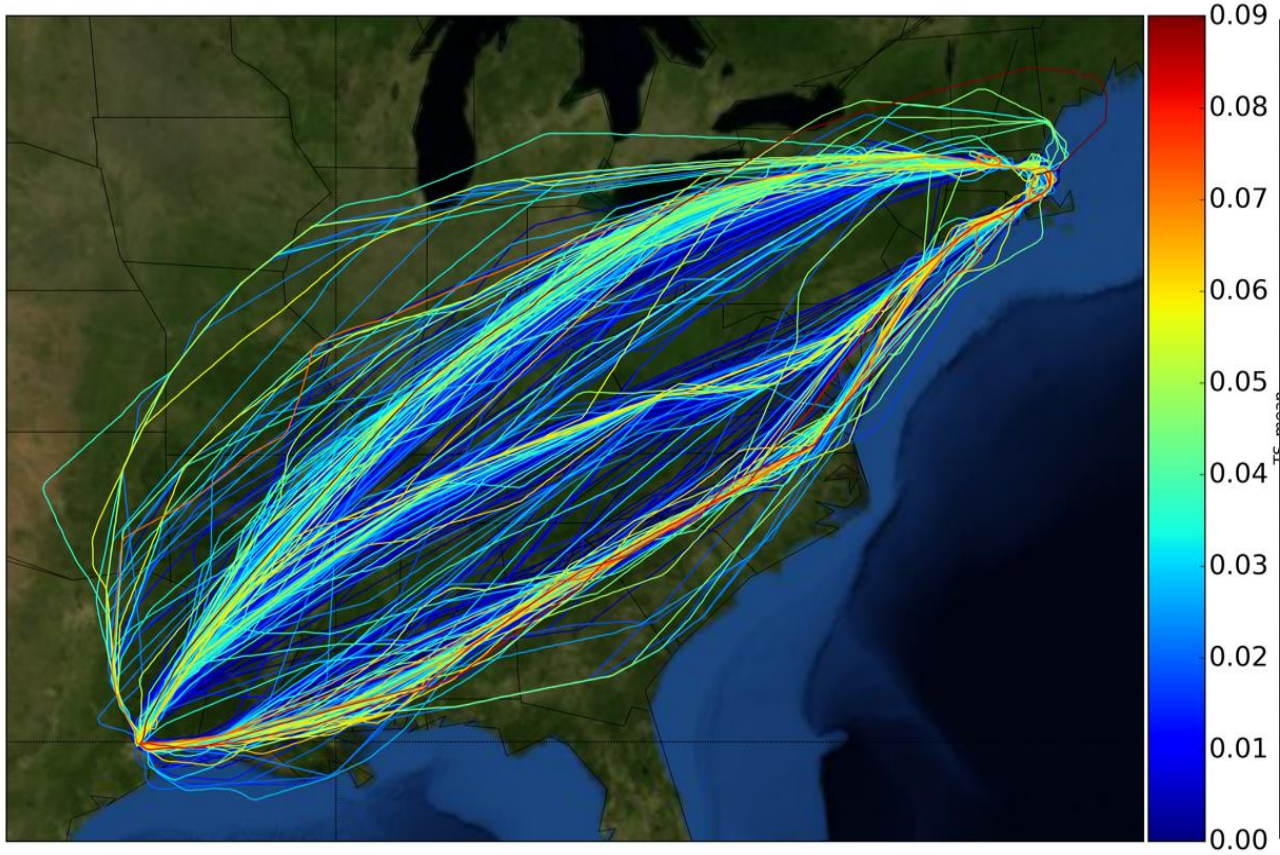
Method – on "Achieved distance"



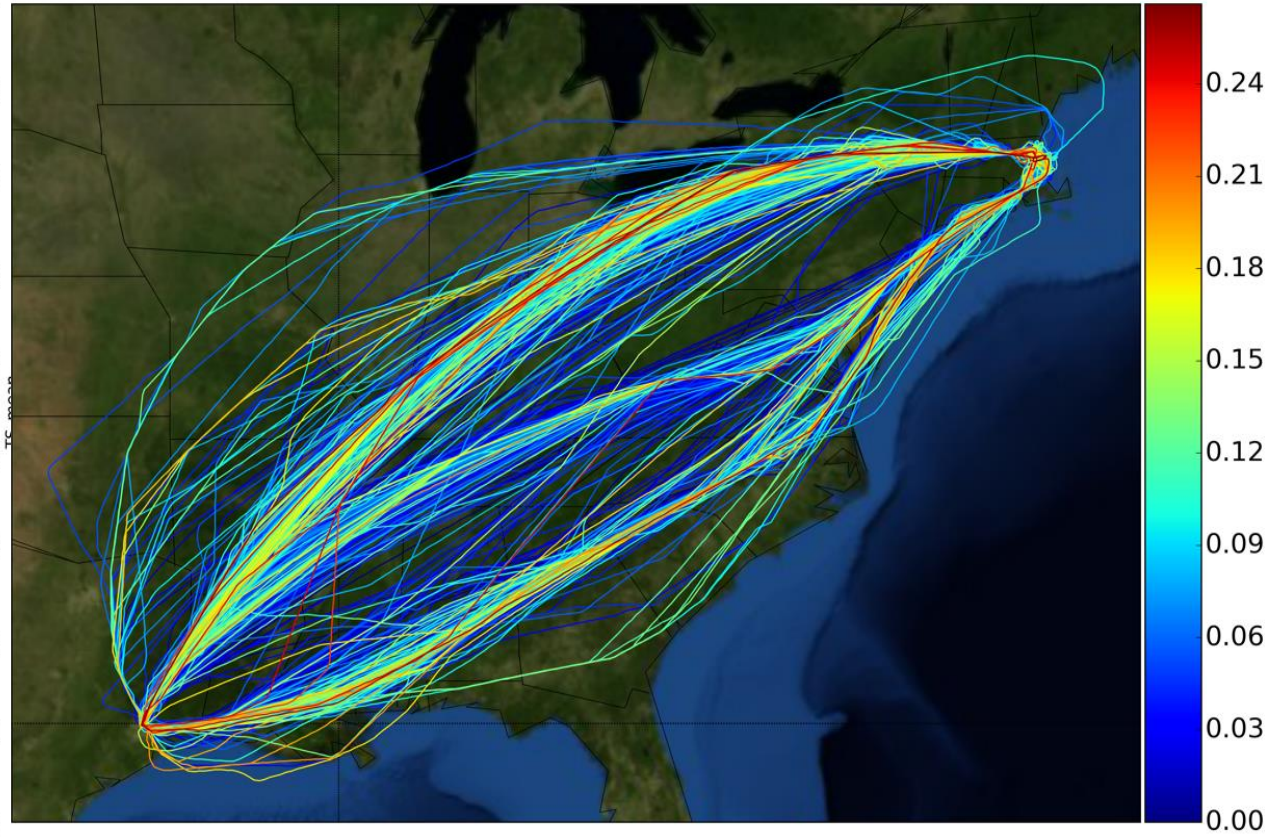
- $H = \frac{H_1 + H_2}{2}$
- Indicate how much closer is the ***Entry point*** to destination and how much further is the ***Exit point*** away from origin.

Composite Weather Exposure

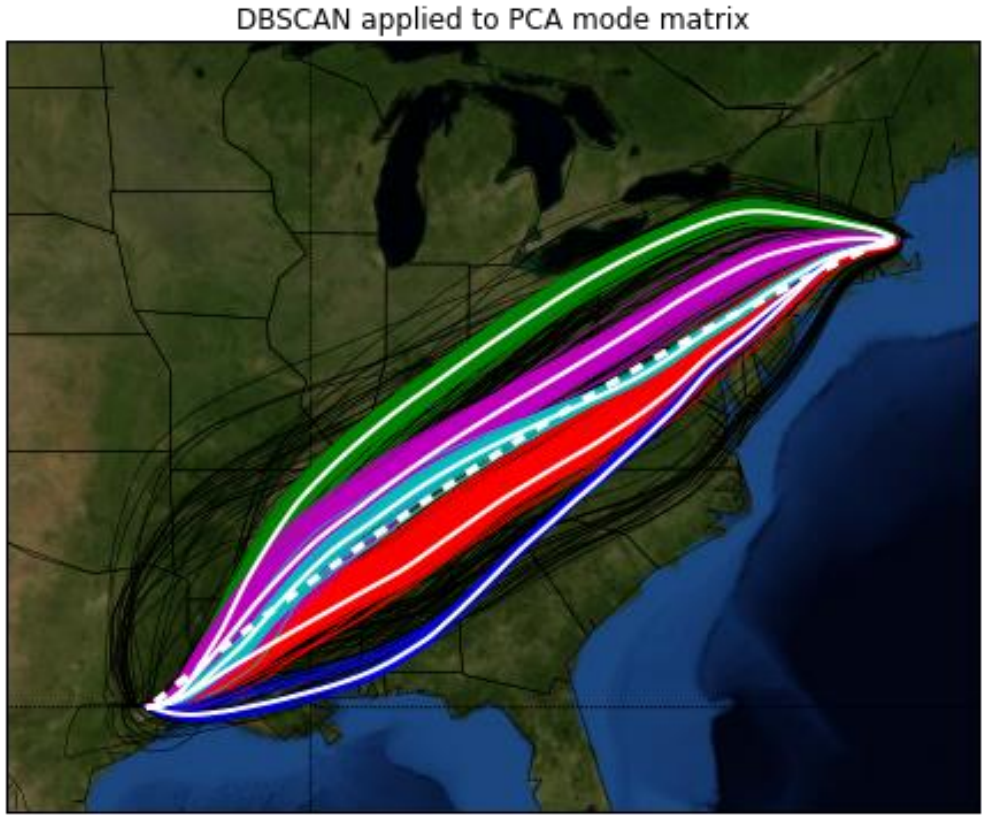
Thunderstorm Exposure

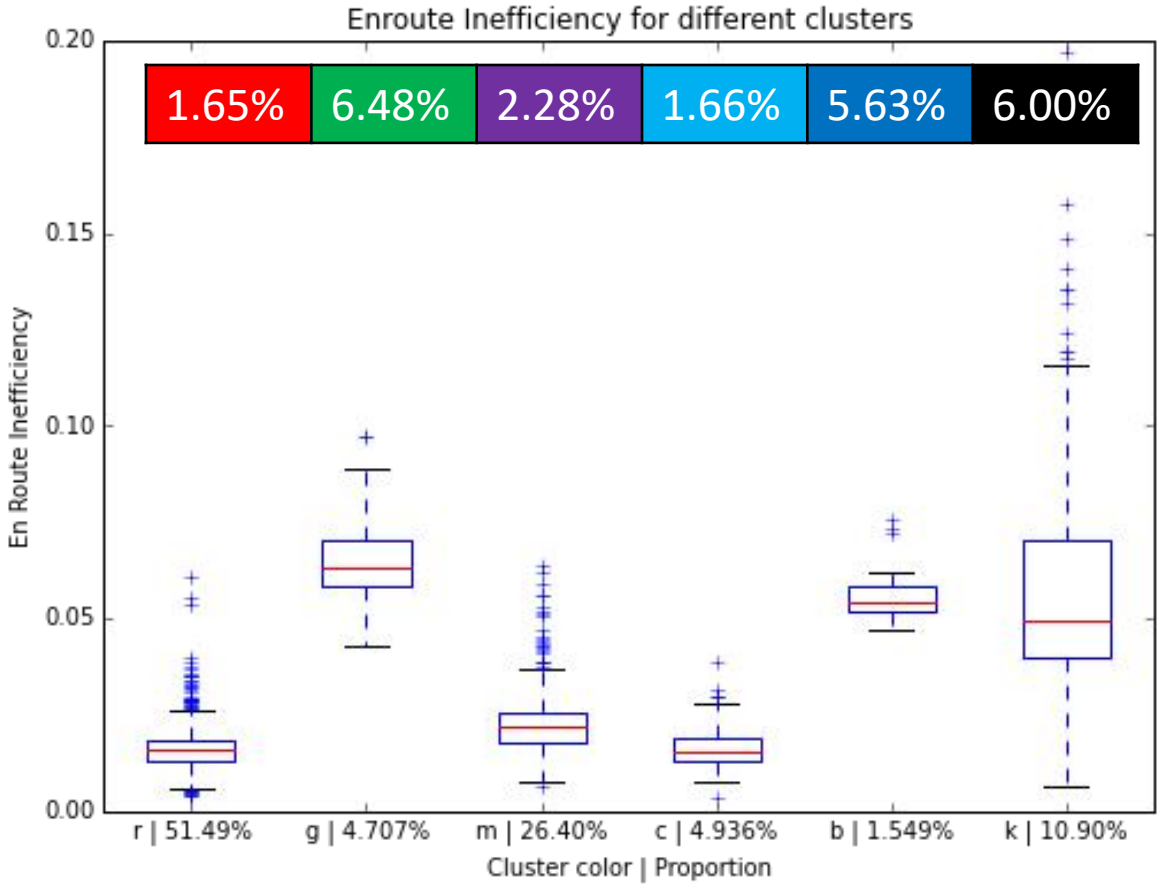


Rain Exposure



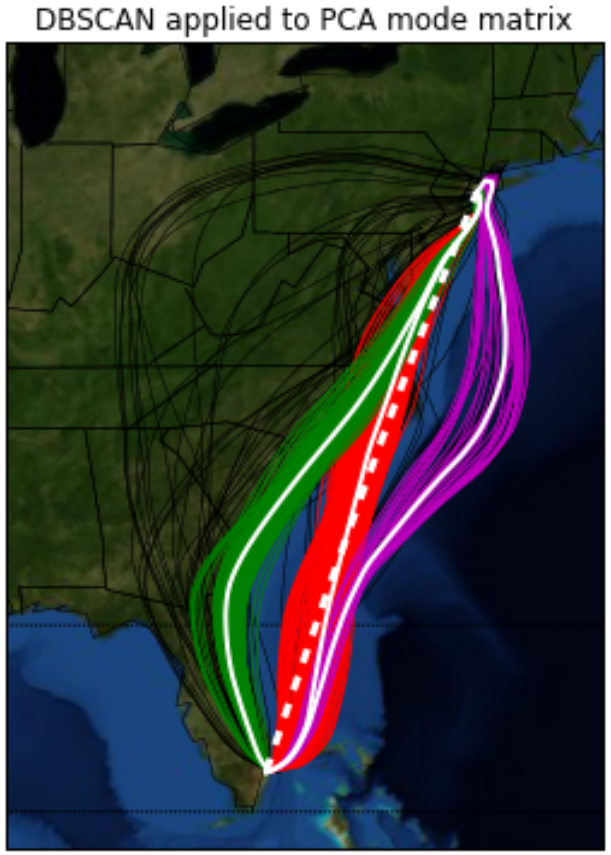
BOS → IAH (1742 of original 1883)

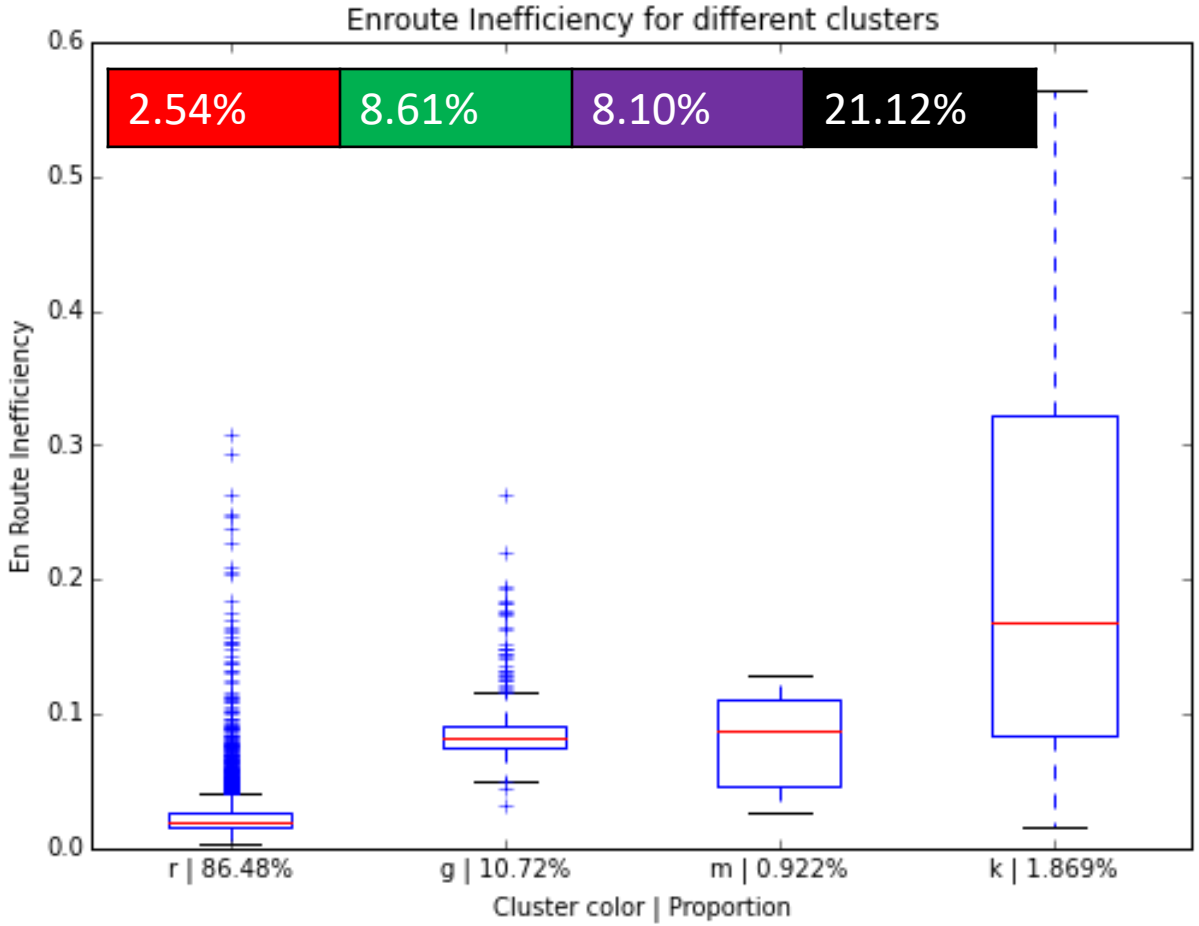




Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory

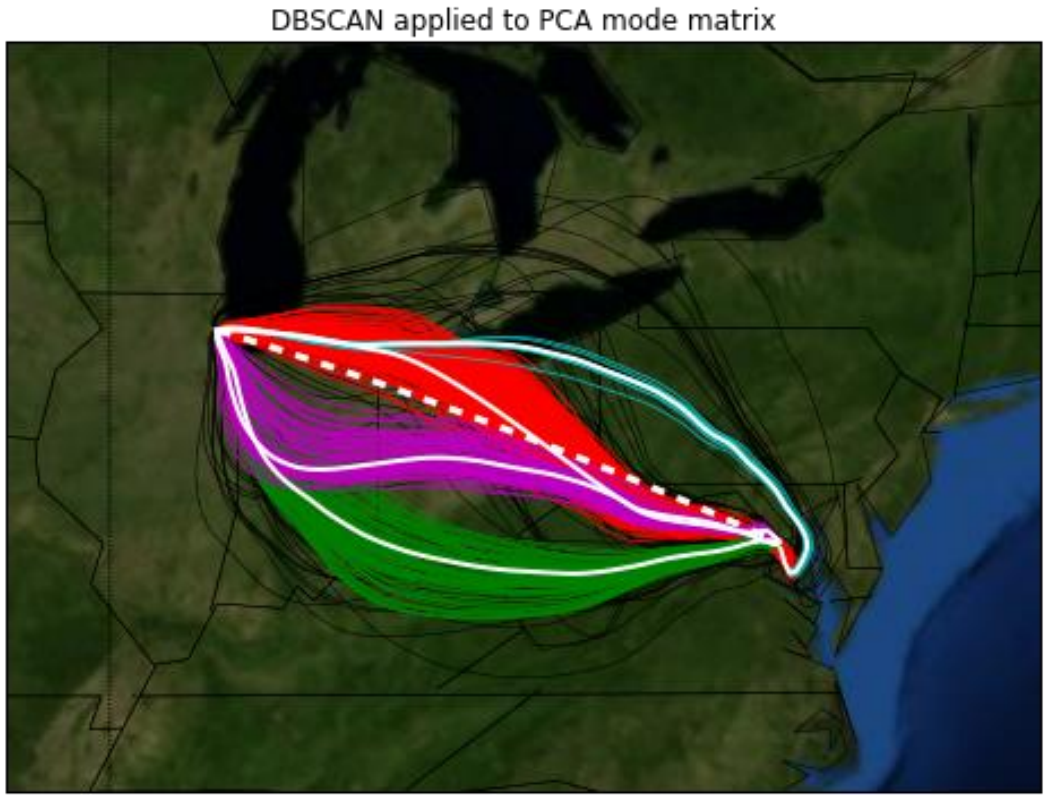
FLL → JFK (4011 of original 4267)

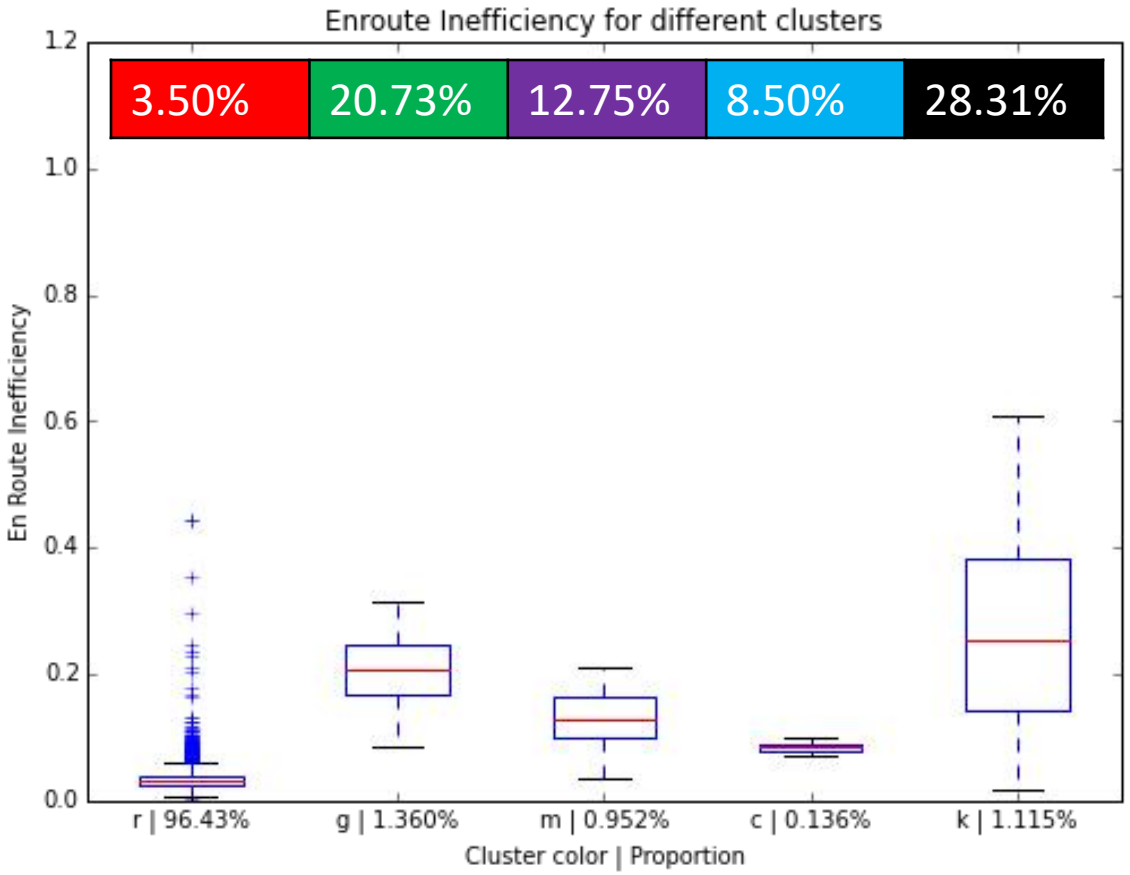




Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory

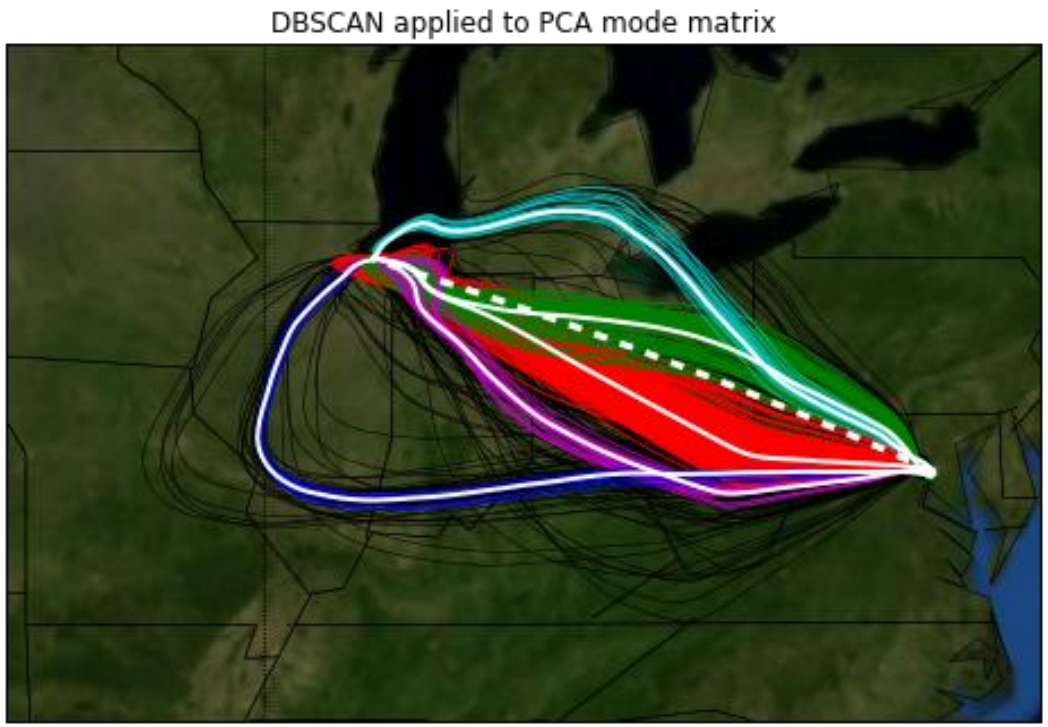
ORD → DCA (7349 of original 7574)

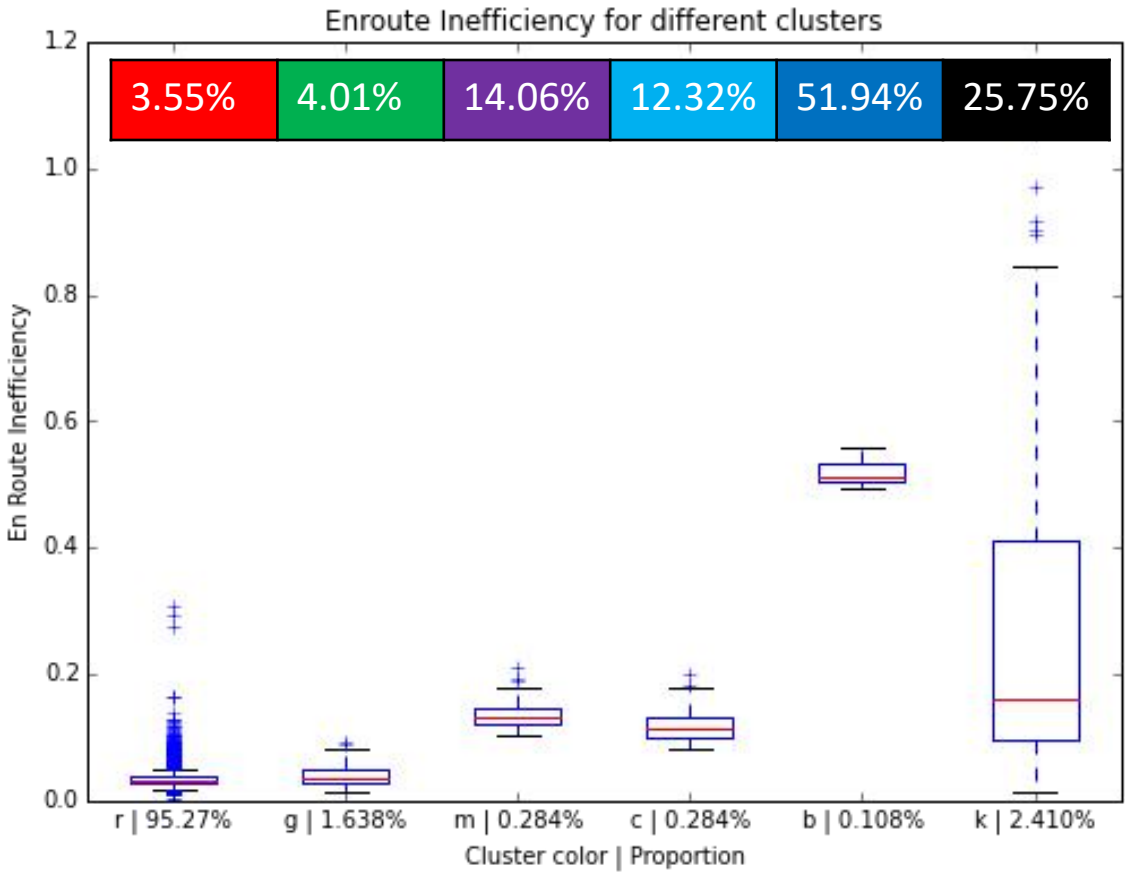




Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory

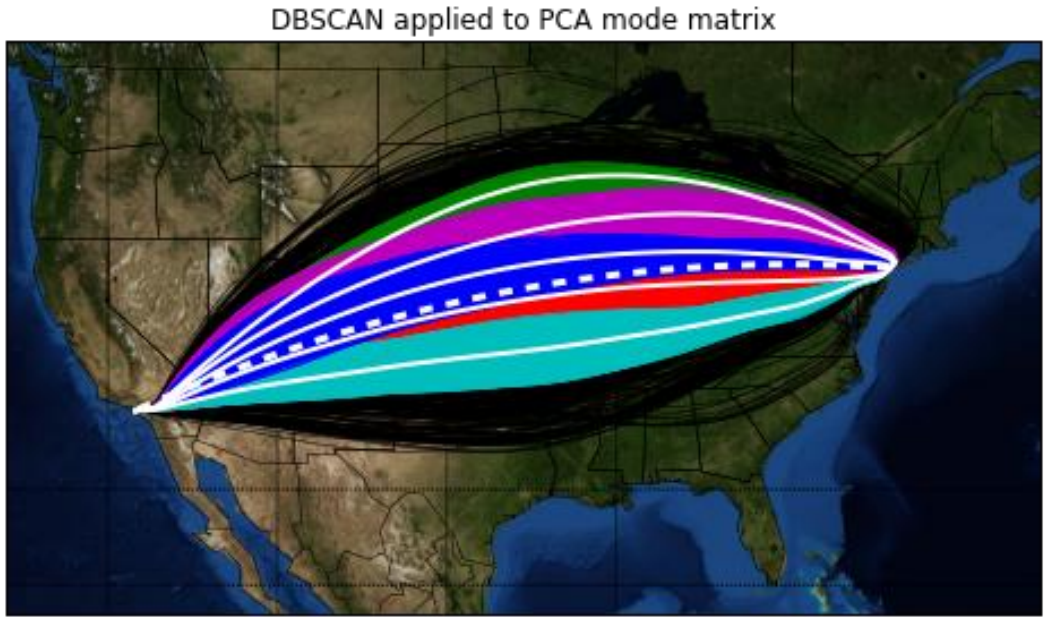
DCA → ORD (7383 of original 7557)

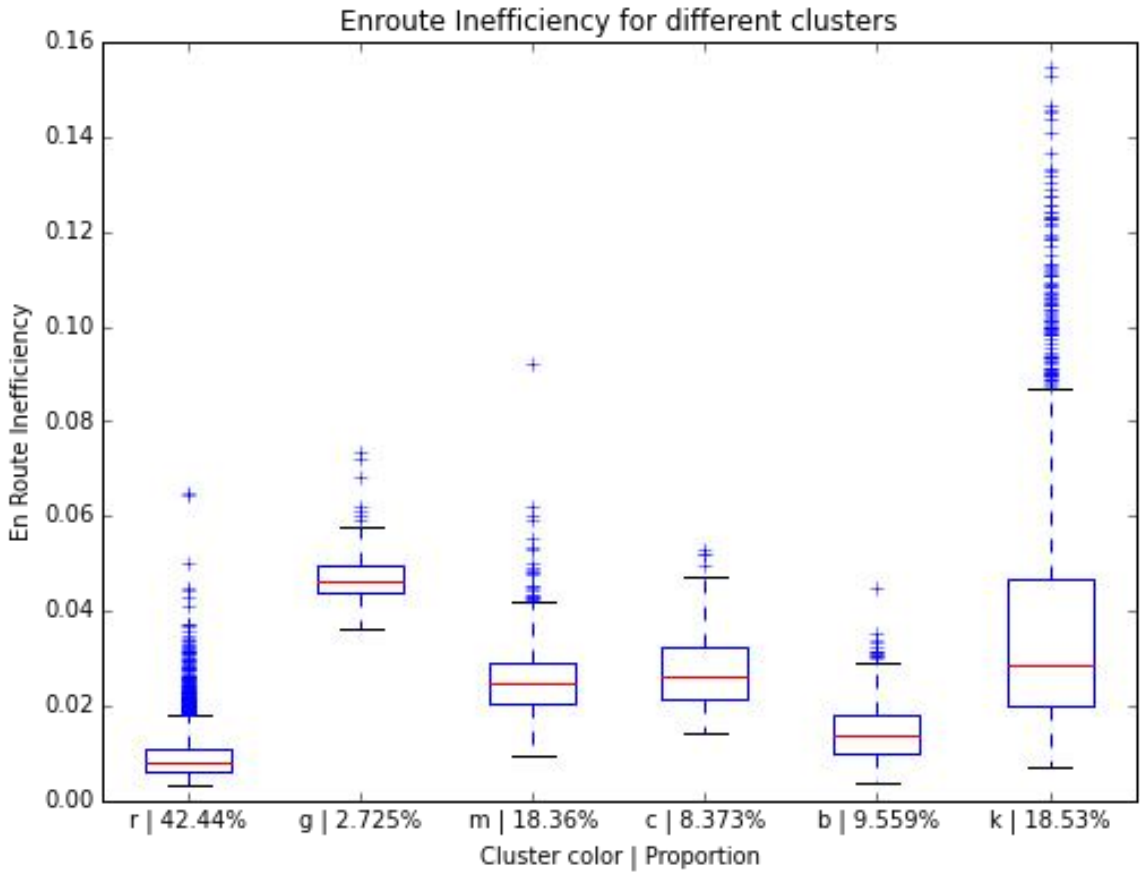




Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory

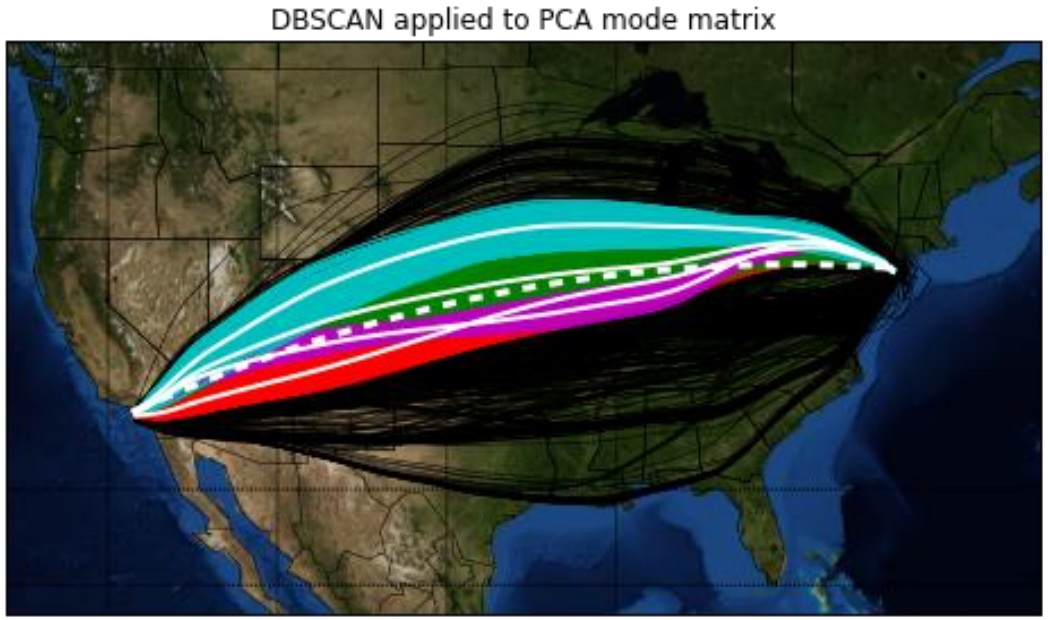
JFK → LAX (10725 of original 11586)

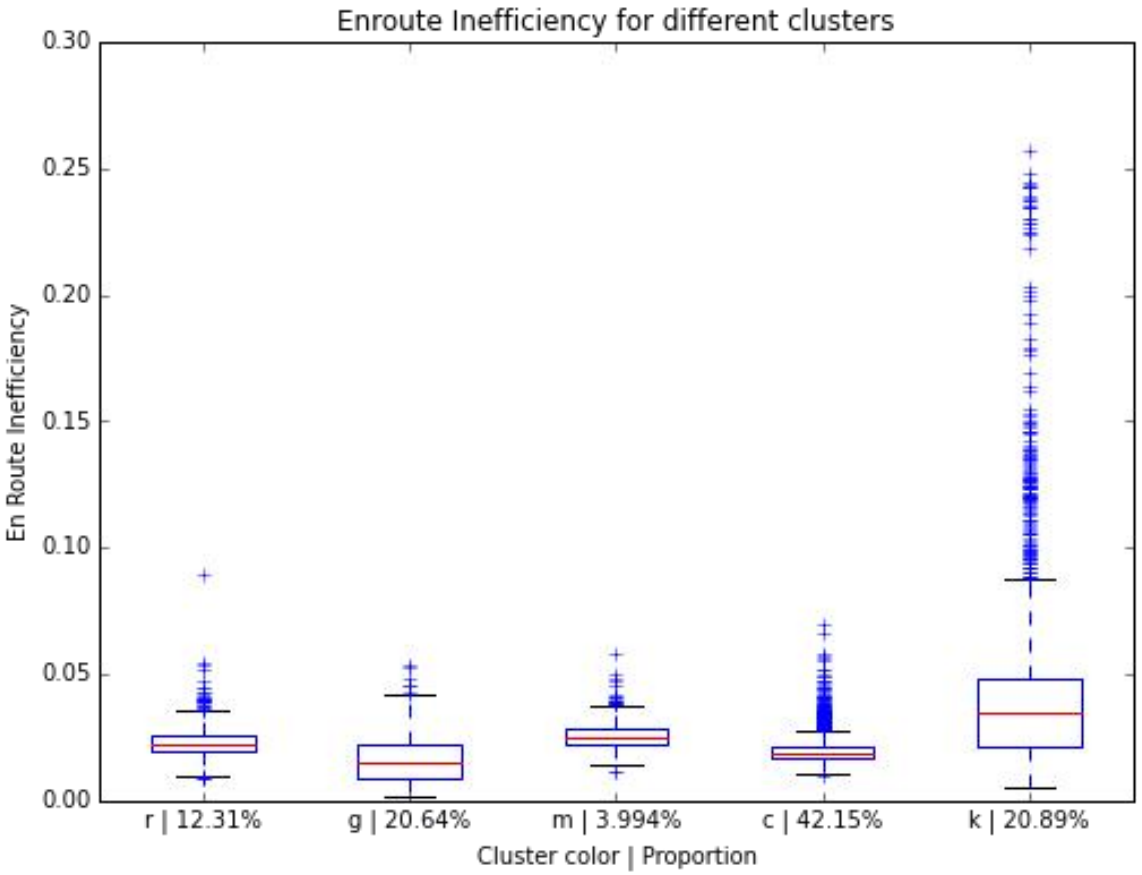




Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory

LAX → JFK (10447 of original 11543)





Black lines are classified as outliers
White Solid Lines are centers of each clusters
White Dashed line is great circle trajectory