Causal Analysis of En Route Flight Inefficiency in the US

Yulin Liu, Mark Hansen

University of California, Berkeley

Michael Ball, David Lovell, Cara Chuang

University of Maryland, College Park

John Gulding

Federal Aviation Administration

NEXTOR Research Symposium 10/26/2018

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

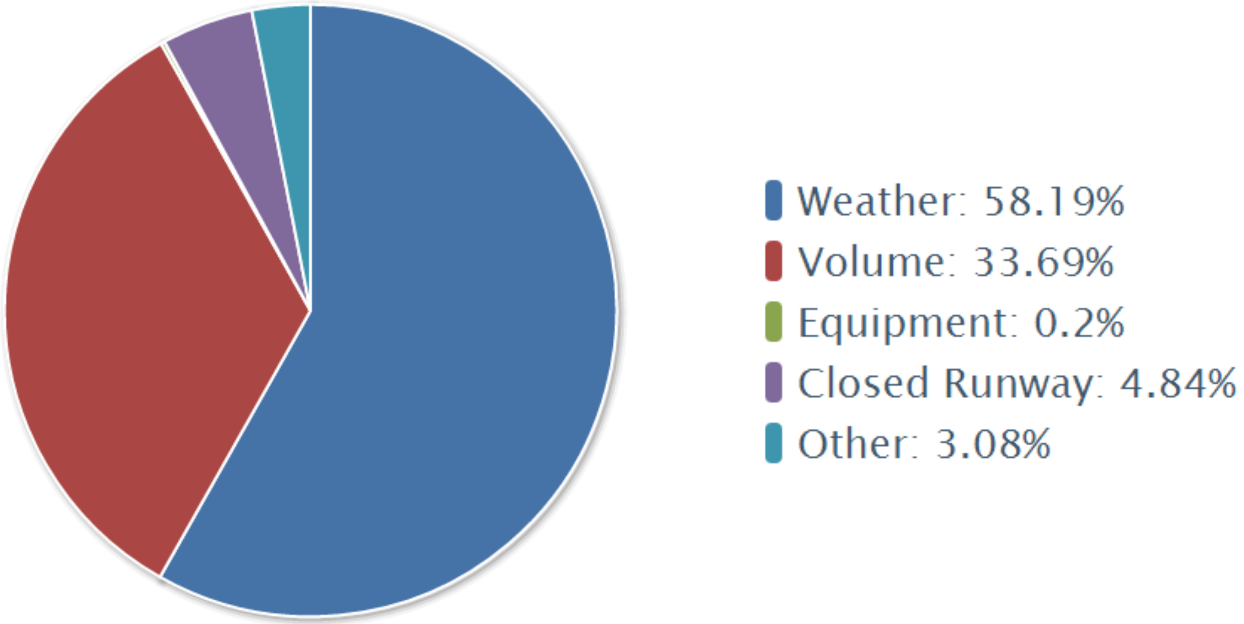
Motivations

- FAA and Eurocontrol have published metrics to evaluate flight en route inefficiency
- Limited understanding of the causal factors behind the inefficiency
- For arrival delay we have:

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

Motivations

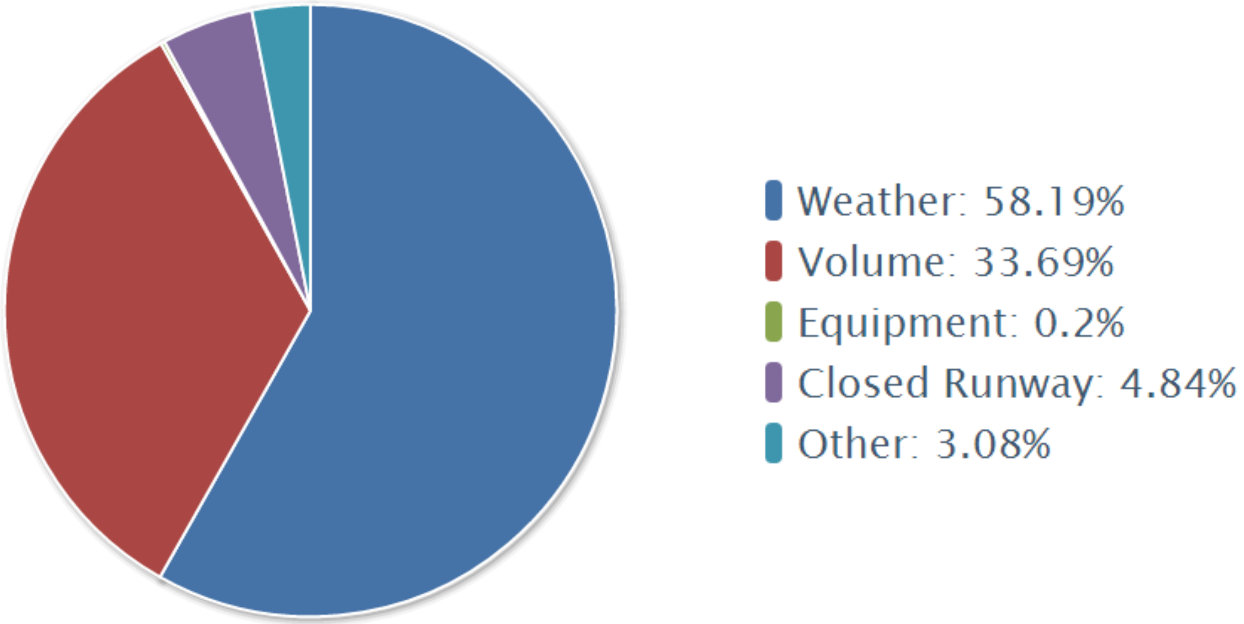
- FAA and Eurocontrol have published metrics to evaluate flight en route inefficiency
- Limited knowledge of the causal factors behind the inefficiency
- For arrival delay we have:



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

Motivations

- FAA and Eurocontrol have published metrics to evaluate flight en route inefficiency
- Limited understanding of the causal factors behind the inefficiency
- For arrival delay we have:
- What about **en route inefficiency**?



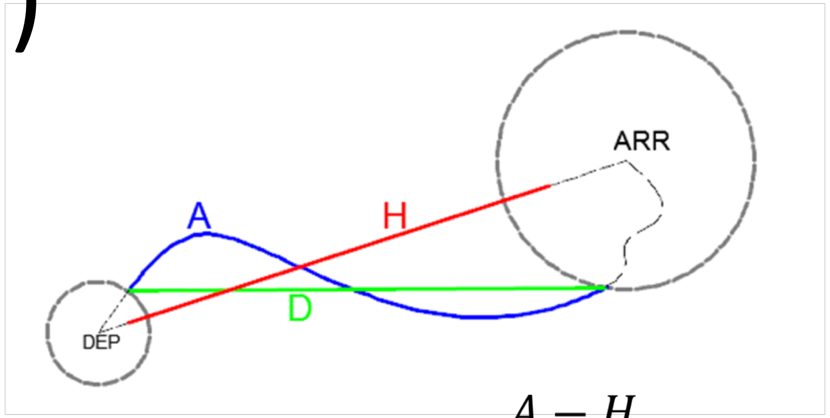
Sources:
http://www.transtats.bts.gov/ot_delay/ot_delaycause1.asp?type=5&pn=1

Project Goal: Quantify Contribution of Causal Factors to En Route Inefficiency

- Convective weather
- Winds
- Traffic Management Initiatives (specifically, MIT restrictions)

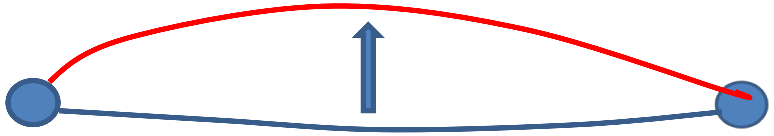
Approach (I)

- Inefficiency is measured by the ground distance of a flight trajectory relative to the "achieved" distance

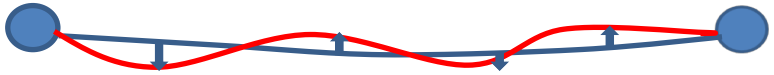


$$Inefficiency = \frac{A - H}{H}$$

- Two types of causal mechanisms
  - Impact on the flight route



  - Impact on the flight trajectory



Approach (II)

- Apply trajectory clustering algorithm to raw trajectory data for selected OD pairs to identify clusters and construct nominal routes
- Map nominal routes with convective weather, wind and Miles-in-Trail (MIT) data
- Estimate and apply statistical models to quantify the contributions of convection, wind and MIT
  - Route assignment model
  - Trajectory inefficiency model

Limitations

- Do not consider flight operations within the terminal area
- Focus on six city pairs as case studies
  - IAH↔BOS
  - FLL↔JFK
  - LAX↔JFK
- Inefficiency metric assumes that the ideal route is great circle
- Winds are considered as a source of “inefficiency”—a longer distance route may be taken because of favorable winds

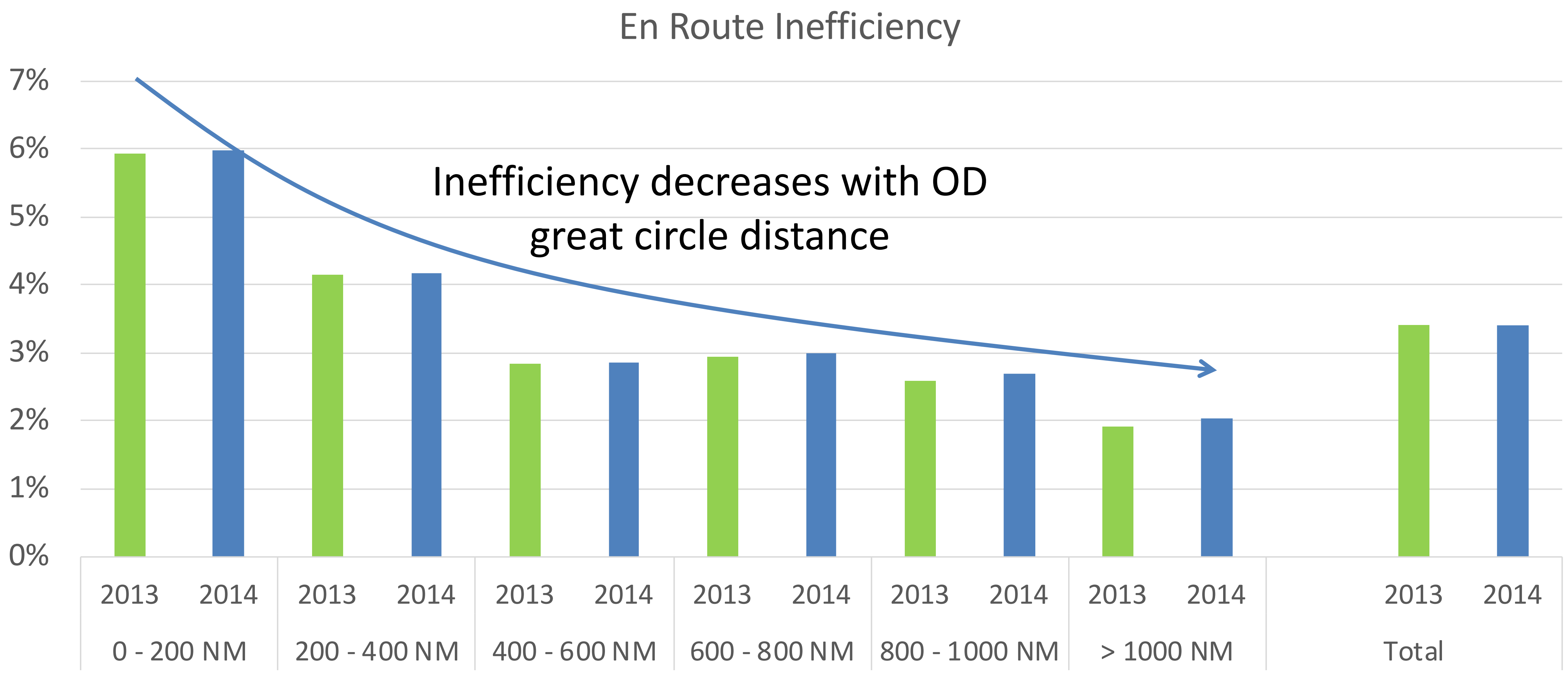
Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

Data Sources

- Traffic Flow Management System (TFMS)
  - Flight Event Data
    - Flight level records, including en route inefficiency, aircraft type and etc.
  - Flight Track Data
    - 60-second update
    - Currently we focus on six pairs: IAH $\longleftrightarrow$ BOS, JFK $\longleftrightarrow$ LAX and FLL $\longleftrightarrow$ JFK
- National Traffic Management Log (NTML)
  - Miles-In-Trail (MIT) Data
- Quality Controlled Local Climatological Data (QCLCD)
  - Hourly summary of convective weather (ground-based)
- North American Mesoscale (NAM) data
  - Hourly wind forecast data
  - $0.1° \times 0.1° \times 25\ mbar$

En Route Inefficiency vs Great Circle Distance



Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

Finding Nominal Routes

- We define *Nominal Routes* as the set of representative trajectories for a given OD pair
- Nominal routes help us better understand the NAS route structure
- Finding nominal routes allows time-dependent variables such as TMIs and convective weather to be calculated in a more efficient manner
- We estimate the ***hypothetical*** exposure of a particular flight to weather, MIT restrictions, and wind, if it ***had*** used any one of the nominal routes, assuming its ***actual*** departure time.

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.

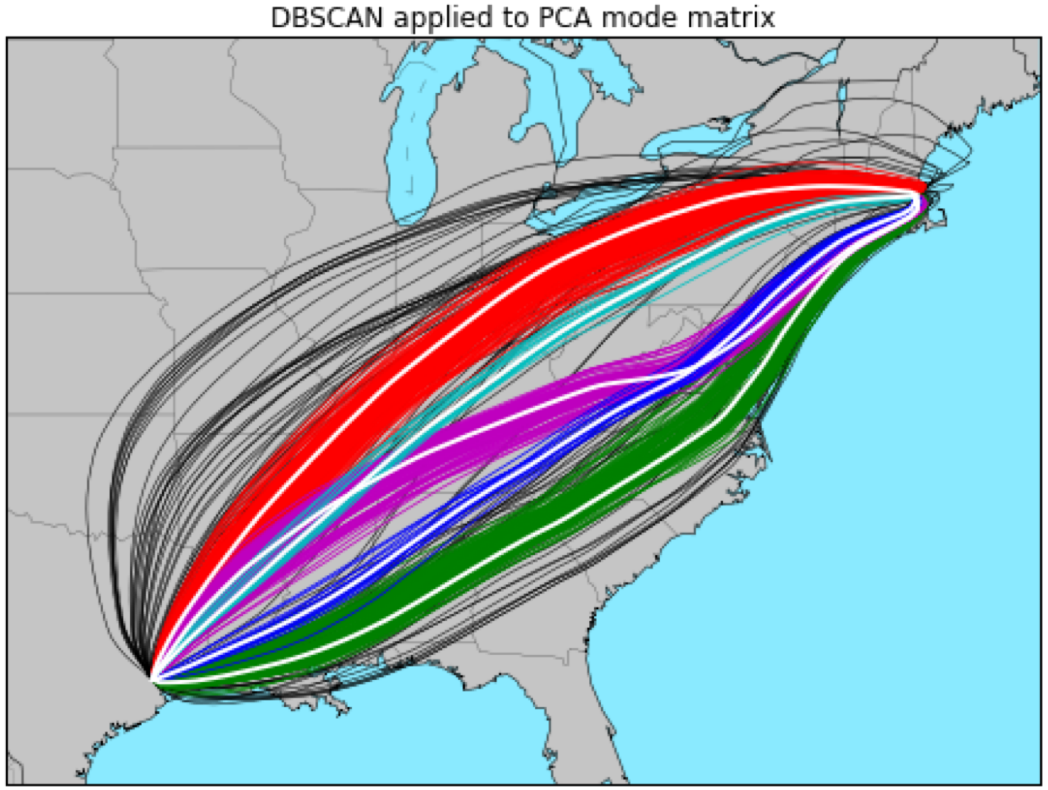
- ***Step 2: Principal Component Analysis (PCA)***
  - Dimension reduction & Trajectory smoothing;
  - First five components can capture more than 90% of variations.

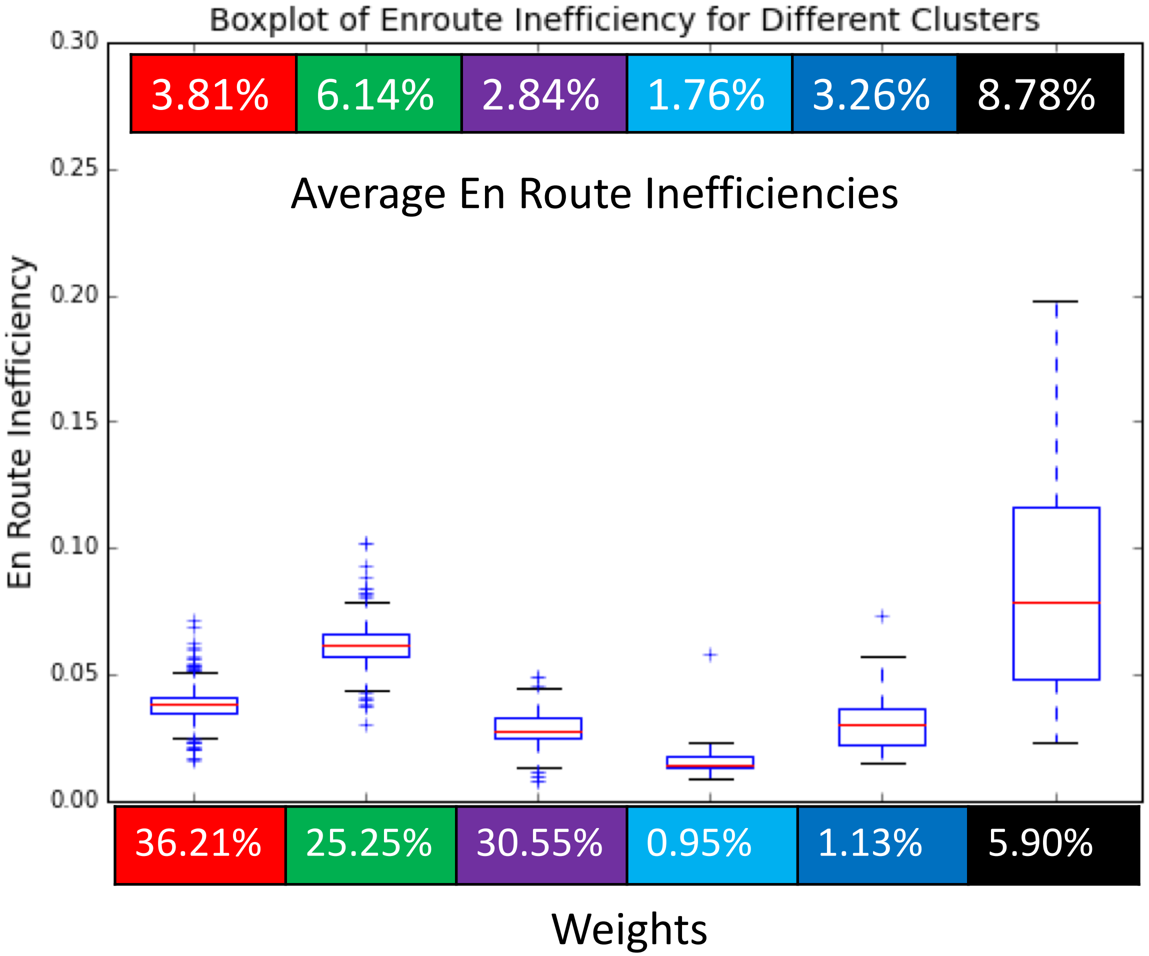
- ***Step 3: DBSCAN Clustering***
  - Trajectory classifications;
  - DBSCAN algorithm is applied to the PCA components to get representative clusters.

From Clusters to Nominal Trajectories

- Each cluster contains multiple trajectories
- For each cluster we find the trajectory that has the smallest total dissimilarity (based on PCA) to the other trajectories in the same cluster (p-median problem)
- The selected trajectories are termed the “nominal trajectories” and represent their clusters in the subsequent analysis

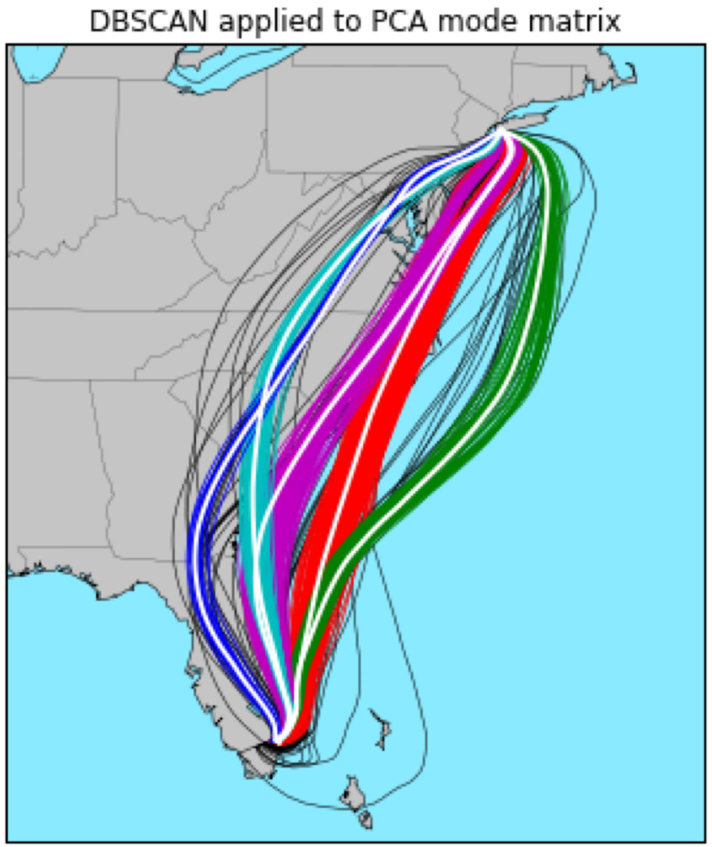
IAH → BOS (1679 of original 1817)

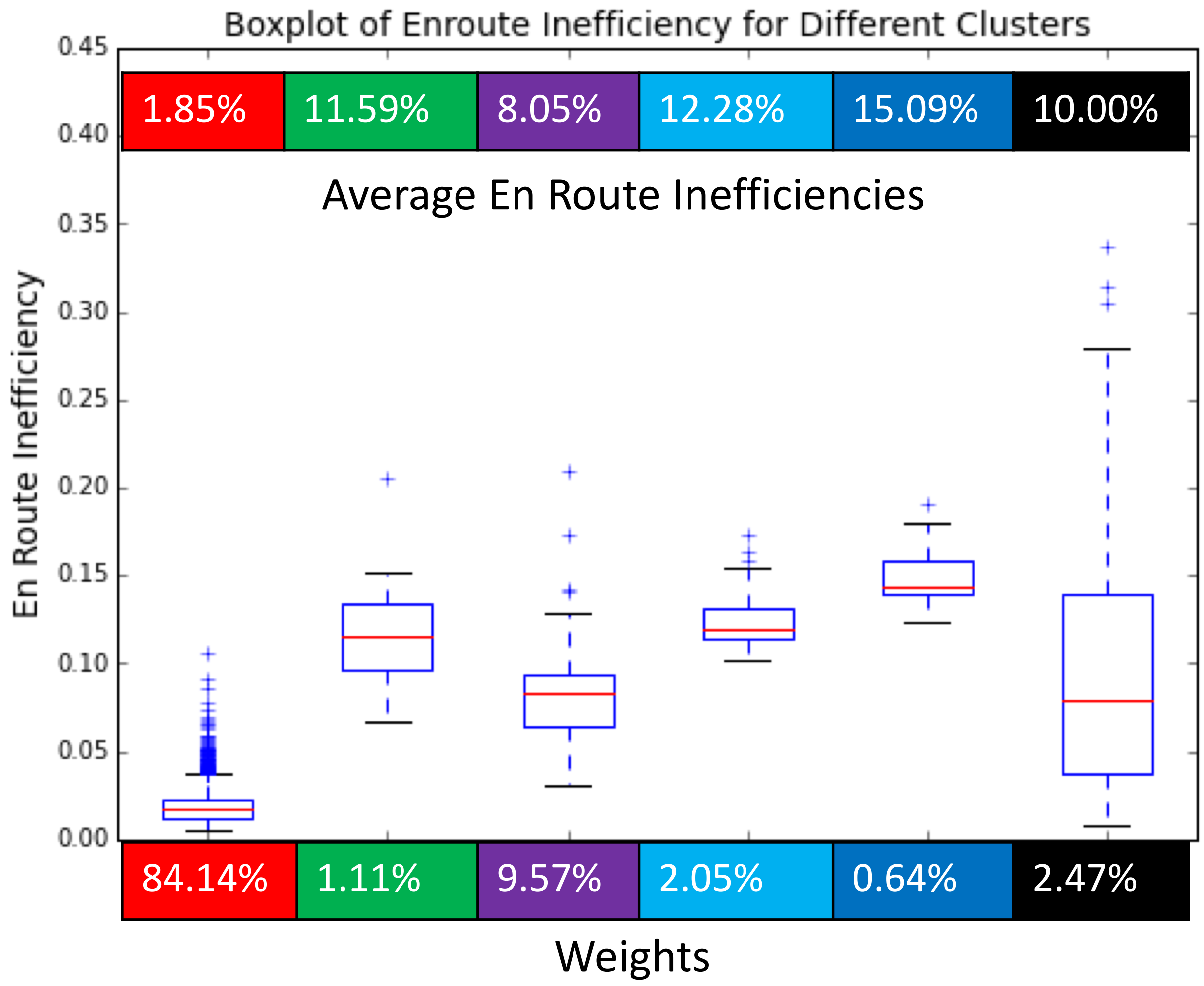




Black curves are classified as outliers
White Solid curves are Nominal Routes

JFK → FLL (4043 of original 4273)





Black curves are classified as outliers
White Solid curves are Nominal Routes

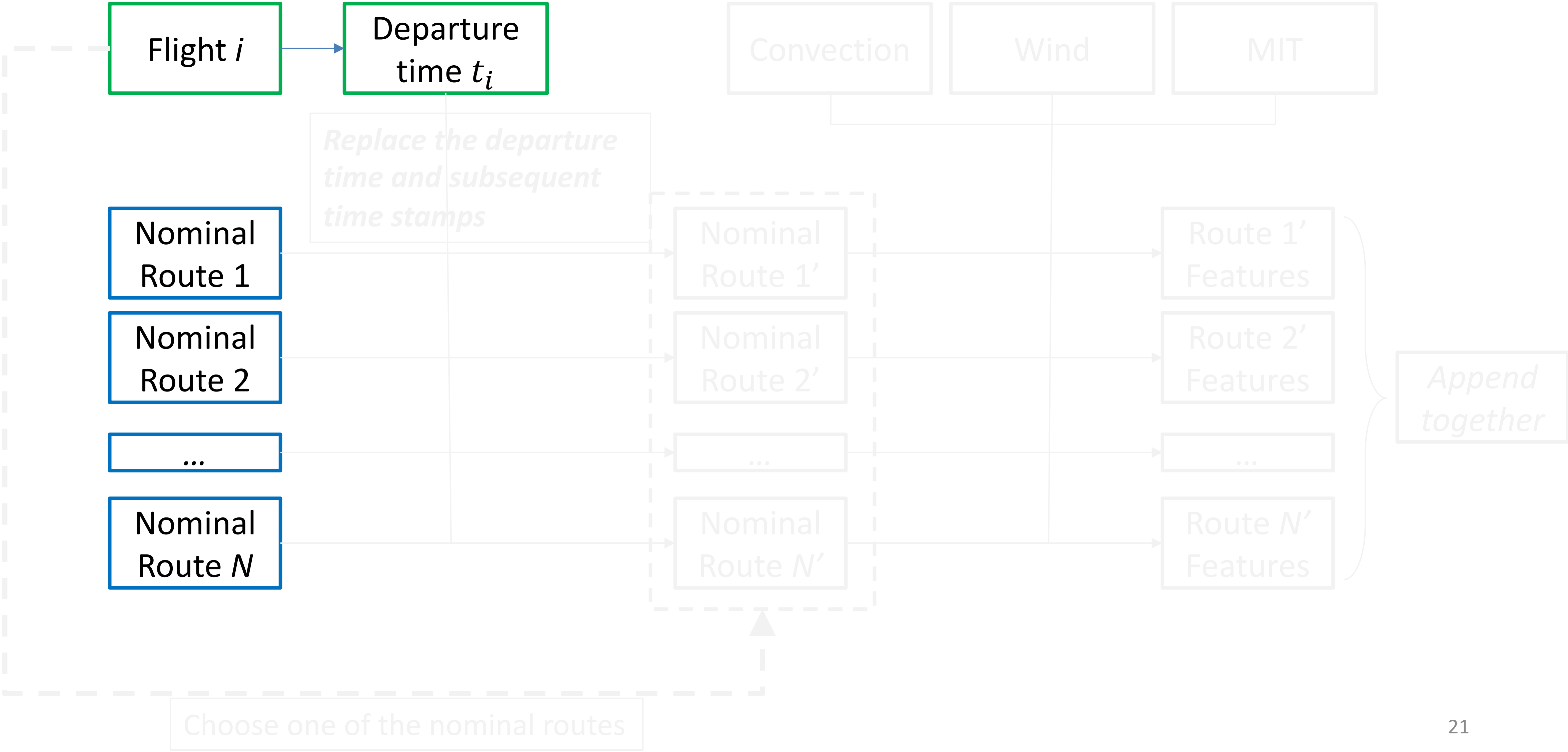
Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

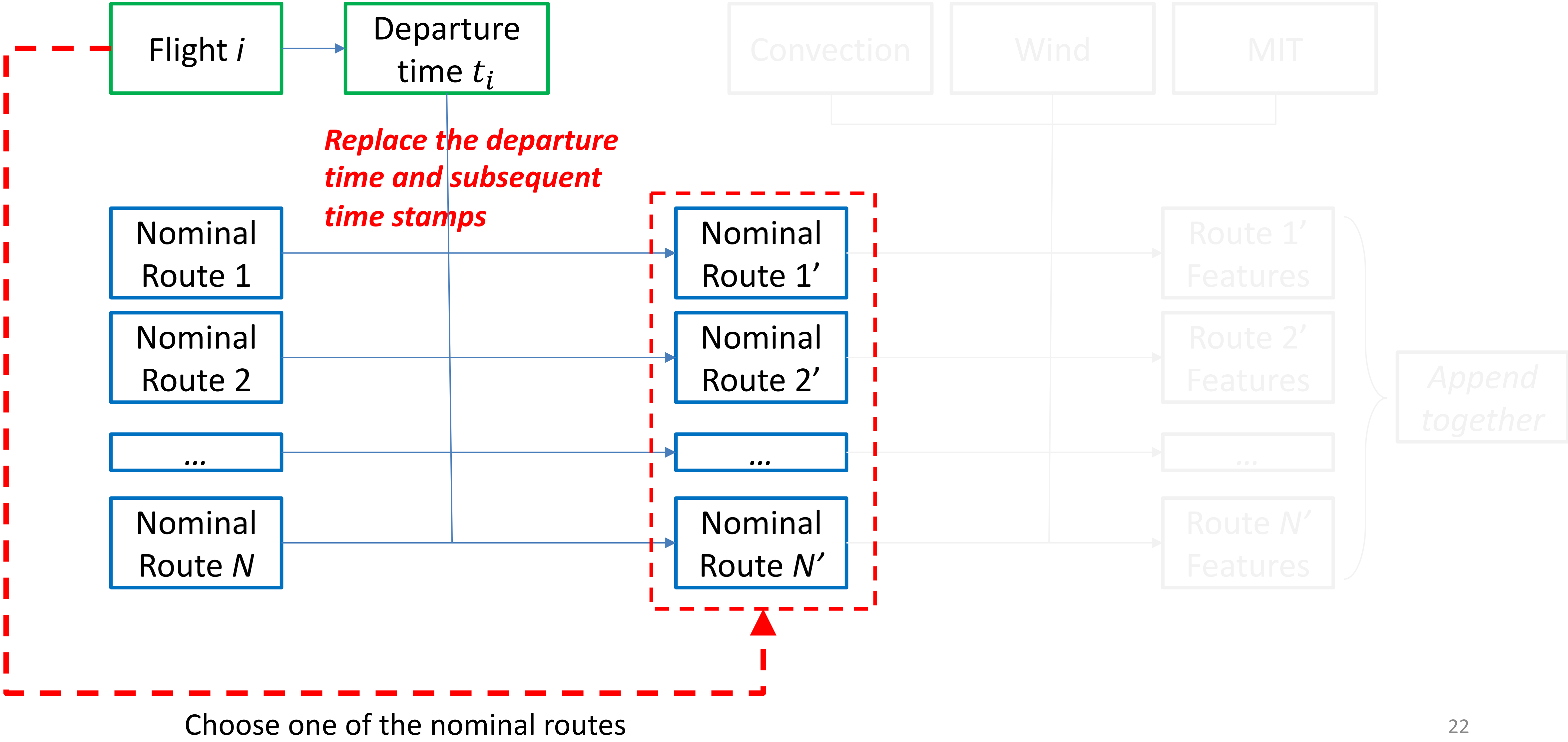
Purpose of Mapping Causal Factors

- Consider any specific flight between a given OD pair, e.g. UA 1578 from IAH to BOS departing 4:50 pm on 7/7/2014
- Determine conditions that flight would have encountered if it had followed any of the nominal trajectories between IAH and BOS
- Conditions include:
  - En route weather
  - Winds
  - Miles in Trail Restrictions

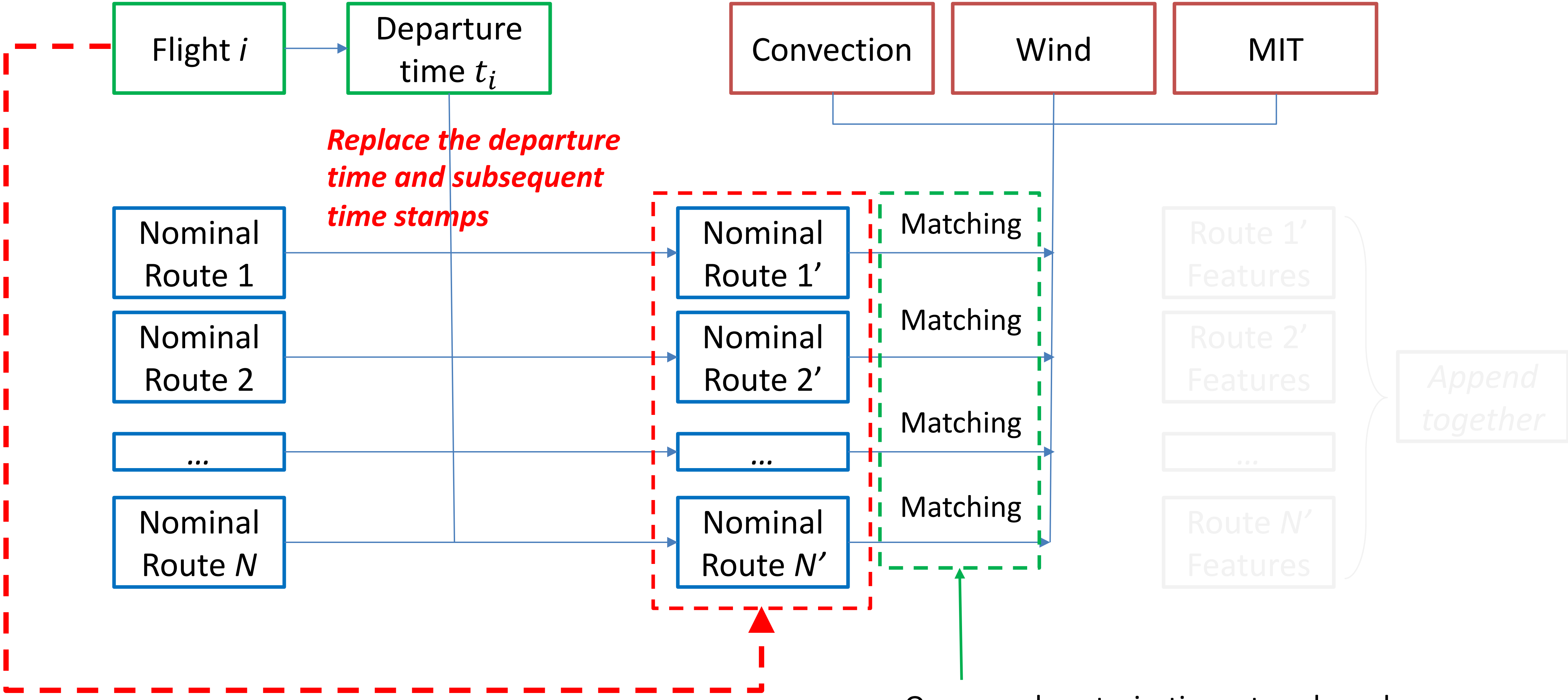
Mapping Causal Factors (I)



Mapping Causal Factors (II)



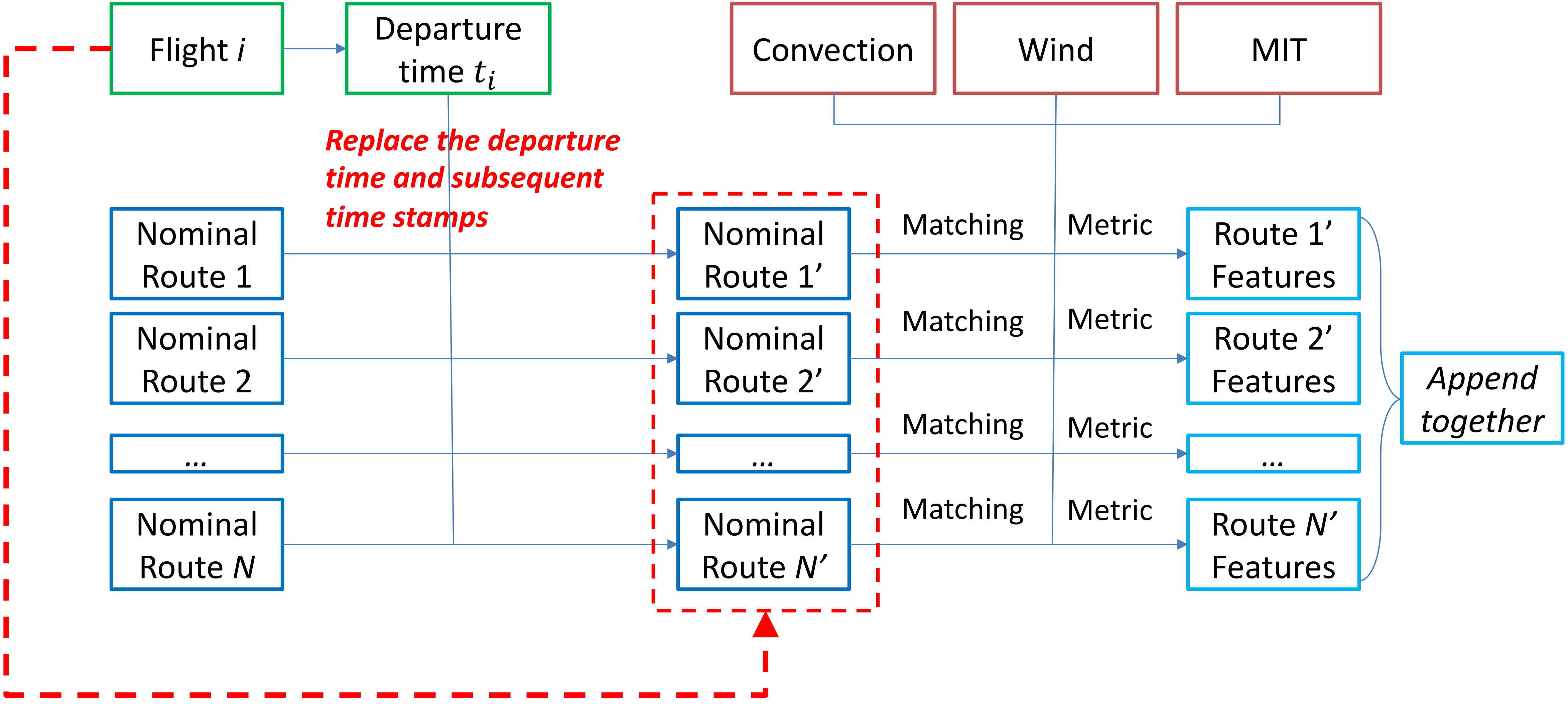
Mapping Causal Factors (III)



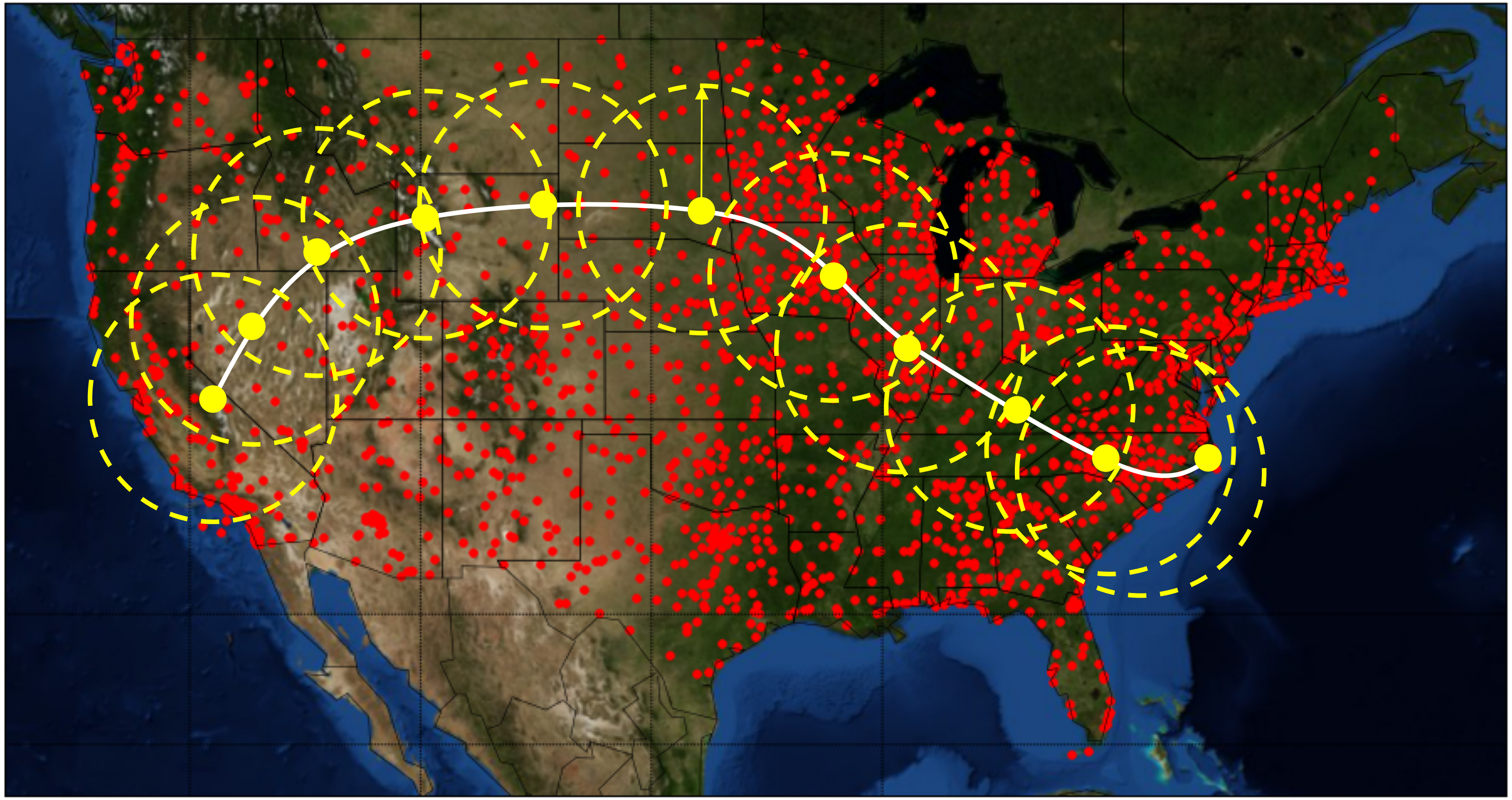
Choose one of the nominal routes

Query-and-vectorization – tree-based query
Time-space trade-off – batch mode operation

Mapping Causal Factors (IV)

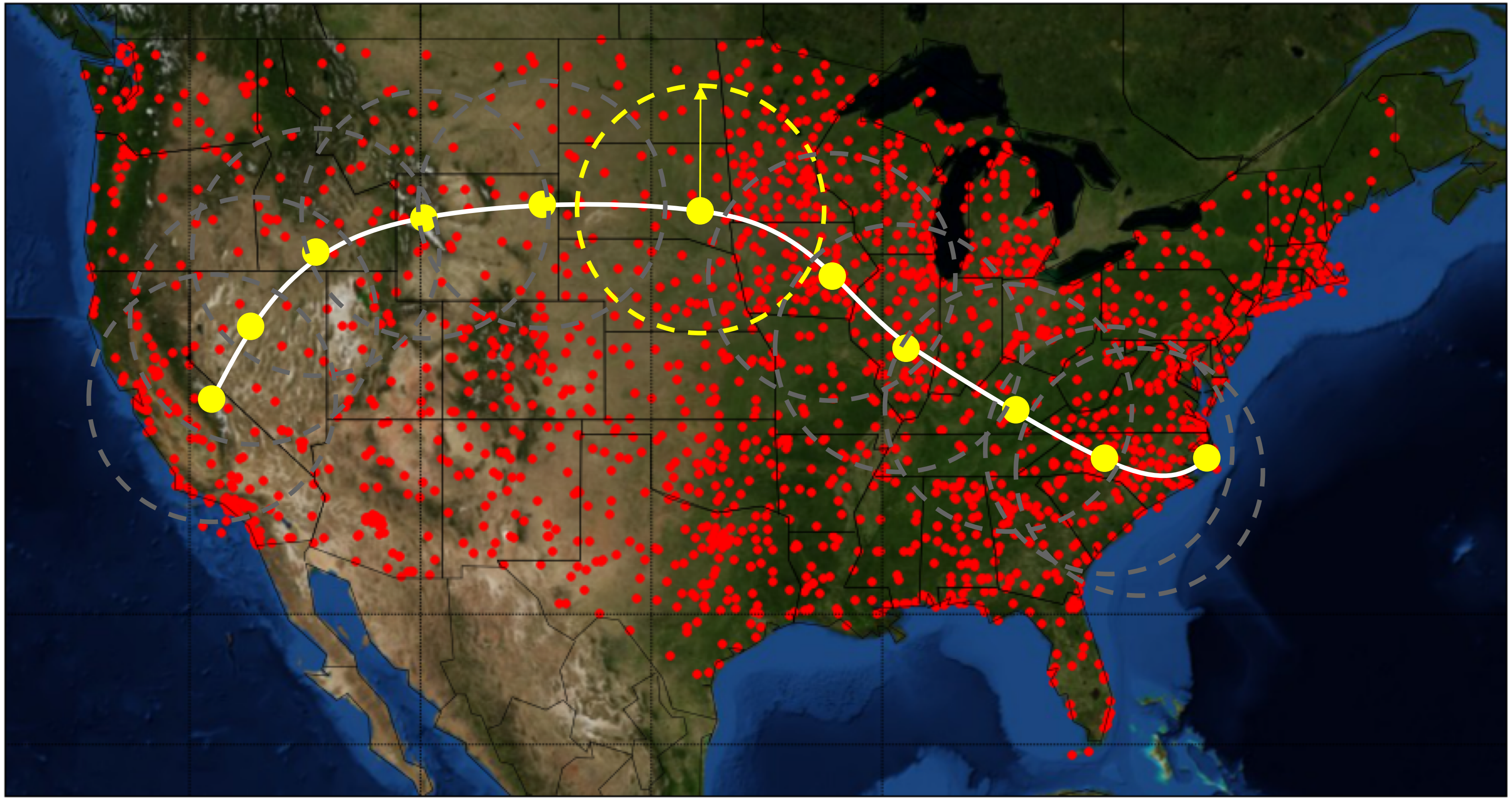


Matching – Convective Weather



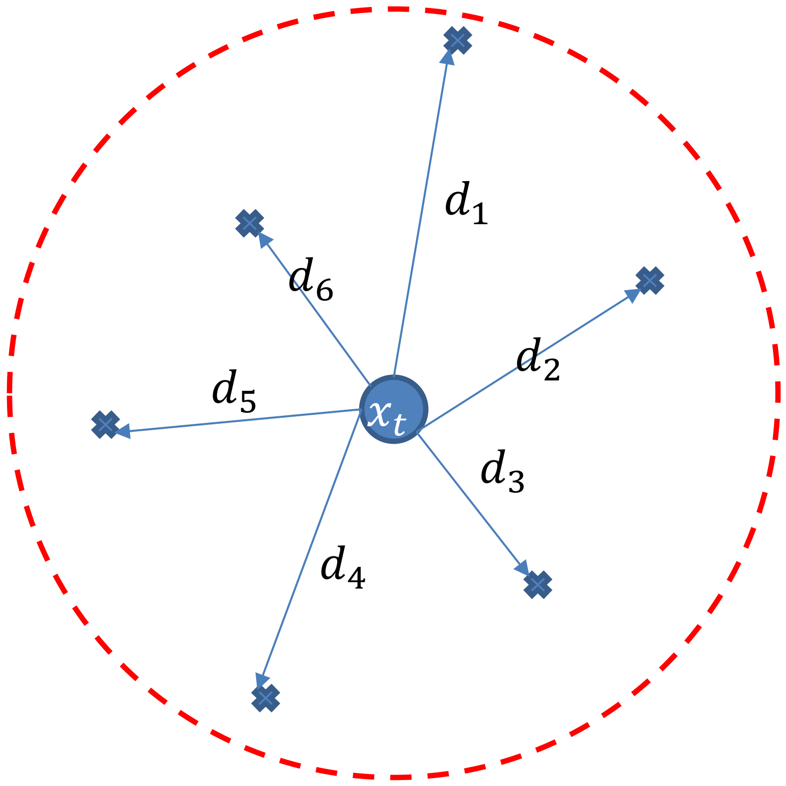
- For each track point on the trajectory, find all the stations within a circle with radius r (150 nmi).

Matching – Convective Weather



- Weighted average the weather variable (binary) for stations within the circle, and the weight is the proportional to the inverse of the distance.
- Metric: average of the weather exposure for all track points along the route.

Matching – Convective Weather



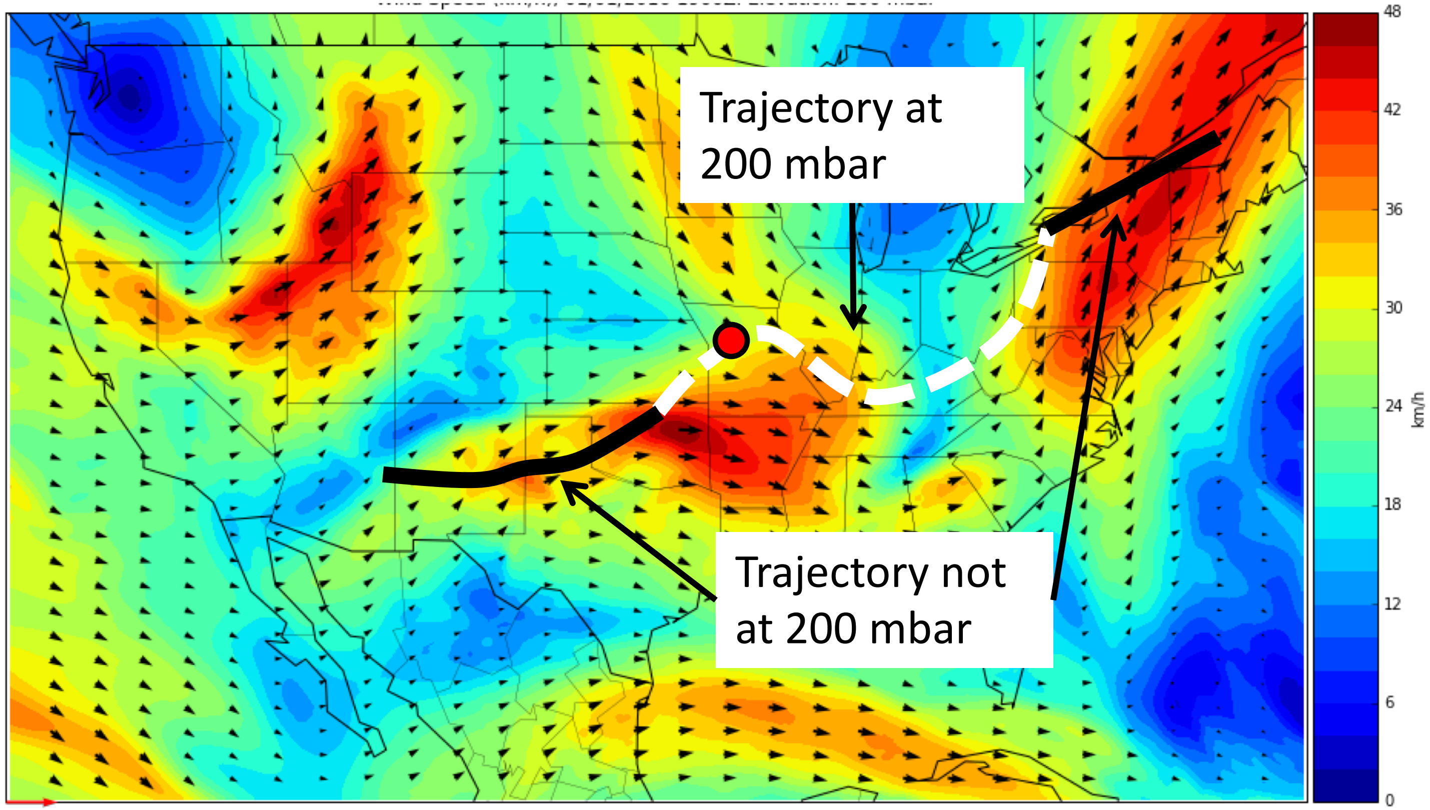
$$Wx_t = \sum_i I_i \cdot \frac{d_i}{\sum_j d_j}$$

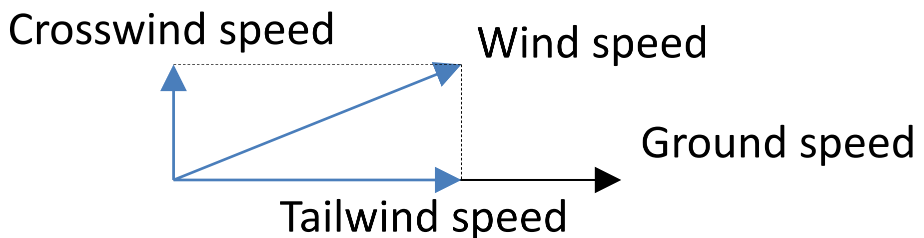
$$Wx = \frac{1}{T} \cdot \sum_t Wx_t$$

- Weighted average the weather variable (binary) for stations within the circle, and the weight is the proportional to the inverse of the distance.
- Metric: average of the weather exposure for all track points along the route.

Matching – Wind

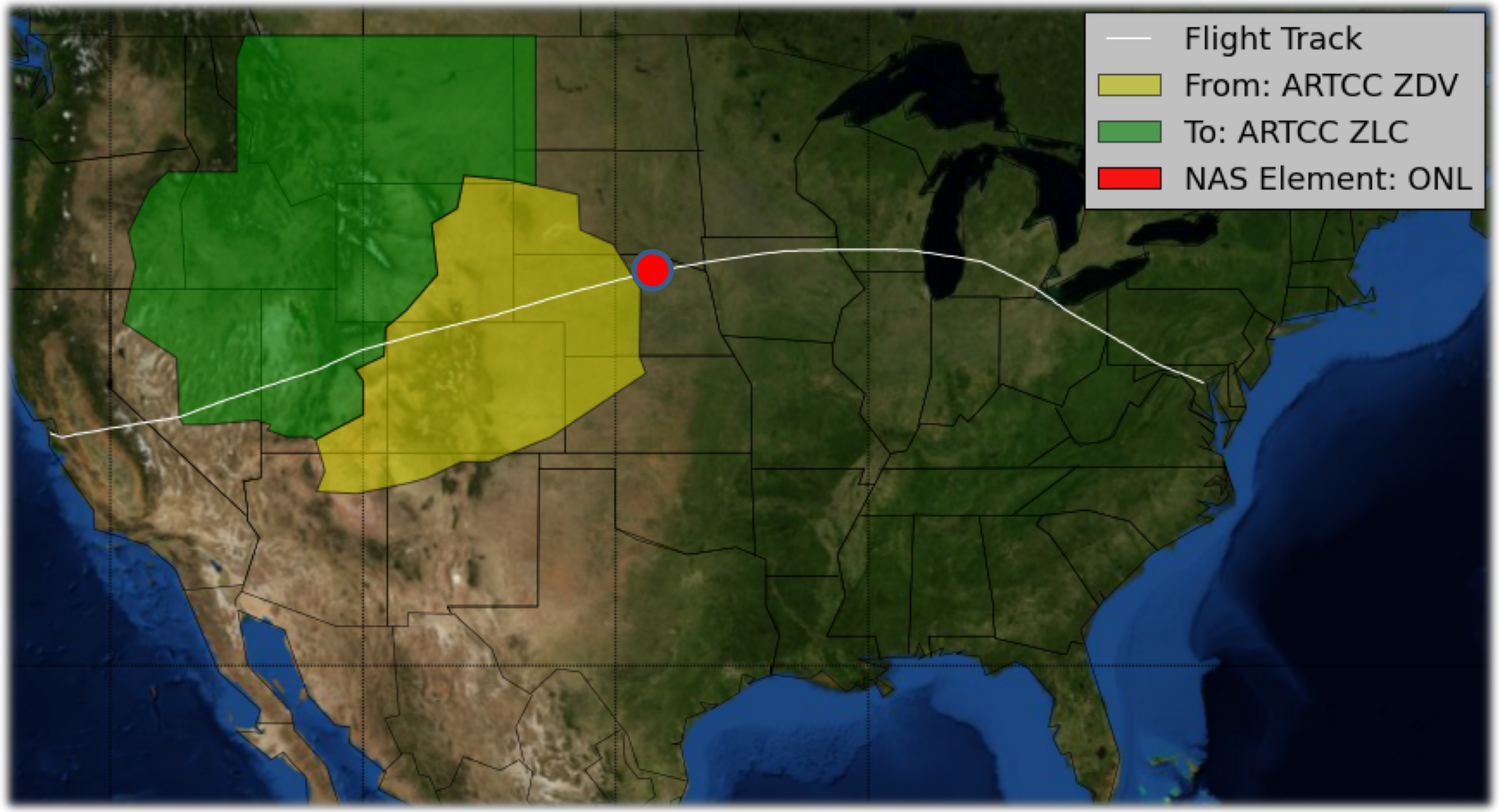
Wind Field Diagram (m/s) @ 200 mbar (~ 38,000 ft.)
02/04/2013 18:00 Zulu





- For each track point, find the nearest 4d reference point of the wind data file
- Assign the wind speed (vertical and horizontal) of the nearest grid to the track point
- Calculate the headwind/tailwind speed for each track point, based on heading derived from previous track point

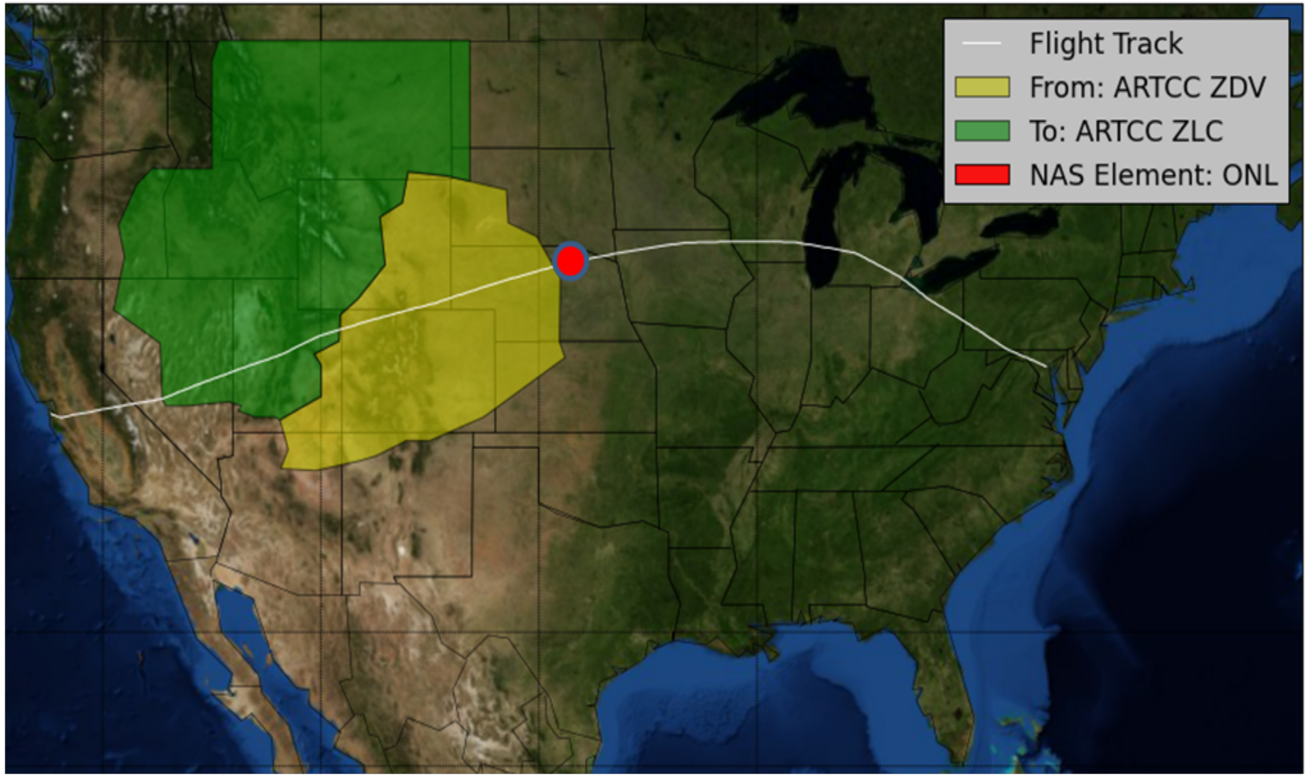
Matching – MIT



- Miles-in-trail specifies the minimal required distance between two consecutive aircrafts.
- Apportion traffic into a manageable flow.

ZDV is trying to protect *ZLC*, which is overloaded, by providing a MIT (e.g., 15 miles) to separate aircrafts through the Navaid *ONL*.

Matching – MIT



- A given nominal route, with adjusted departure time, is assumed to be affected by an MIT if:
  - It crosses the MIT facilities
  - It crosses the NAS element or follows the jet route to which the MIT applies
  - Its crossing time is within the time the MIT is in effect
  - Its crossing altitude is covered by the restriction

Metrics

- Convective weather
  - Average percentage of weather stations along the nominal route reporting weather phenomena of each of the three types – thunderstorm, rain and squalls
  - Range: 0 (no convective weather along route) – 1 (convective weather at every weather station along route)
- Wind
  - Still air distance: The summation of the product of airspeed and time along the nominal route, where the airspeed is calculated by summing the ground speed and the headwind/tailwind speed
  - Range (for IAH → BOS): 1021 nmi. – 1564 nmi.
- MIT
  - Summation of the MIT stringency imposed by all MIT restrictions along the nominal route, where MIT stringency is defined as the product of MIT value (in miles) and MIT duration (in hours)
  - Range (for IAH → BOS): 0 – 2800 mi. x hr.

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

Impact of Causal Factors on Flight En Route Inefficiency

- Objective: Quantify contributions of different causal factors to flight inefficiency
- Causal factors to be considered
  - Weather
  - MIT
  - Wind
- Causal mechanisms
  - Nominal route assignment
  - Flight inefficiency for a given nominal route

Model

Flight *i*

Departure time

Weather exposure/wind/MIT on nominal routes

Probabilistic assignment of nominal routes (Route choice model)

Weighted Average

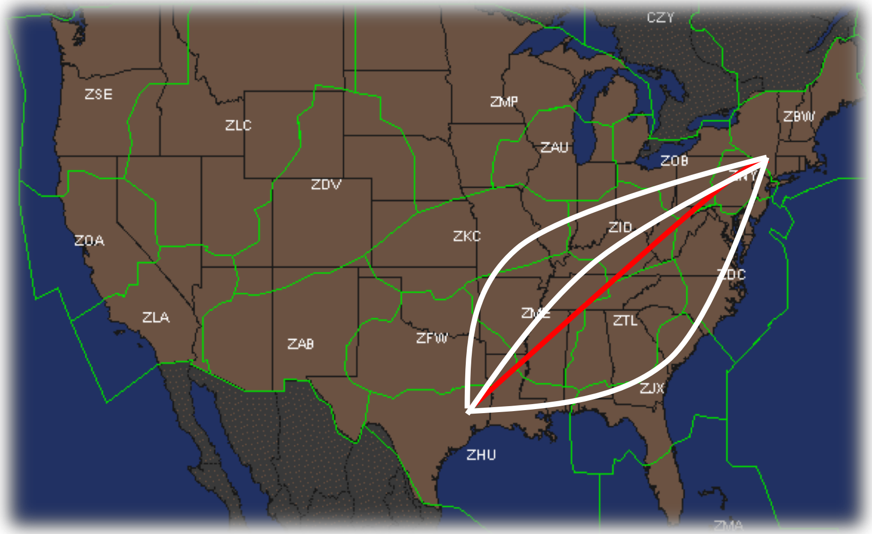
Morning flight indicator

Composite weather exposure/ MIT

Predicted inefficiency under **actual** conditions

Route inefficiency for each nominal route (Linear model)

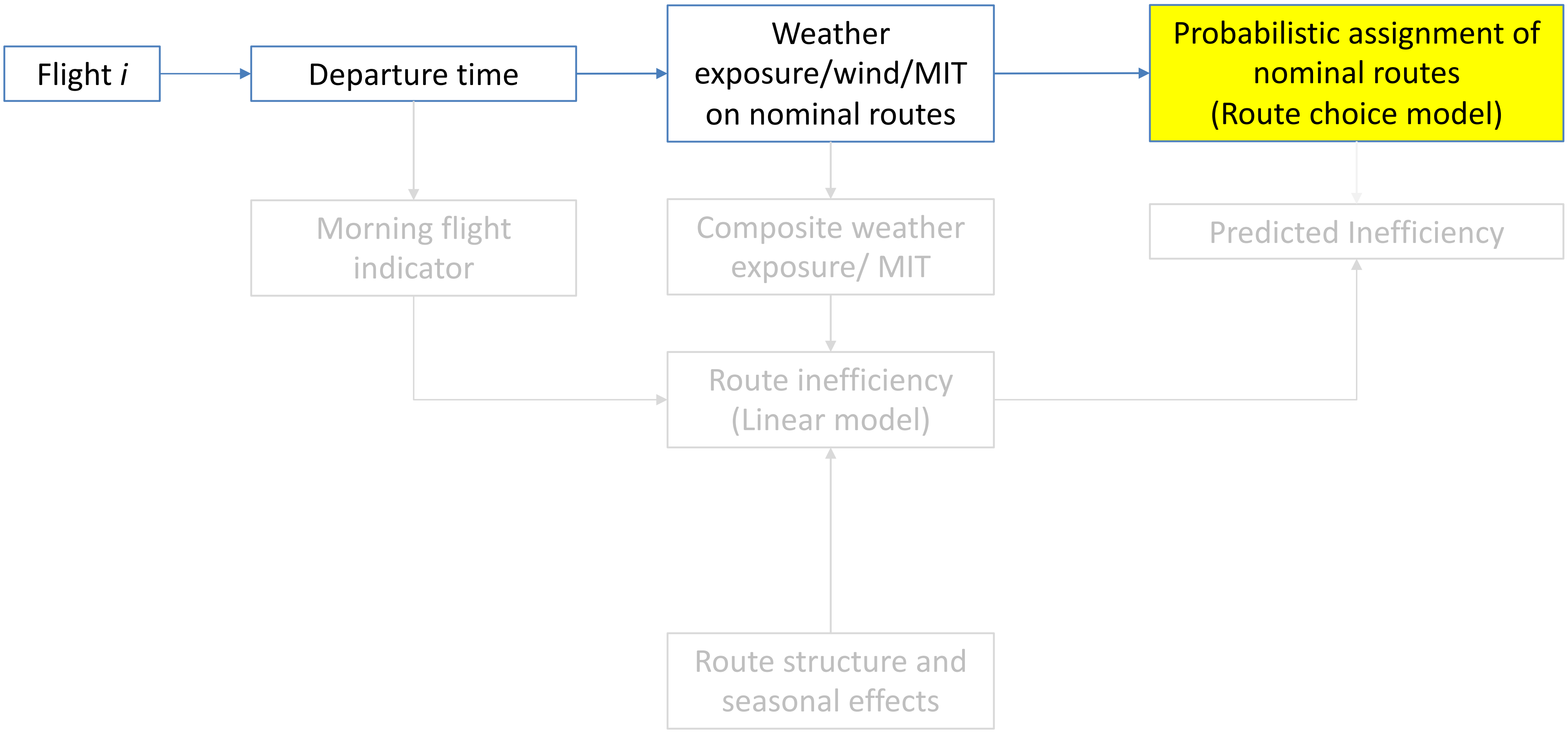
Route structure and seasonal effects



Contribution of Causal Factors

- Rerun models under different counterfactual scenarios
  - Assume no weather
  - Assume no winds
  - Assume no MIT restrictions
  - Assume no weather, winds, or MIT restrictions
- Compare original inefficiency predictions to those based on appropriate counterfactual

Framework


Flight *i*
Departure time
Weather exposure/wind/MIT on nominal routes
Probabilistic assignment of nominal routes (Route choice model)
Morning flight indicator
Composite weather exposure/ MIT
Predicted Inefficiency
Route inefficiency (Linear model)
Route structure and seasonal effects

Route Choice Model

- Model Specification

$$V_0 = ASC_0 + \boldsymbol{\beta_1} \cdot \boldsymbol{TS_0} + \boldsymbol{\beta_2} \cdot \boldsymbol{R} + \boldsymbol{\beta_3} \cdot \boldsymbol{SQ_0} + \boldsymbol{\beta_4} \cdot \boldsymbol{MIT_0} + \boldsymbol{\beta_5} \cdot \boldsymbol{WD_0} + \beta_{7,0} \cdot Season + \beta_{8,0} \cdot MF$$

...

$$V_i = ASC_i + \boldsymbol{\beta_1} \cdot \boldsymbol{TS_i} + \boldsymbol{\beta_2} \cdot \boldsymbol{R_i} + \boldsymbol{\beta_3} \cdot \boldsymbol{SQ_i} + \boldsymbol{\beta_4} \cdot \boldsymbol{MIT_i} + \boldsymbol{\beta_5} \cdot \boldsymbol{WD_i} + \beta_{7,i} \cdot Season + \beta_{8,i} \cdot MF$$

...

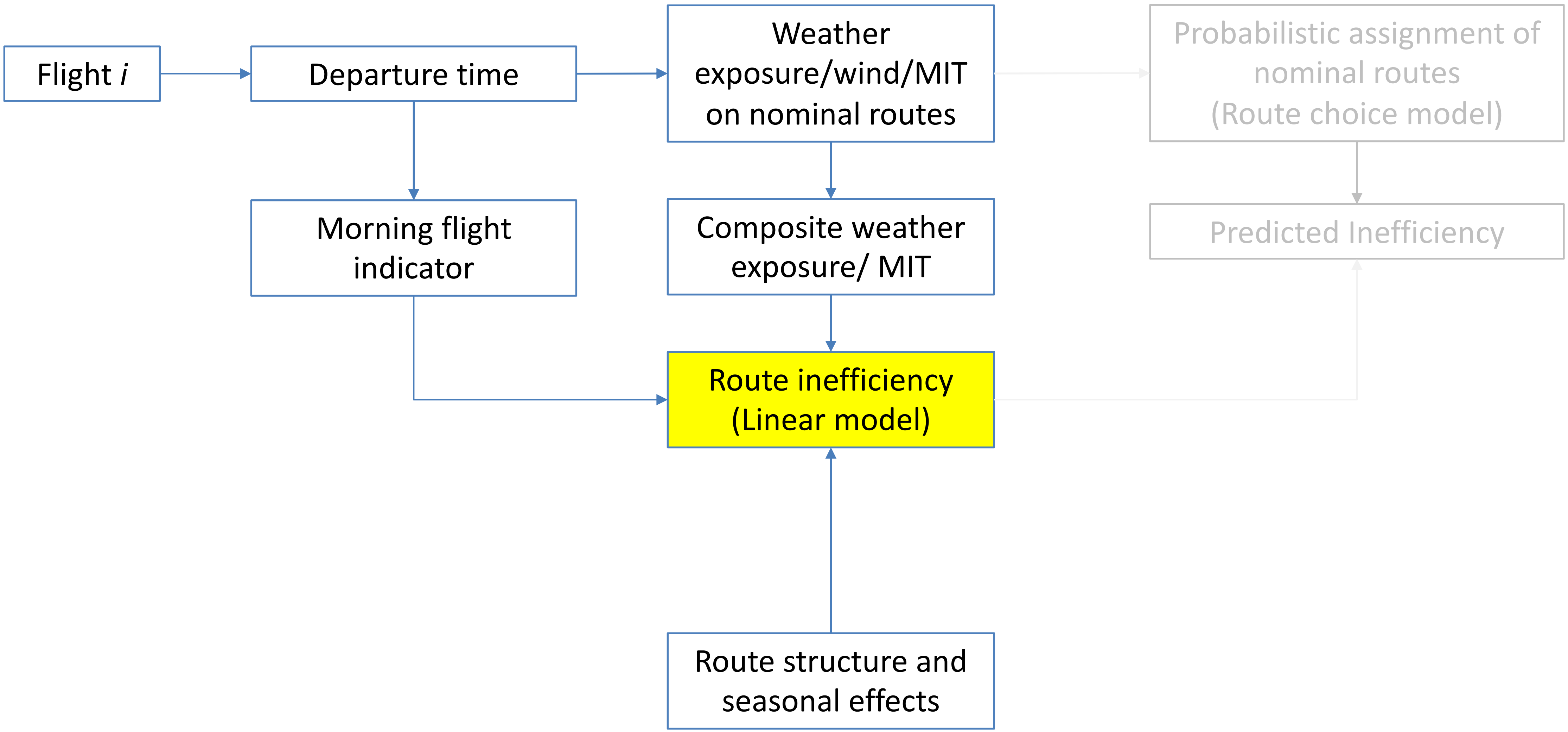
$$V_N = 0 + \boldsymbol{\beta_1} \cdot \boldsymbol{TS_N} + \boldsymbol{\beta_2} \cdot \boldsymbol{R_N} + \boldsymbol{\beta_3} \cdot \boldsymbol{SQ_N} + \boldsymbol{\beta_4} \cdot \boldsymbol{MIT_N} + \boldsymbol{\beta_5} \cdot \boldsymbol{WD_N}$$

- Notation
  - V_i represents the deterministic utility for the i^{th} alternative (nominal route);
  - V_N represents the deterministic utility for ***OUTLIER*** cluster;
  - $Season$: Seasonal fixed effects, 3 dummy variables; Winter: Dec – Feb, Spring: Mar – Apr, Summer: May – Aug, Fall: Sep – Nov;
  - TS: Thunderstorm exposure;
  - R: Rain exposure;
  - SQ: Squall exposure;
  - MIT: MIT stringency;
  - WD: still air distance

Route Choice Model Estimation Results

- Weather variables highly significant for all airport pairs; flights avoid routes with more weather exposure
- Winds highly significant except for JFK<->FLL; flights take routes with more favorable winds
- Miles-in-trail restrictions significant only for JFK<->FLL; flights avoid routes with stronger MIT restrictions

Framework


Flight *i*
Departure time
Weather exposure/wind/MIT on nominal routes
Probabilistic assignment of nominal routes (Route choice model)
Morning flight indicator
Composite weather exposure/ MIT
Predicted Inefficiency
Route inefficiency (Linear model)
Route structure and seasonal effects

Linear Model

Model Specification

$$Log(Inefficiency(\%)) = \beta_0 + \beta_1 \cdot TS + \beta_2 \cdot R + \beta_3 \cdot SQ + \beta_4 \cdot MIT + \beta_5' \cdot X_{NRoute} + \beta_6' \cdot Season + \beta_7 \cdot MF$$

- TS:Thunderstorm exposure
- R: Rain exposure
- SQ: Squall exposure
- MIT: MIT stringency

Averaged across all nominal routes with weights based on cluster percentages

- X_{NRoute}: Fixed effects of nominal routes for different airport pairs
- $Season$: Seasonal fixed effects, winter as baseline; Winter: Dec – Feb, Spring: Mar – May, Summer: Jun – Aug, Fall: Sep – Nov
- MF: Binary variable. 1 if local departure hour is before 12 pm.

Estimation Results

- R squares range from 0.38 to 0.71
- Thunderstorm exposure is highly significant for all six pairs
- Inefficiency increases from 0 to the maximal thunderstorm exposure in the data are 1.4-2.2%
- Rain, squall and MIT are significant only for certain airport pairs
- The estimates of cluster membership are highly significant, indicating that the between cluster effect is important

Contribution of Factors

- We would like to understand how much those factors contribute to flight en route inefficiency.
- Both models are used to calculate the contributions

$$\%\Delta = \frac{\sum_{C_L}\left[\overline{\mathbb{E}(Inefficiency|F, C_L)P(C_L|F)} - \overline{\mathbb{E}(Inefficiency|F=\mathbf{0}, C_L)P(C_L|F=\mathbf{0})}\right]}{\sum_{C_L}\overline{\mathbb{E}(Inefficiency|F=\mathbf{0}, C_L)P(C_L|F=\mathbf{0})}}$$

  - $\mathbb{E}(\cdot)$: Predicted inefficiency based on route inefficiency model for a given flight
  - C_L: Cluster *L*
  - F: Vector of factor variable (i.e., $[Wx, WD, MIT]$)
  - $P(C_L|F)$: Probability that a given flight is assigned to route cluster *L* given F

Contribution of Factors

City Pair	Convective Weather	Miles-In-Trail	Wind	Combined Effect
IAH → BOS	4.10%	1.72%	4.91%	10.67%
BOS → IAH	10.15%	-	2.60%	11.71%
FLL → JFK	4.34%	12.57%	-	16.91%
JFK → FLL	6.54%	6.68%	-	13.35%
JFK → LAX	11.37%	0.83%	11.72%	25.11%
LAX → JFK	9.51%	4.67%	3.94%	18.87%

- Convection accounts for 4% - 11% of en route inefficiency for the 6 pairs.
- Wind accounts for up to 12% of en route inefficiency.
- MIT accounts for up to 13% of en route inefficiency.
- Combined effect of wind, convective weather and MIT range from 11% to 25%.

Outline

- Introduction
- Data Sources and Preliminary Statistical Analysis
- Identifying Nominal Trajectories
- Mapping Causal Factors to Trajectories
- Causal Analysis of Flight Inefficiency
- Conclusions

Conclusions

- Contributions of weather, wind, and MITs to en route inefficiency have been estimated for US domestic flights
- Our approach considers effects related to assignment of flights to routes and variation of inefficiency for given routes
- Model estimates have expected signs and, in general, are statistically significant
- Estimated contributions of above factors to total inefficiency:
  - From convective weather: 5 – 15%
  - From wind: 0 – 12%
  - From MIT: 0 – 5%
  - All factors together: 12% – 25%

Ongoing Research

- Additional causal factors
  - Monitor alerts (indicate sectors predicted to have excess demand)
  - Airspace flow programs (ground hold flights to meter demand through airspace with reduced capacity)
  - NAS route structure
- Model refinements
  - Convective weather metrics based on better data
  - Improved exposure metrics
  - Improved statistical modeling
- Generative flight route model
  - Use deep learning methods to predict routes of specific flights based on causal factors information
  - Use counterfactuals to find impacts of particular factors

Thanks!

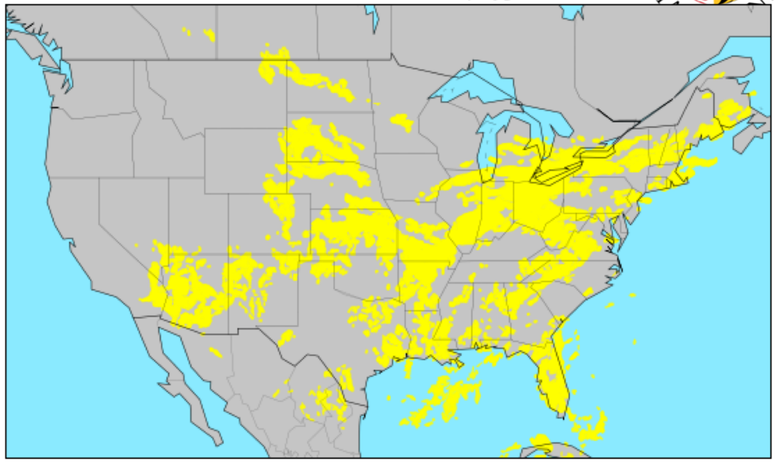
Q&A

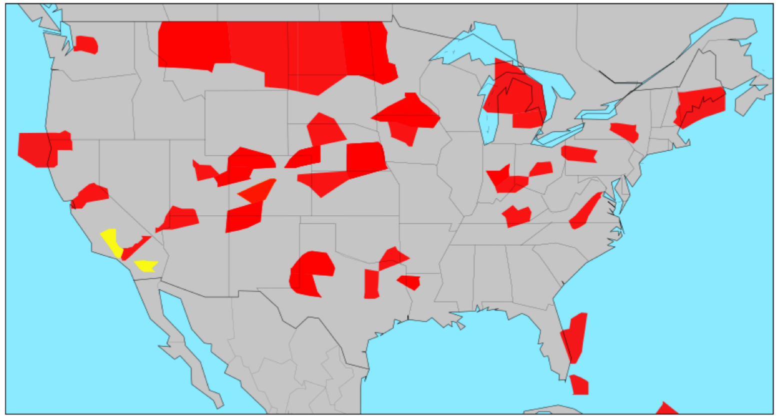
mhansen@ce.berkeley.edu

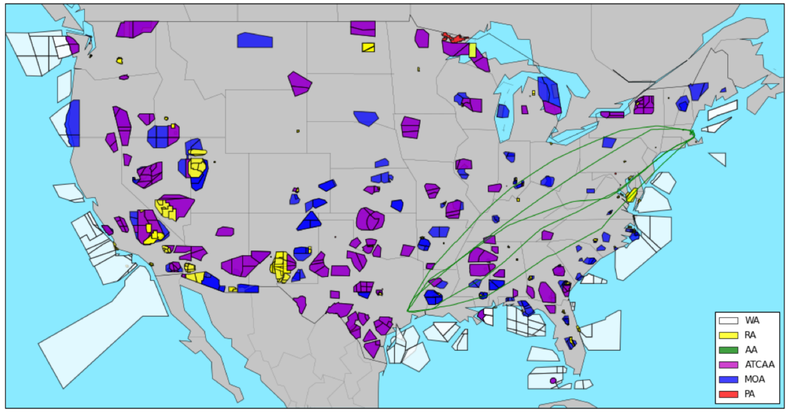
Ongoing Research

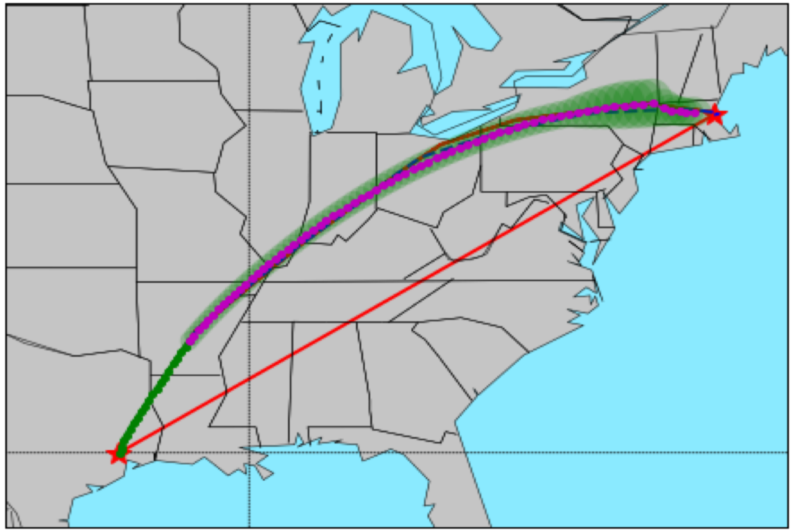
- Improving prediction accuracy
  - High-fidelity convective weather data from NCWF
  - More features to add
    - Special Use Airspace (SUA)
    - Monitor Alert (MA)
    - Airspace Flow Program (AFP)
- Individual flight trajectory prediction tool
  - Deep generative neural networks

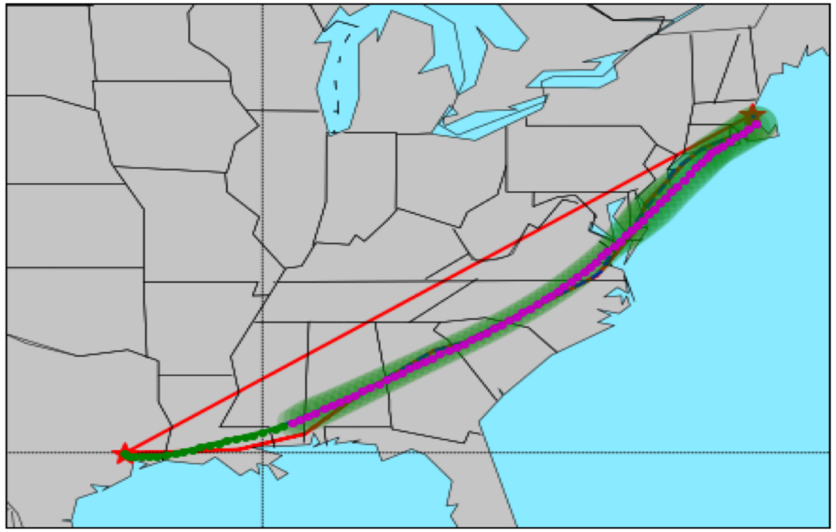
NCWF convective weather polygon



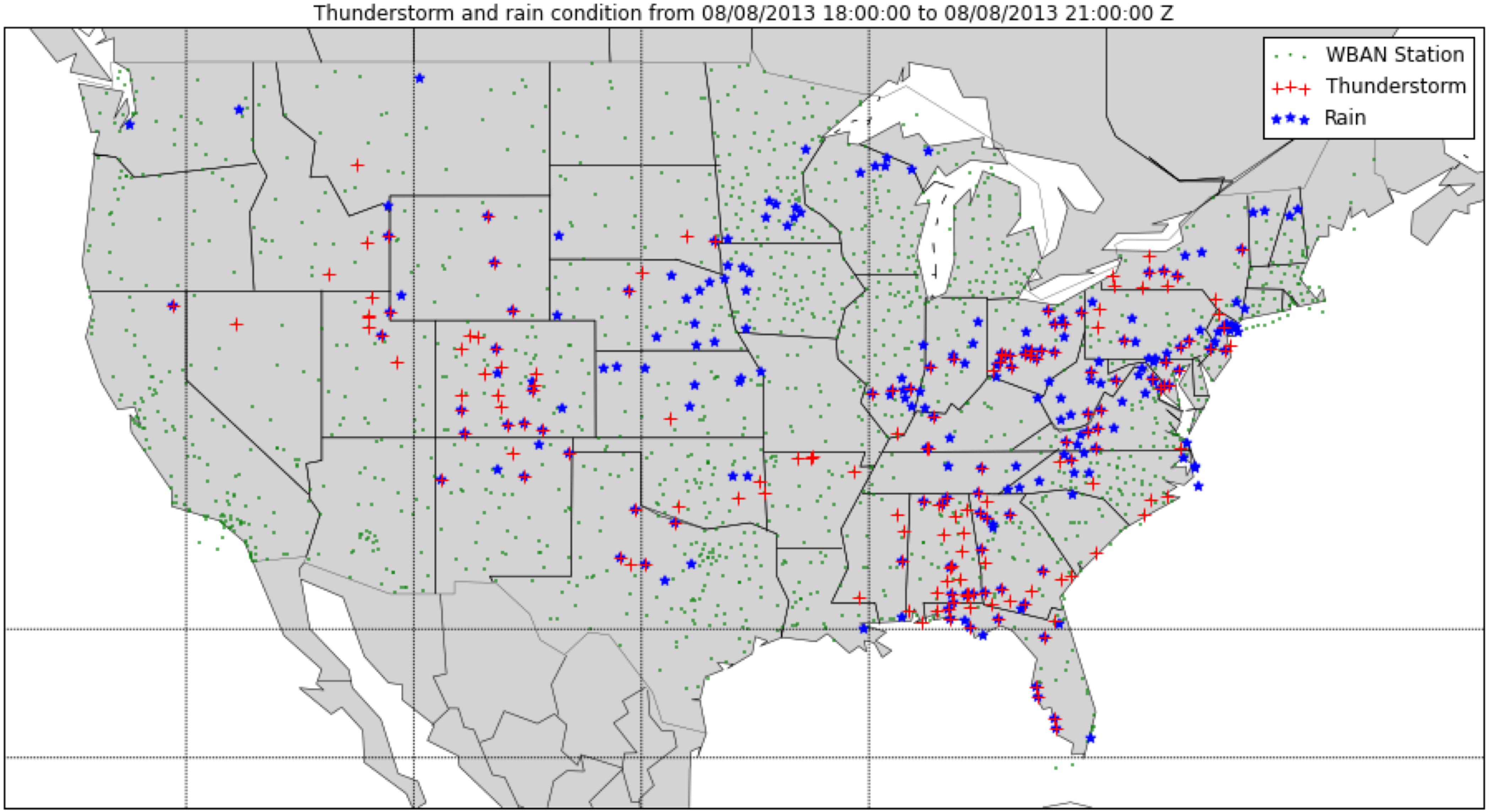




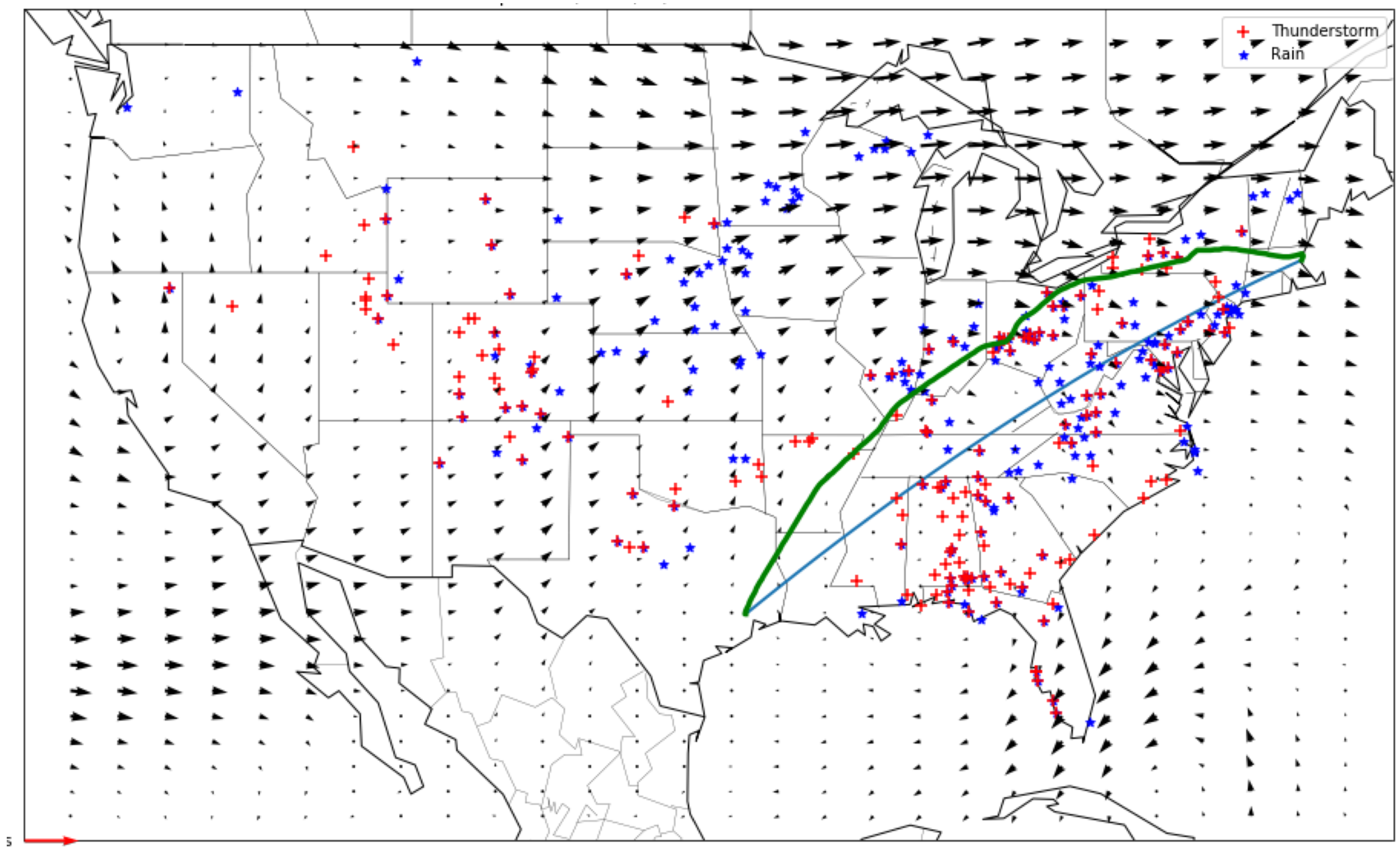




Convective Weather Data



Introduction (III)



Why green?

Estimation Results

- Probability for flight i to choose nominal route k, given causal factor variables F_x and other control variables X: $P(Y_i = k|F_x, X) = \frac{e^{V_k}}{\sum_{j=1}^{K} e^{V_j}}$
- The estimation results indicate that thunderstorm, rain, wind distance and MIT appear to be the most significant factors that influence the strategic routing

Var.	Airport Pair (Est./ Std.)					
	IAH BOS	BOS IAH	FLL JFK	JFK FLL	JFK LAX	LAX JFK
TS	-31.52*** (4.64)	-43.10*** (5.00)	-17.92*** (3.12)	-9.92*** (3.078)	-22.72*** (2.38)	-39.64*** (2.89)
R	-9.25*** (1.36)	-11.04*** (1.41)	-3.92*** (1.49)	-6.67*** (1.15)	-5.86*** (0.71)	-5.62*** (0.68)
SQ					-211.03** (81.95)	-146.69* (79.04)
WD	-16.43*** (1.096)	-11.76*** (1.38)			-13.64*** (0.52)	-12.78*** (0.64)
MIT			-1.38*** (0.47)	-1.39*** (0.35)		

*Standard errors in parentheses; * p<.1, ** p<.05, ***p<.01*

- Estimates indicate that thunderstorm, rain, squall, headwind and MIT will degrade the probability of a flight to choose an alternative

Estimation Results

Var.	Airport Pair (Est./ Std.)					
	IAH BOS	BOS IAH	FLL JFK	JFK FLL	JFK LAX	LAX JFK
TS	25.76*** (2.87)	11.61*** (2.68)	5.49** (2.30)	7.27*** (1.27)	30.72*** (1.66)	41.35*** (2.24)
R	1.35* (0.81)				1.55*** (0.45)	3.70*** (0.52)
SQ					226.72*** (65.32)	
MIT		1.06*** (0.32)	0.86** (0.36)	1.29*** (0.25)		4.10*** (0.30)
BH		0.15** (0.06)	1.05*** (0.10)	0.13** (0.05)	0.04* (0.02)	0.54*** (0.03)

Var.	Airport Pair (Est./ Std.)					
	IAH BOS	BOS IAH	FLL JFK	JFK FLL	JFK LAX	LAX JFK
CL_r	-4.56*** (0.15)	-4.14*** (0.10)	-18.04*** (0.35)	-8.03*** (0.17)	-2.69*** (0.03)	-1.85*** (0.06)
CL_g	-2.22*** (0.15)	0.58*** (0.16)	-12.14*** (0.37)	1.42*** (0.30)	1.13*** (0.07)	-2.31*** (0.05)
CL_n	-5.42*** (0.16)	-3.49*** (0.11)	-12.82*** (0.59)	-1.92*** (0.19)	-1.15*** (0.04)	-1.64*** (0.09)
CL_c	-7.03*** (0.36)	-4.18*** (0.16)		2.21*** (0.25)	-1.08*** (0.05)	-2.31*** (0.04)
CL_b	-5.14*** (0.35)	-0.4 (0.25)		5.11*** (0.36)	-2.23*** (0.04)	
Spring	-0.04 (0.10)	0.09 (0.09)	0.26** (0.13)	0.14** (0.07)	-0.03 (0.03)	0.24*** (0.05)
Summer	0.19* (0.10)	-0.04 (0.09)	0.21 (0.15)	-0.08 (0.09)	-0.21*** (0.04)	-0.20*** (0.05)
Fall	0.05 (0.10)	0.08 (0.09)	-0.04 (0.14)	-0.03 (0.08)	0.01 (0.03)	0.01 (0.05)
Const.	8.00*** (0.17)	5.56*** (0.12)	19.74*** (0.37)	9.65*** (0.18)	3.48*** (0.04)	3.39*** (0.05)
R^2	0.63	0.65	0.54	0.76	0.53	0.33
Obs.	1664	1732	3988	4021	10637	10367

*Standard errors in parentheses; * p<.1, ** p<.05, ***p<.01*