Big Aviation Data Mining for Robust, Ultra-Efficient Air Transportation

Technical Monitor:
Sarah D'Souza, Systems Analysis Office, NASA Ames Research Center

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MIT ICAT

MIT International Center for Air Transportation

NASA LEARN
Phase 1 Outbrief
16 February 2016

NARI
NASA AERONAUTICS RESEARCH INSTITUTE

Team Members

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Kajal Claypool
Data Architectures

Emily Clemons
Analytics

Rich DeLaura
Co-PI

Yan Glina
Analytics

Rich Jordan
Analytics

Alex Proschitsky
Data Architectures

Tom Reynolds
Co-PI

Ngaire Underhill
Analytics

MIT ICAT

Hamsa Balakrishnan
Analytics, Grad student advisor

John Hansman
Analytics, Grad student advisor

Jacob Avery
Analytics, Grad student

Cal Brooks
Analytics, Grad student

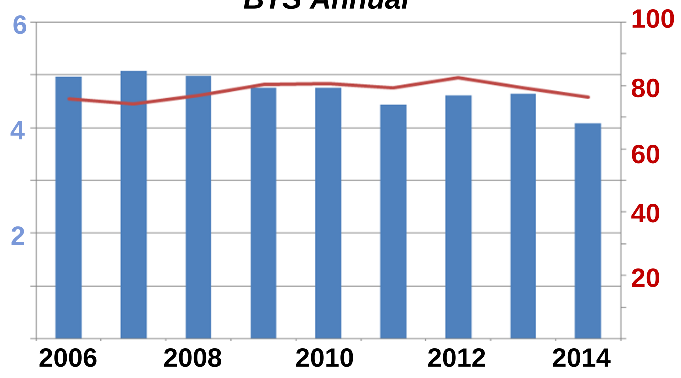
Mayara Conde Rocha Murca
Analytics, Grad student

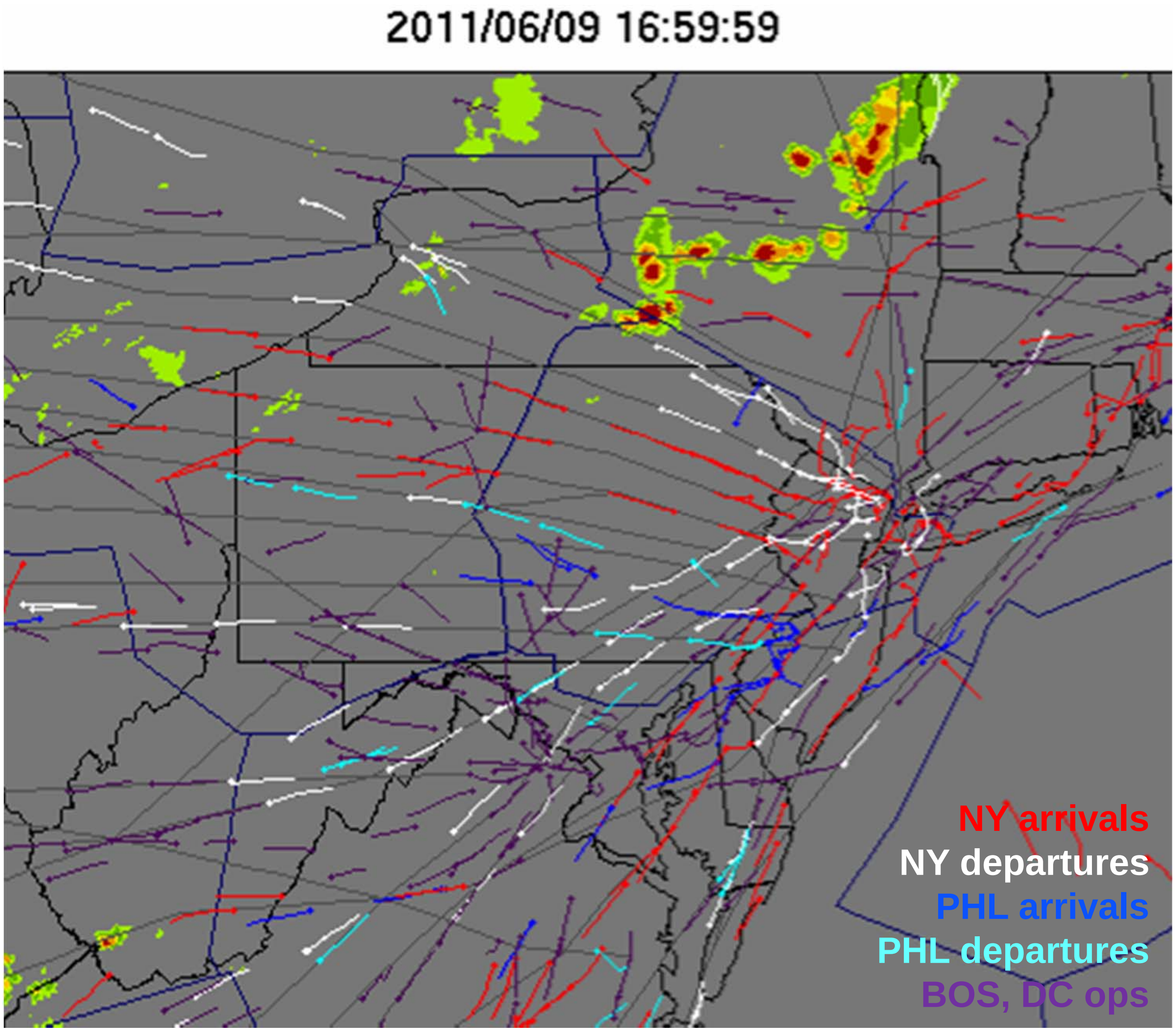
Karthik Gopalakrishnan
Analytics, Grad student

Air Transportation System Challenges

- **Air transportation system is very safe, but efficiency & robustness challenges remain**
- **Most inefficiencies caused by capacity & demand imbalances at range of spatial & temporal scales**

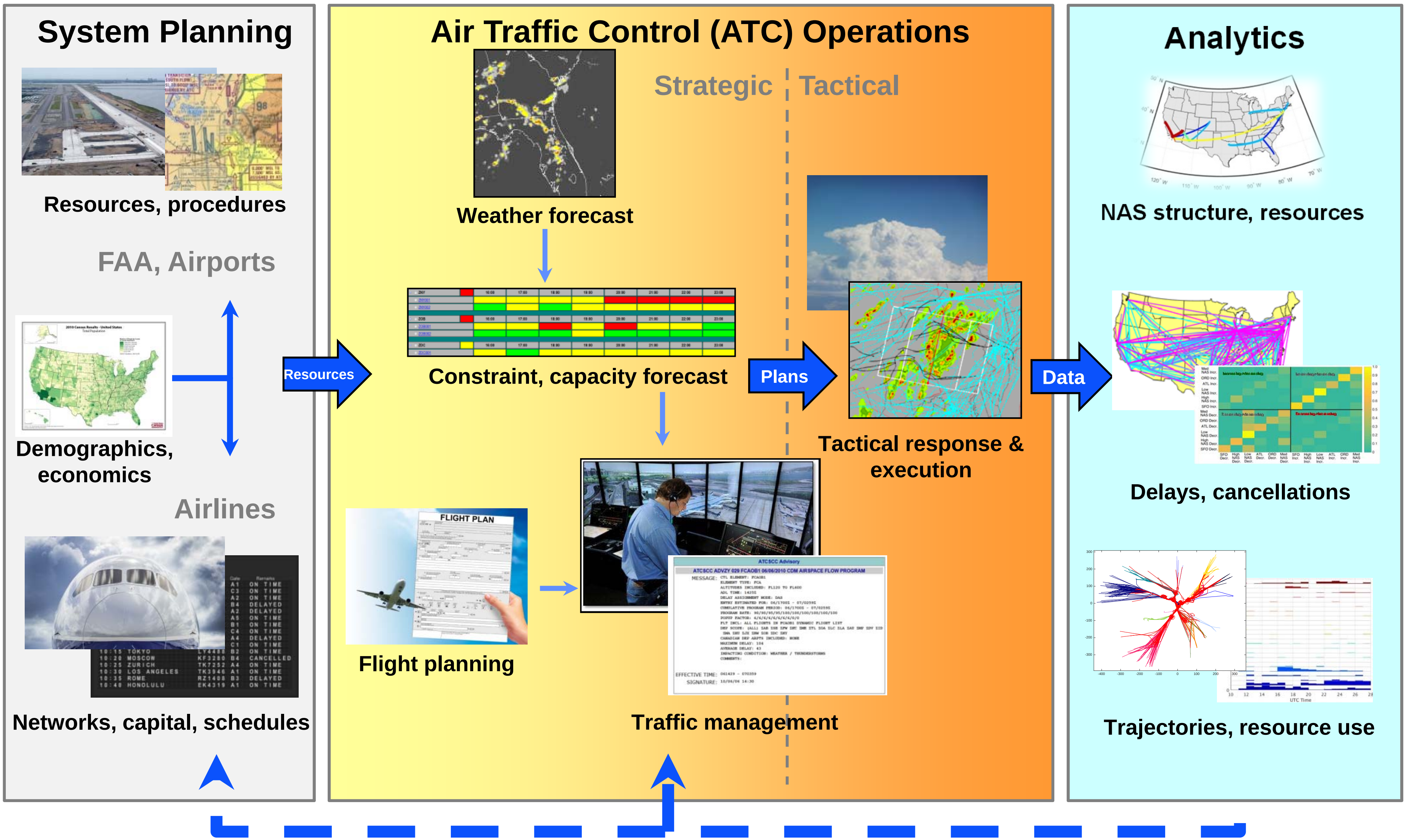
Millions of departures / % on time
BTS Annual





National Airspace System (NAS)

…in a single slide



LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Space, Time, Data, and Impacts

Scale / scope of impact

NAS+

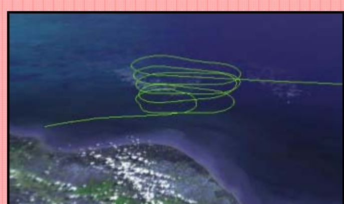
NAS

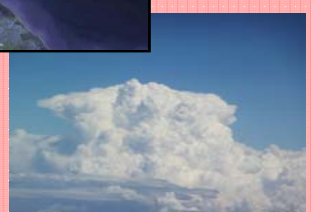
Regional

Local

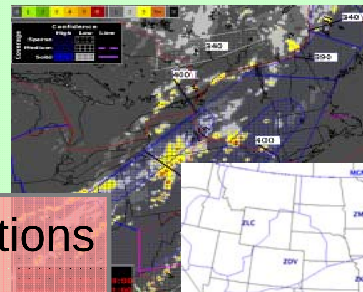
Flight trajectories
Local / regional high resolution weather
Tactical decisions
Etc.

Tactical ATC Operations





Strategic ATC Operations



Delay
Cancellation
Traffic management planning
National / regional weather
Route planning, Etc.

Airline Network Planning
FAA System Planning

Schedule
Economic
Demographic
Climatology, Etc.

Sample data types
Explored
Example future

Minutes Hours Days Years

Planning / operational horizon

Goal: Demonstrate Big Data analytic framework for aviation across spatial/temporal scales

Data Descriptions

Data Description	Spatial Extent	Spatial Resolution	Temporal Extent	Temporal Resolution
Planning				
Flight operations	NAS-wide	Airport pair (>300 BTS airports)	2000 - 2014	Annual
Strategic ATC Operations				
Flight delays, cancellations	NAS-wide	Airport pair (>300 BTS airports)	2008 - 2014	Annual, Seasonal, Daily, Hourly
Traffic Management Initiatives	NAS-wide	N/A	2008 - 2014	Daily
Tactical ATC Operations				
Flight trajectories	Regional (NY, DFW, SFO metro)	~5 miles	2013 - 2015	1 minute
Weather radar mosaics	Regional (NY, DFW, SFO metro)	1 km	2013 - 2015	2.5 minute
Convective weather impacts	NY metro	Individual route	2013 - 2015	5 minute
Terminal wind impacts	NY metro	Individual terminal	2013 - 2015	hourly

BTS = Bureau of Transportation Statistics

Anatomy of the Big Data Analysis Framework

'Raw'
system data

Size of data…

Aggregate into compact mathematical representation

Derive descriptive metrics

Identify patterns of system behavior

Size of insight…

Enabled insights, applications, solutions

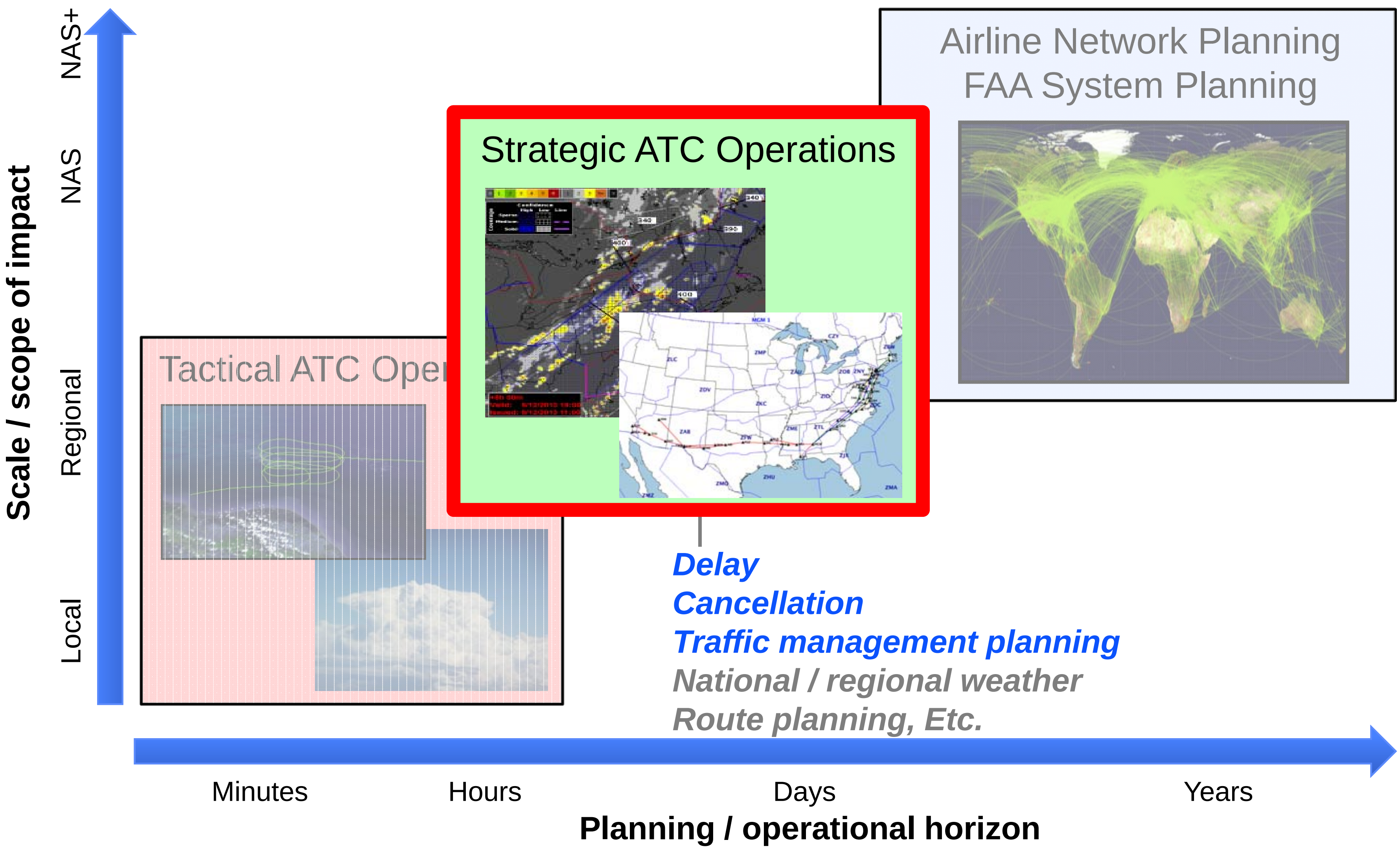
- Identify anomalies, interesting behaviors
- Develop descriptive, predictive models
- Evaluate performance, identify best practices

Analytics must be scalable, generalizable, and interpretable

Outline

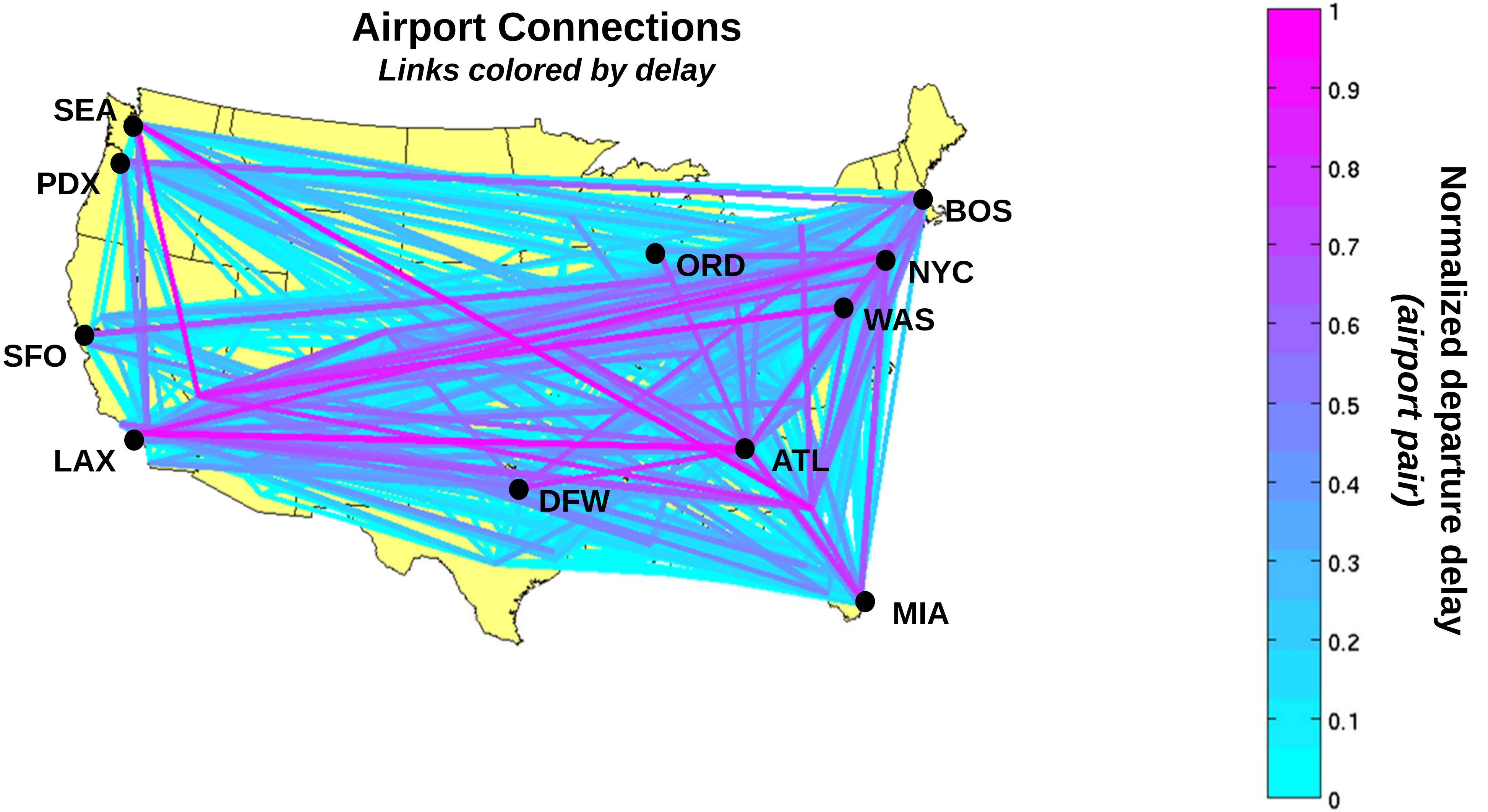
- **Motivation: Air transportation system challenges and Big Data opportunities**
- **Technical approach & Selected results:**
  - **Strategic ATC Operations**
  - **Tactical ATC Operations**
  - **Airline Network Planning**
- **Summary of innovations, Potential impacts and Next step recommendations**
- **Distribution / Dissemination & Acknowledgements**

Space, Time, Data, and Impacts



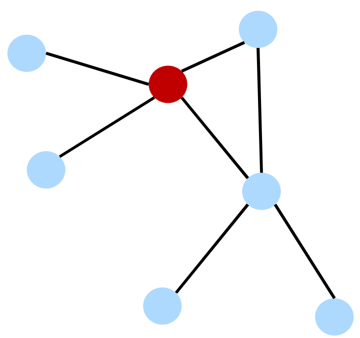
NAS-Wide Operational Network

At a glance…

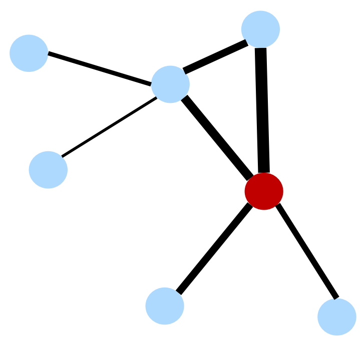


Strategic ATC Operations:
Analyzing the NAS-Wide Network

Adjacency matrix



Demand-weighted adjacency matrix



KEY: Airport — Flight connection

Eigencentrality:

Airport connectivity

Application:

Network structure

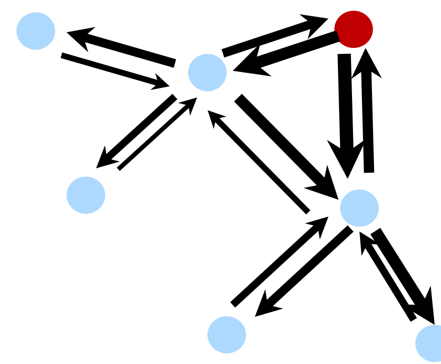
Eigencentrality:

Airport throughput

Application:

Network capacity

Delay, cancellation weighted adjacency matrix



HUB: Sends delay	AUT: Receives delay	DYNAMIC
High (Low)	High (Low)	Inbound, outbound delay balanced
High	Low	Delay propagator
Low	High	Delay reducer

Hub, authority metrics:

Asymmetrical propagation of delay, cancellation

Application:

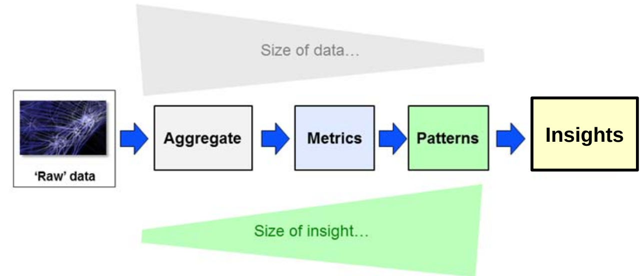
Propagation of weighting metric (delay, cancellation, etc.)

Goal: Characterize and model NAS-wide network dynamics and performance

Approach: Apply novel adjacency matrix weightings and metrics to define NAS-*wide states* that characterize propagation of disruptions

Delay State Identification: Methodology

Framework key:



United States Department of Transportation
Office of the Assistant Secretary for Research and Technology
Bureau of Transportation Statistics

Flight delays, cancellations (2008-2014)

Aggregate (daily, hourly) weighted connectivity matrices (delay, cancellation) → **Calculate Hub, Authority scores for major airports** → **Cluster into propagation patterns** →

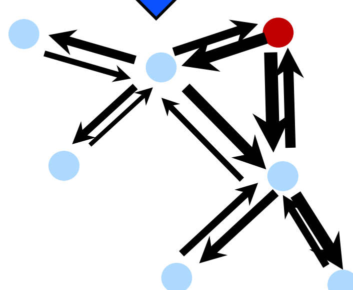
Daily Delay / Cancellation States

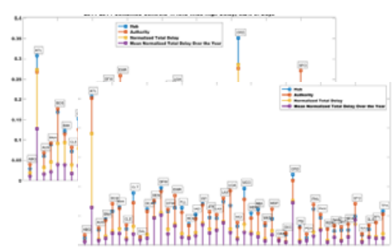
Post-event performance evaluation

Hourly Delay / Cancellation States

Dynamic delay propagation for predictive modeling

…

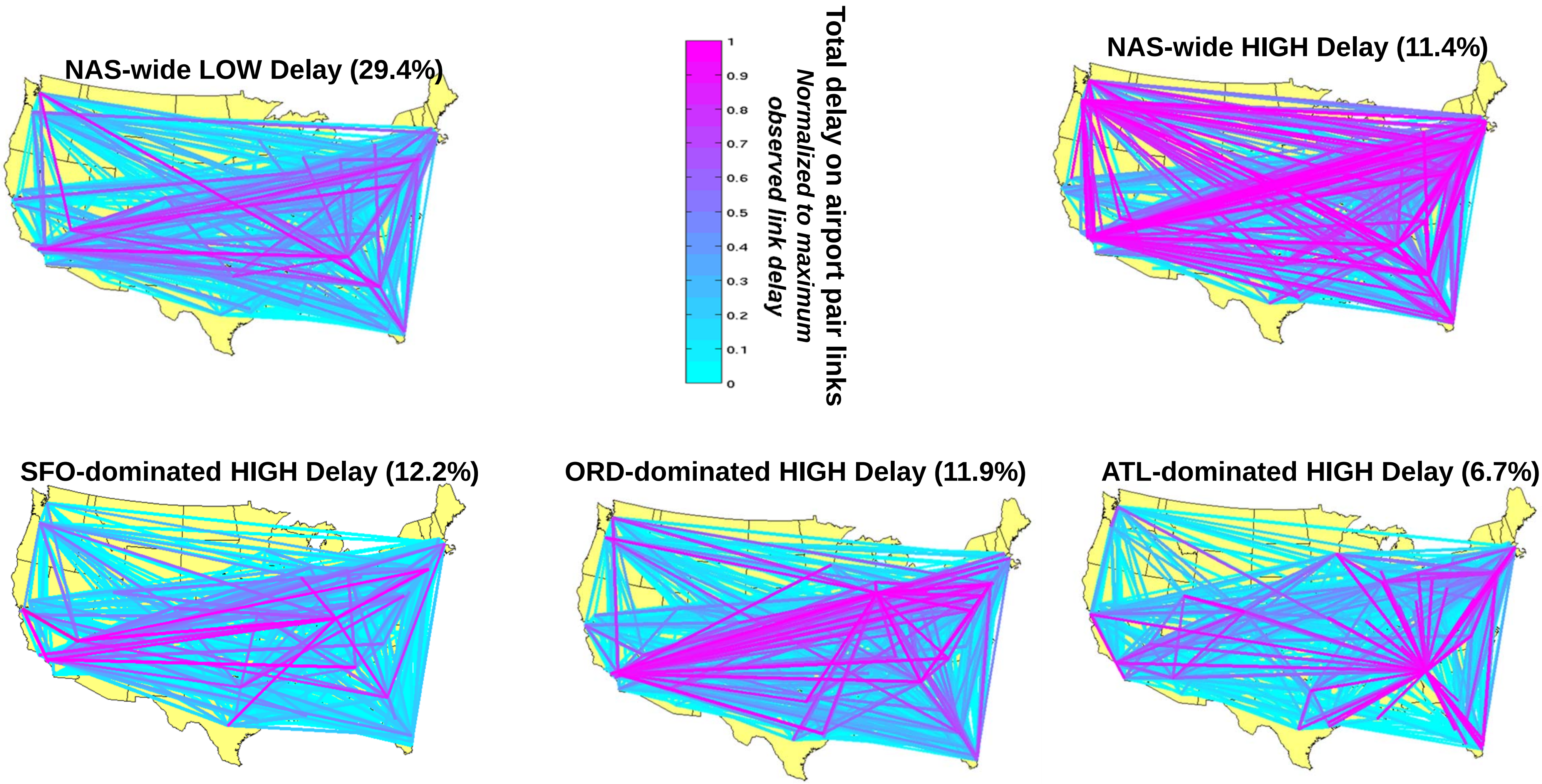




…

Delay Distribution by Daily Delay State

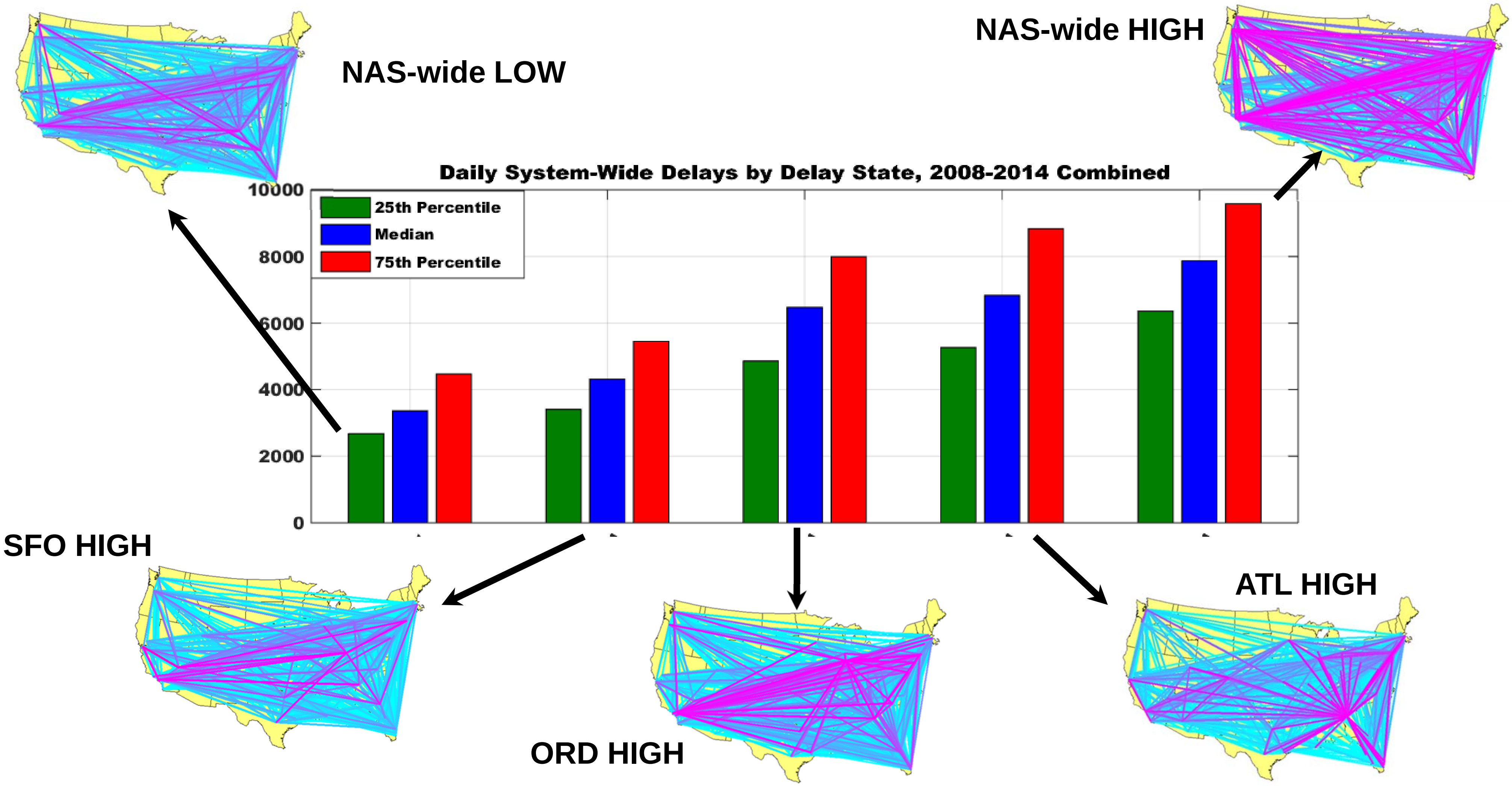
Selected (5 of 12) *Persistent* Delay States (2008-2014)



Daily Delay States provide insights into the scale and propagation of delay

NAS-Wide Delays by Daily Delay State

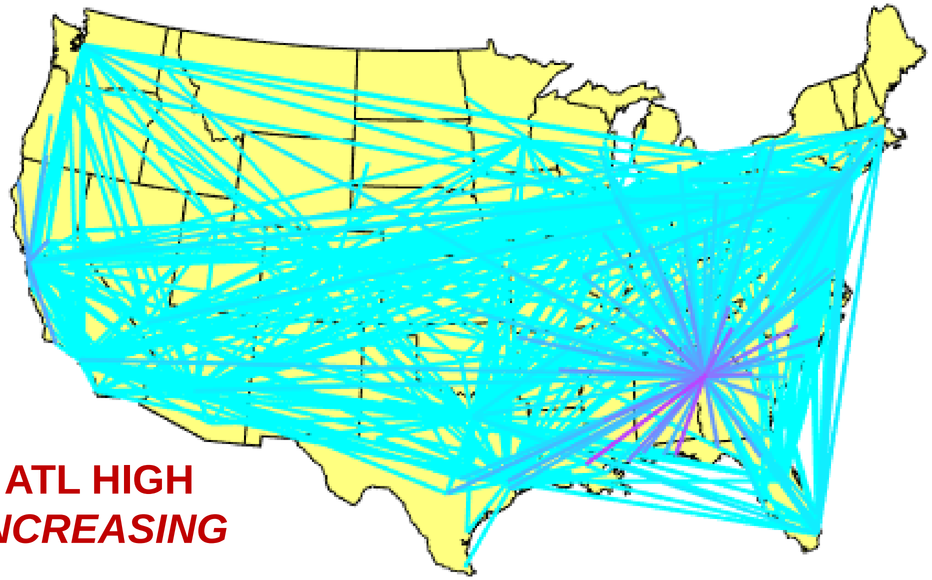
2008 - 2014



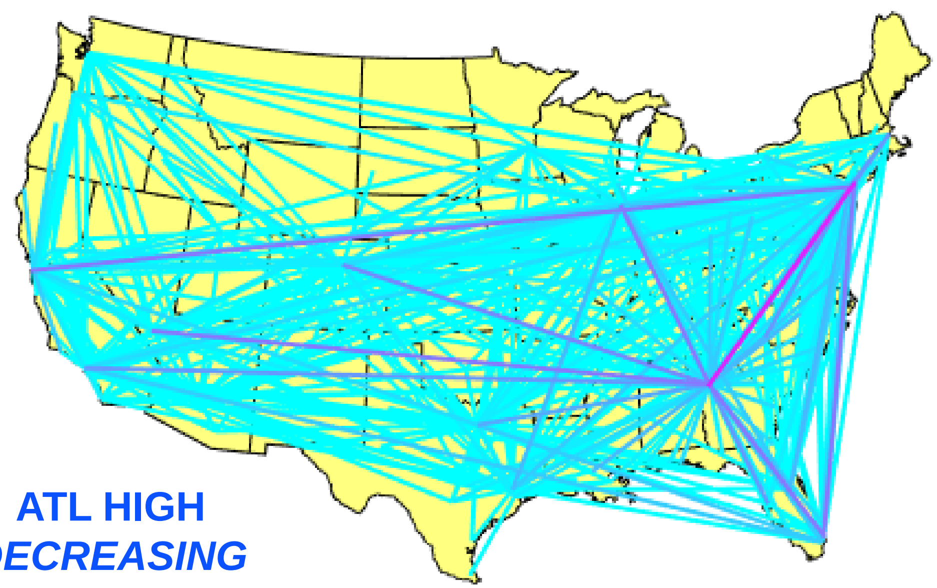
Total delay is similar (but propagation is not) in single-airport dominated states
Total delay in NAS-wide states tends to the extremes

Hourly Delay States

Capturing Dynamics of Delay Propagation

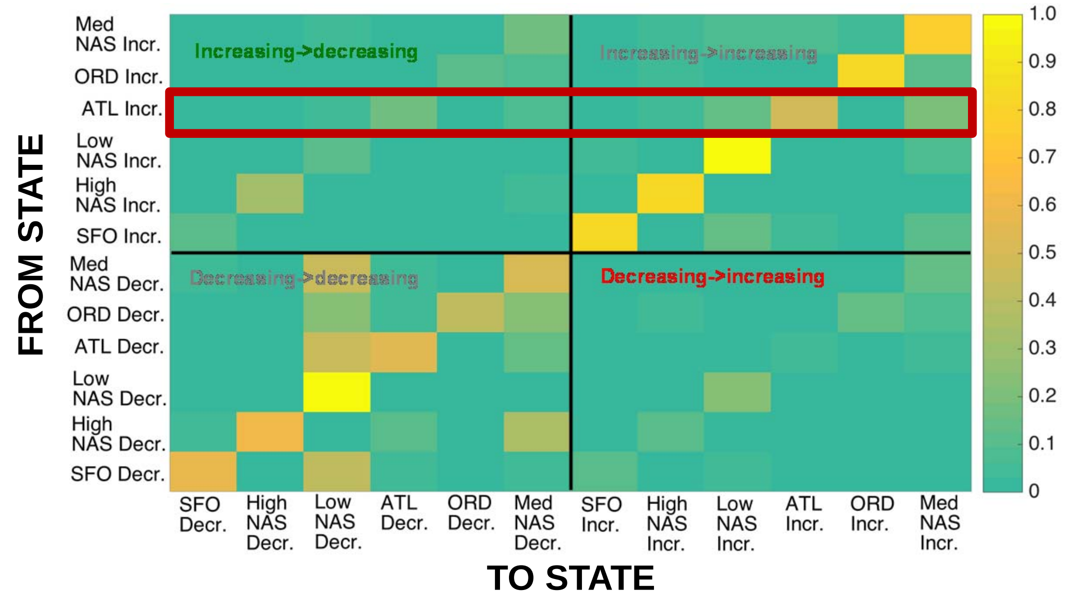


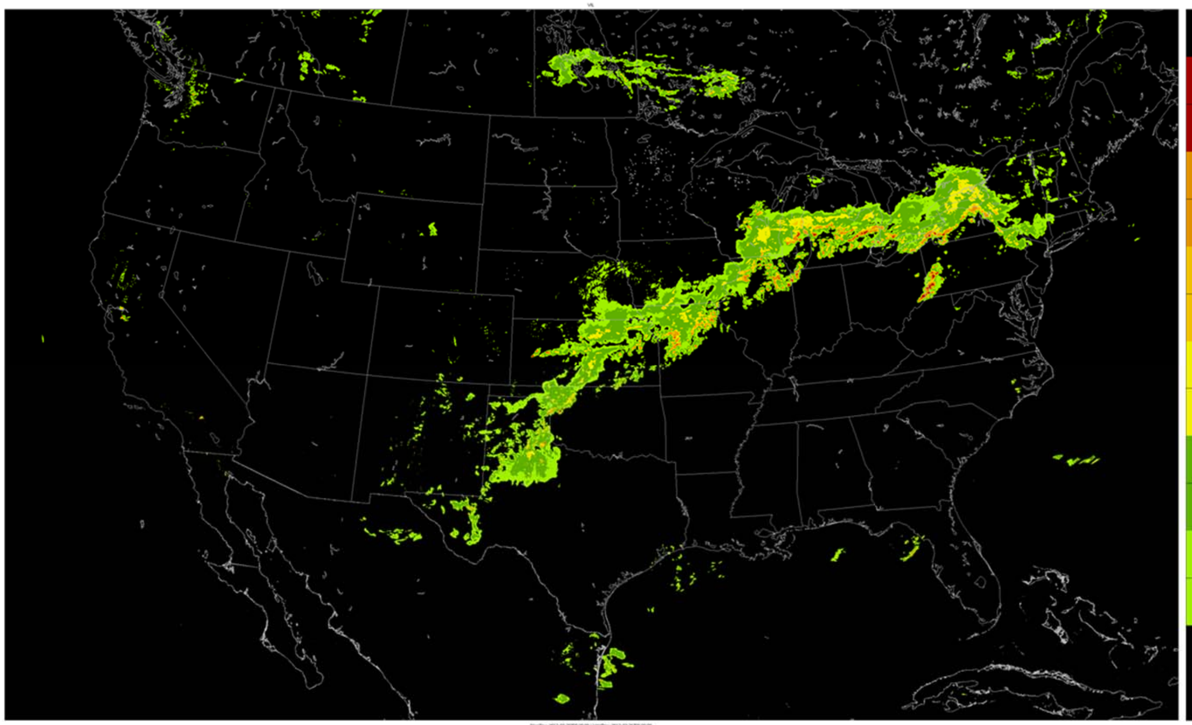
ATL HIGH
INCREASING



ATL HIGH
DECREASING

- **Hourly Delay States capture delay *propagation structure, magnitude,* and *trends***
  - **Local delays build and spread**
  - **Propagation is widest as delays peak and begin decrease**
- **Observed Hourly Delay State *transition probabilities,* and *dwell times* can be calculated**





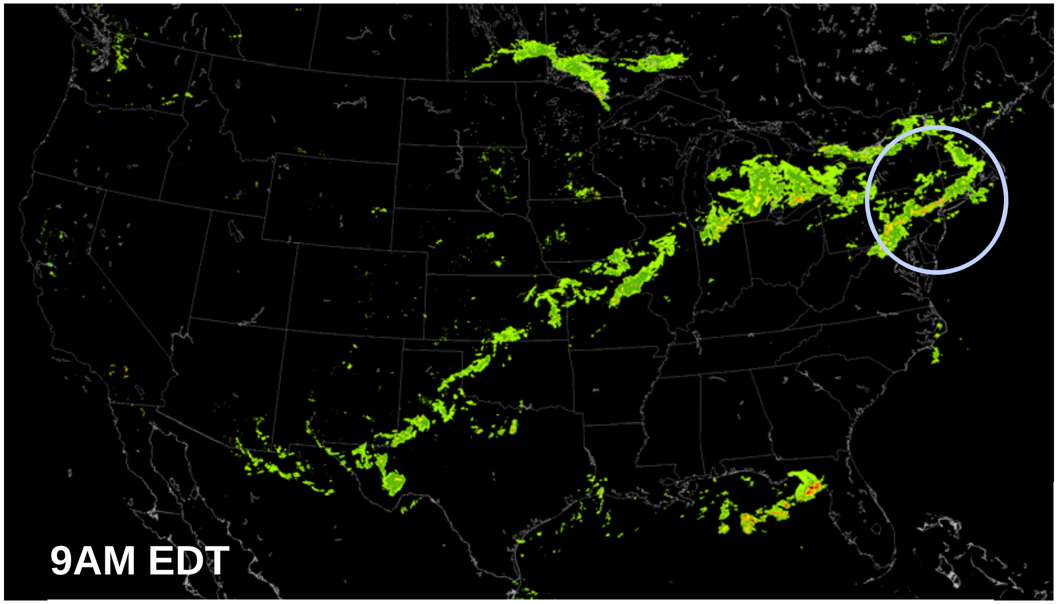
04:00 EDT

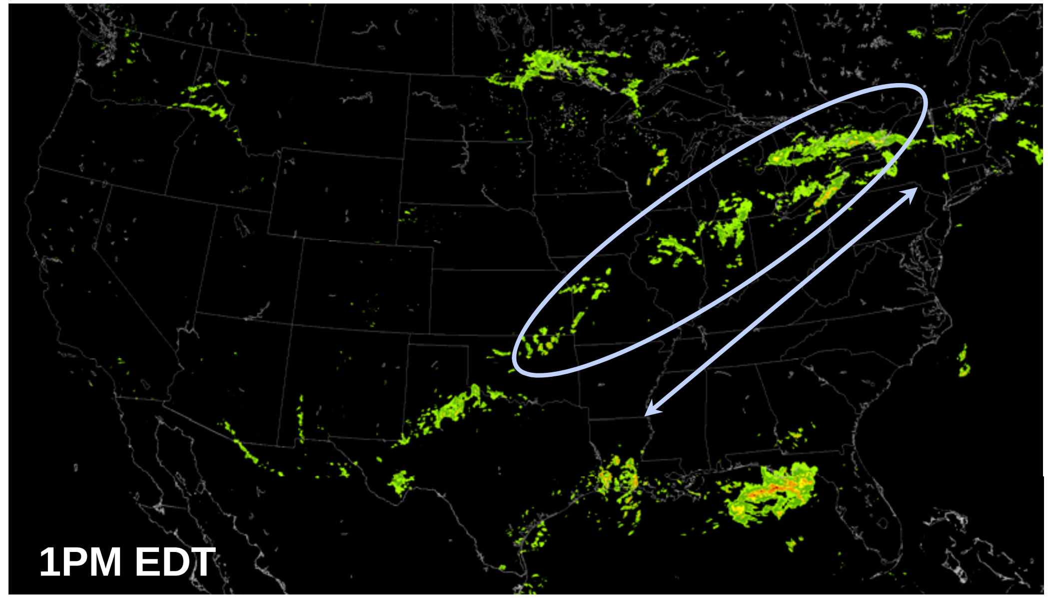
July 26, 2012

Day	Delay	Cancelled
July 26, 2012	26808 hours	554
Avg: 2008-2014	13054 hours	295

Network Dynamics Case Study

26 July, 2012



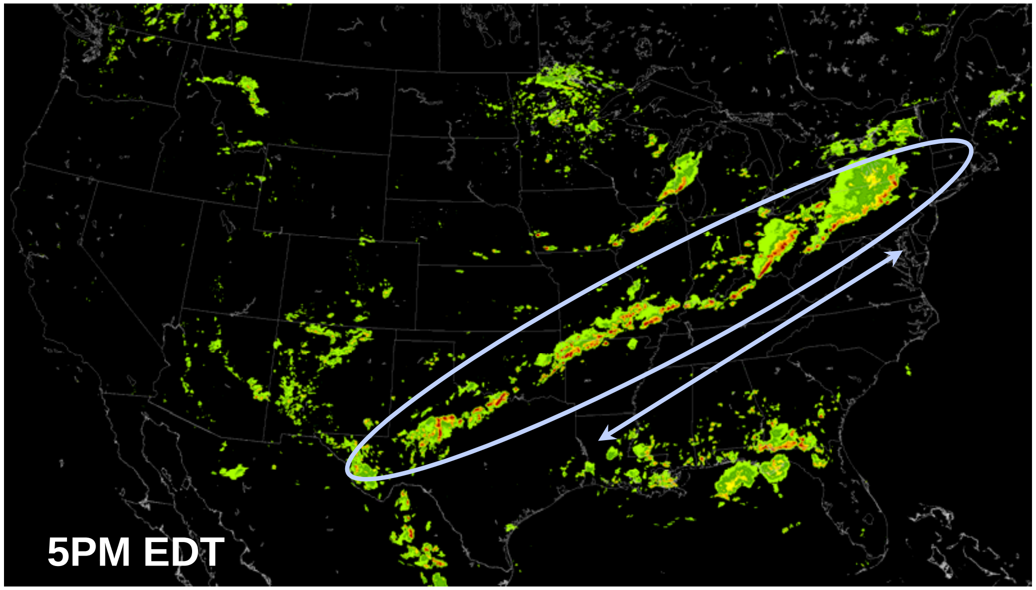


NY Ground Delay Program (GDP) to reduce demand as thunderstorms impact local operations

NY GDP continues & delays persist and propagate as weather dissipates and major traffic corridors clear

Network Dynamics Case Study

26 July, 2012



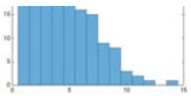
Delays rapidly increase storms bisect the NAS (but coastal corridor remains clear)

Delay growth and propagation appear to be driven by weather-related airspace constraints and control decisions with long time constants

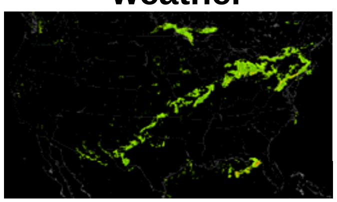
Delay State dwell times, transition probabilities provide insight into NAS system response times

Strategic ATC Operations: Next Steps

Delay states
Dwell times
Observed transition probabilities



Forecast, observed weather



Traffic management decisions

ATCSCC Advisory

ATCSCC ADVZY 029 FCAOB1 06/06/2010 CDM AIRSPACE FLOW PROGRAM

MESSAGE: CTL ELEMENT: FCAOB1
ELEMENT TYPE: FCA
ALTITUDES INCLUDED: FL120 TO FL600
ADL TIME: 1425Z
DELAY ASSIGNMENT MODE: DAS
ENTRY ESTIMATED FOR: 06/1700Z - 07/0259Z
CUMULATIVE PROGRAM PERIOD: 06/1700Z - 07/0259Z
PROGRAM RATE: 90/90/95/95/100/100/100/100/100/100
POPUP FACTOR: 6/6/6/6/6/6/6/6/0/0
FLT INCL: ALL FLIGHTS IN FCAOB1 DYNAMIC FLIGHT LIST
CANADIAN DEP ARPTS INCLUDED: NONE
MAXIMUM DELAY: 104
AVERAGE DELAY: 43
IMPACTING CONDITION: WEATHER / THUNDERSTORMS
COMMENTS:

EFFECTIVE TIME: 061429 - 070359
SIGNATURE: 10/06/06 14:30

Delay Propagation Modeling

Markov Jump Linear System

$$\vec{x}(t+1) = \Gamma_{m(t)}\vec{x}(t),$$

$$\pi_{ij}(t) = \mathrm{pr}[m(t+1) = j | m(t) = i]$$

$\vec{x}(t)$ Vector of airport delays at time *t*

$m(t)$ Delay state at time *t*

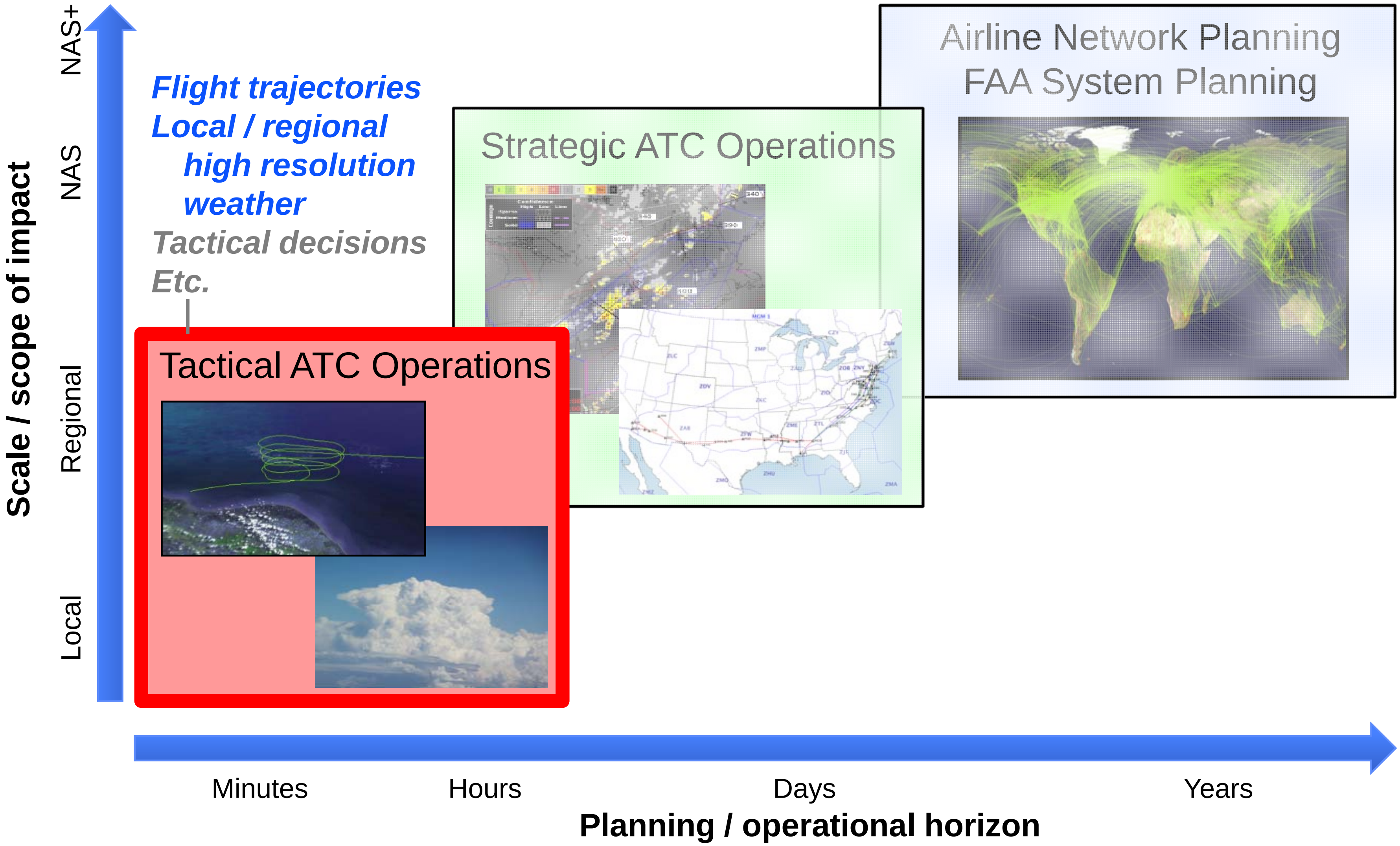
$\Gamma_{m(t)}$ Delay-state dependent system matrix
Derived from network delay matrix

π_{ij} Probability of transition from delay state i to state j

Delay / demand prediction modeling

Control strategy assessment

Space, Time, Data, and Impacts

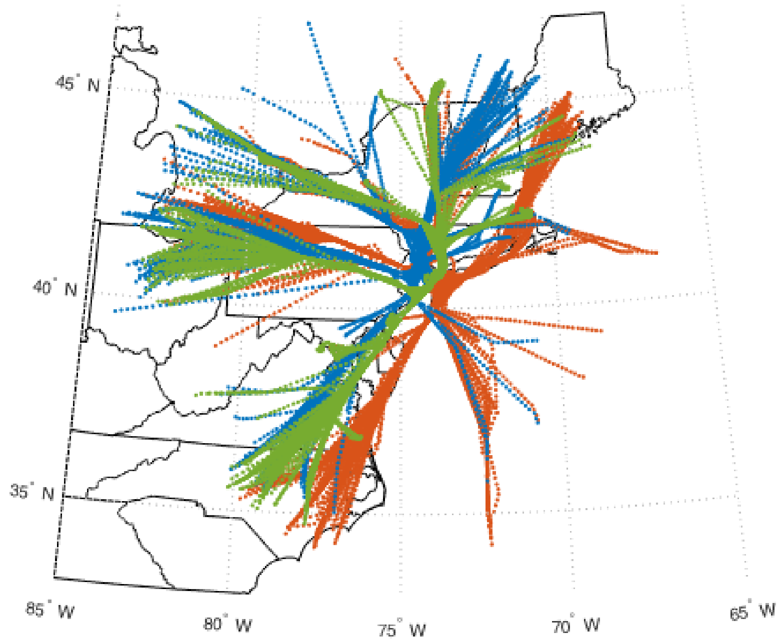


Tactical ATC Operations

NY Metro Focus

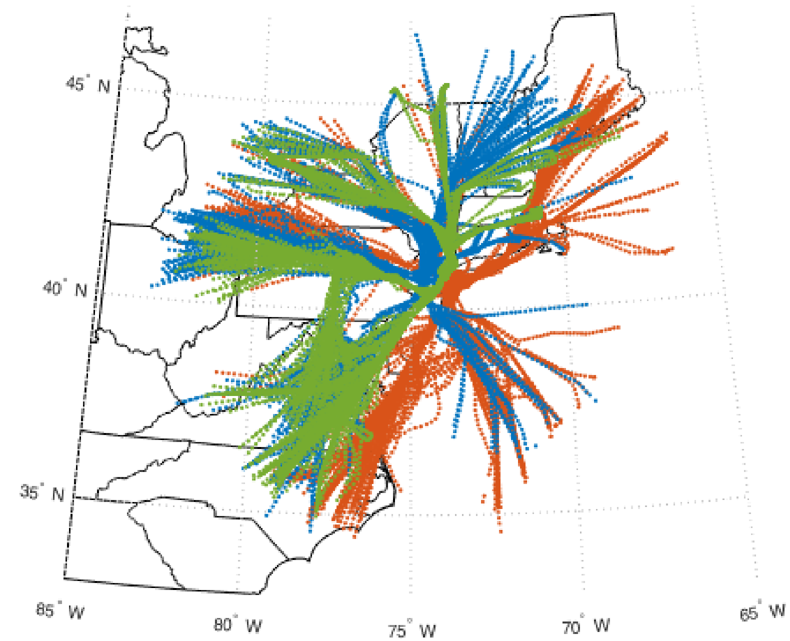
NY Metro Arrival Trajectories

Fair weather operations



Key:
LGA
EWR
JFK

Convective weather operations



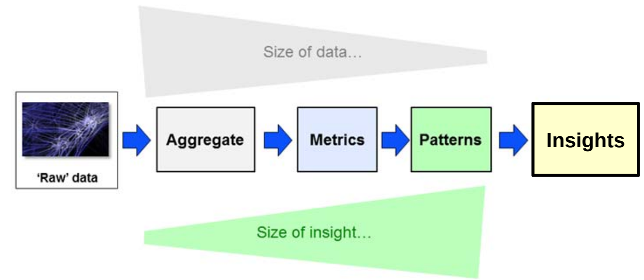
Goal: Develop a generalizable method to characterize tactical use of terminal and transition airspace to guide airspace design and support operational best practices

Approach: Identify patterns of arrival / departure resource use through trajectory analysis and link them to constraints and outcomes

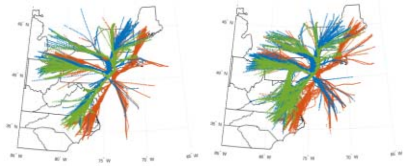
'arrival (departure) resource' = routinely used arrival (departure) path

Tactical ATC Operations: Methodology

Framework key:



Observed trajectories

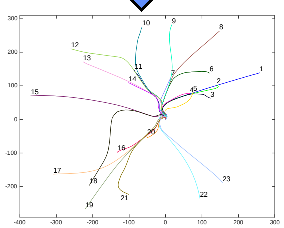


13 day training set

57 day weather impact dataset
1000 day pattern dataset (2013-2015)

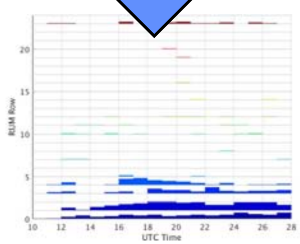
Resource Identification

Cluster trajectories using DBSCAN



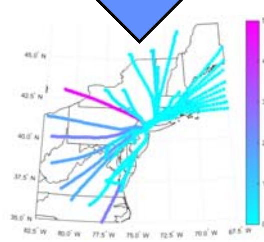
Resource Use

Assign trajectories to resources using Random Forest & identify non-conforming trajectories



Operational Patterns

Cluster Resource Use Vectors to identify patterns of hourly use



<u>Daily Resource Use Matrices</u>

Post-event analysis of operational dynamics

<u>Hourly Resource Use Vectors</u>

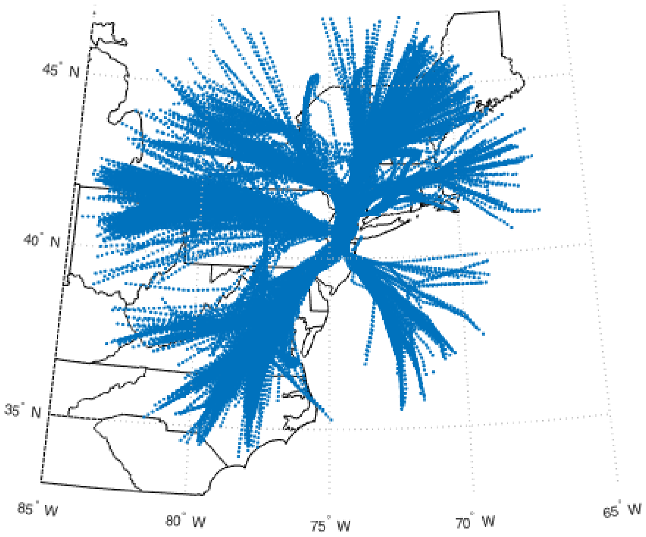
Real time operational dynamics

<u>Hourly Resource Use Patterns</u>

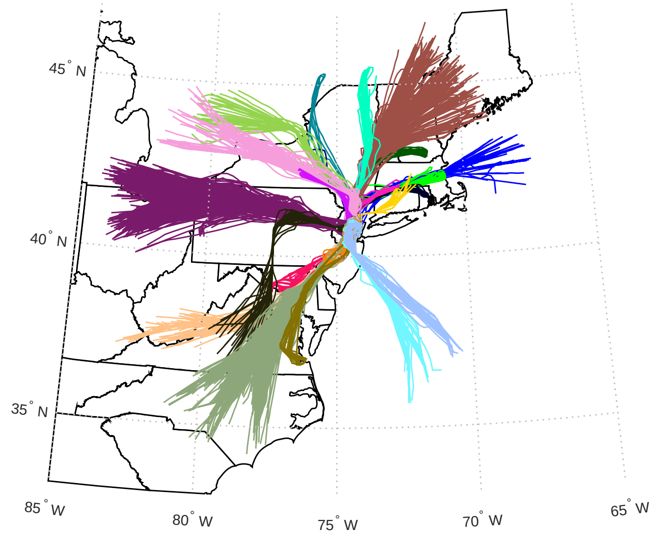
Predictive modeling

Resource Identification

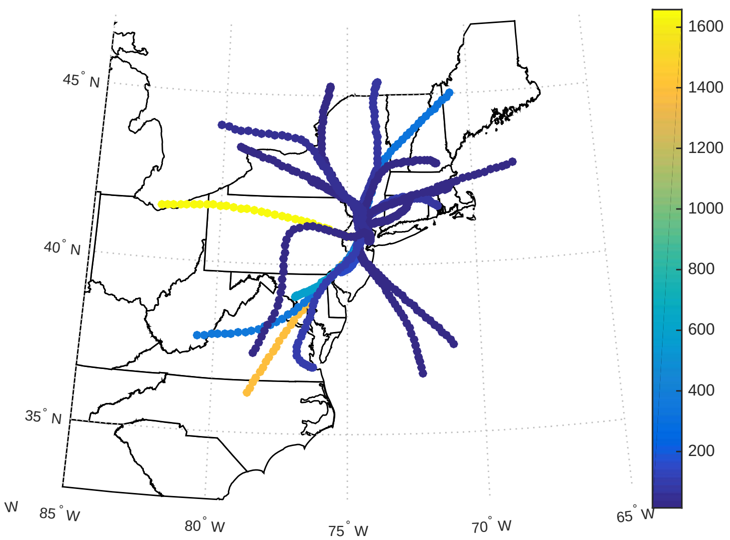
‘Emergence’ of EWR Arrival Resources



13 days of arrivals…



…23 clusters…



…23 cluster centroids = Arrival Resources

- **Cluster algorithm parameterization involves tradeoffs between compactness, separability, and dissimilarity of clusters**
- **Resulting clusters captured ~92% of all trajectories**

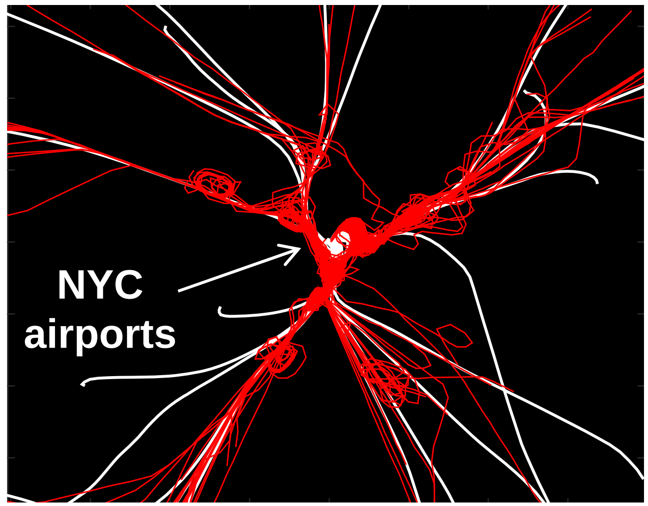
Resource Assignment and Non-conformance: JFK Arrivals

Trajectories assigned to Arrival Resources
(all conforming)

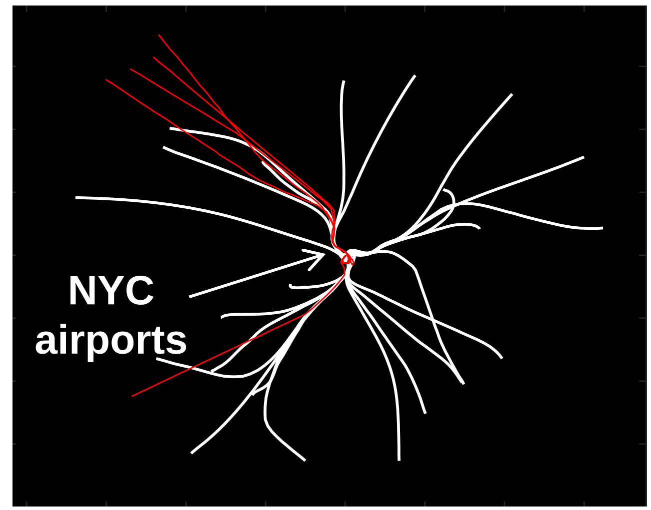
October 8, 2014

Illustrations of non-conformance

Non-conforming trajectories
Arrival resources



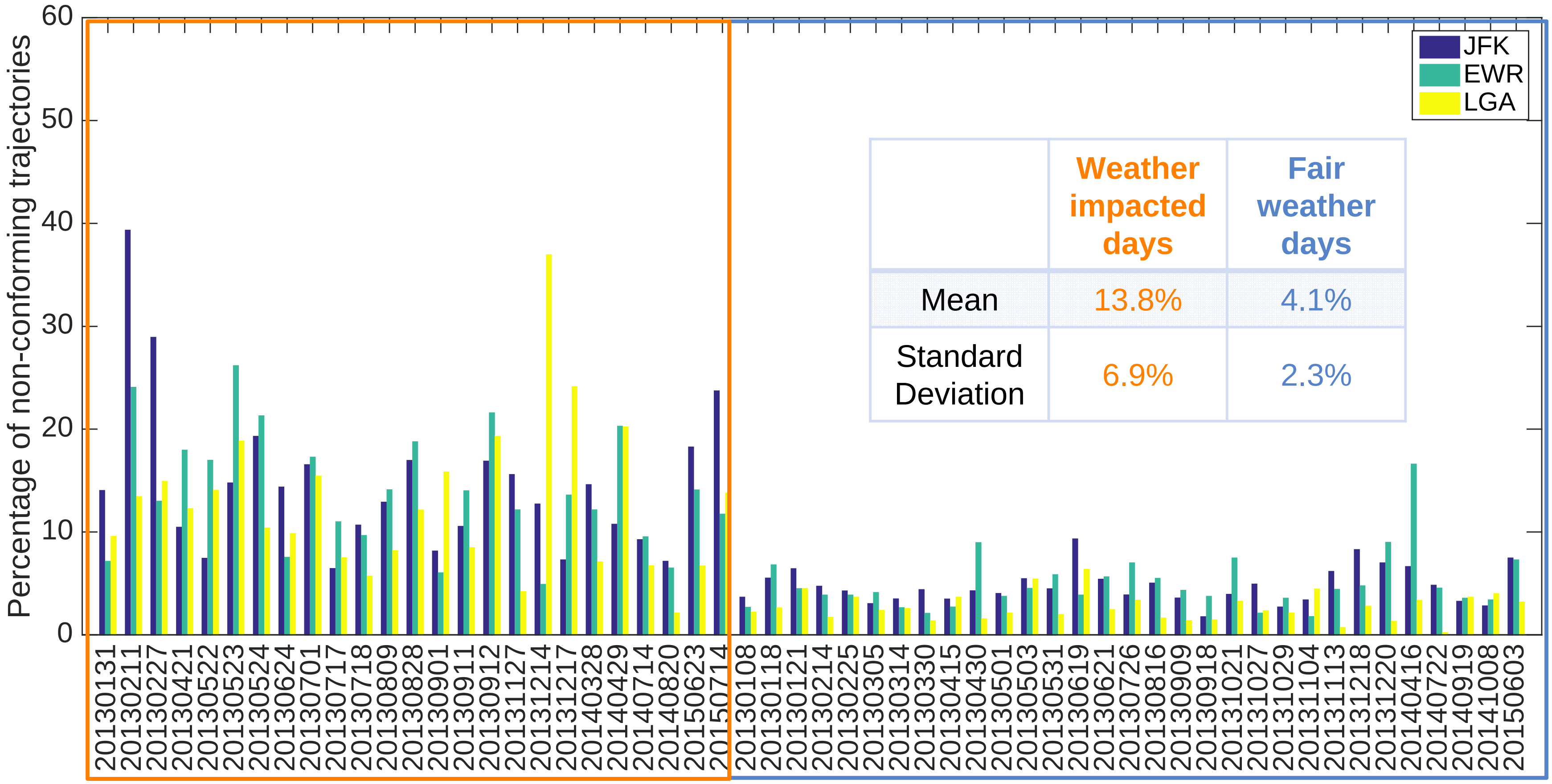
February 11, 2013



September 9, 2013

- **Random Forest trajectory classification assigns individual trajectories to resources and identifies non-conforming trajectories**
- **Non-conforming trajectories take many forms**
  - **Dynamically alter flow structure**
  - **Workload consequences for Air Traffic Control?**

Non-conformance and Weather

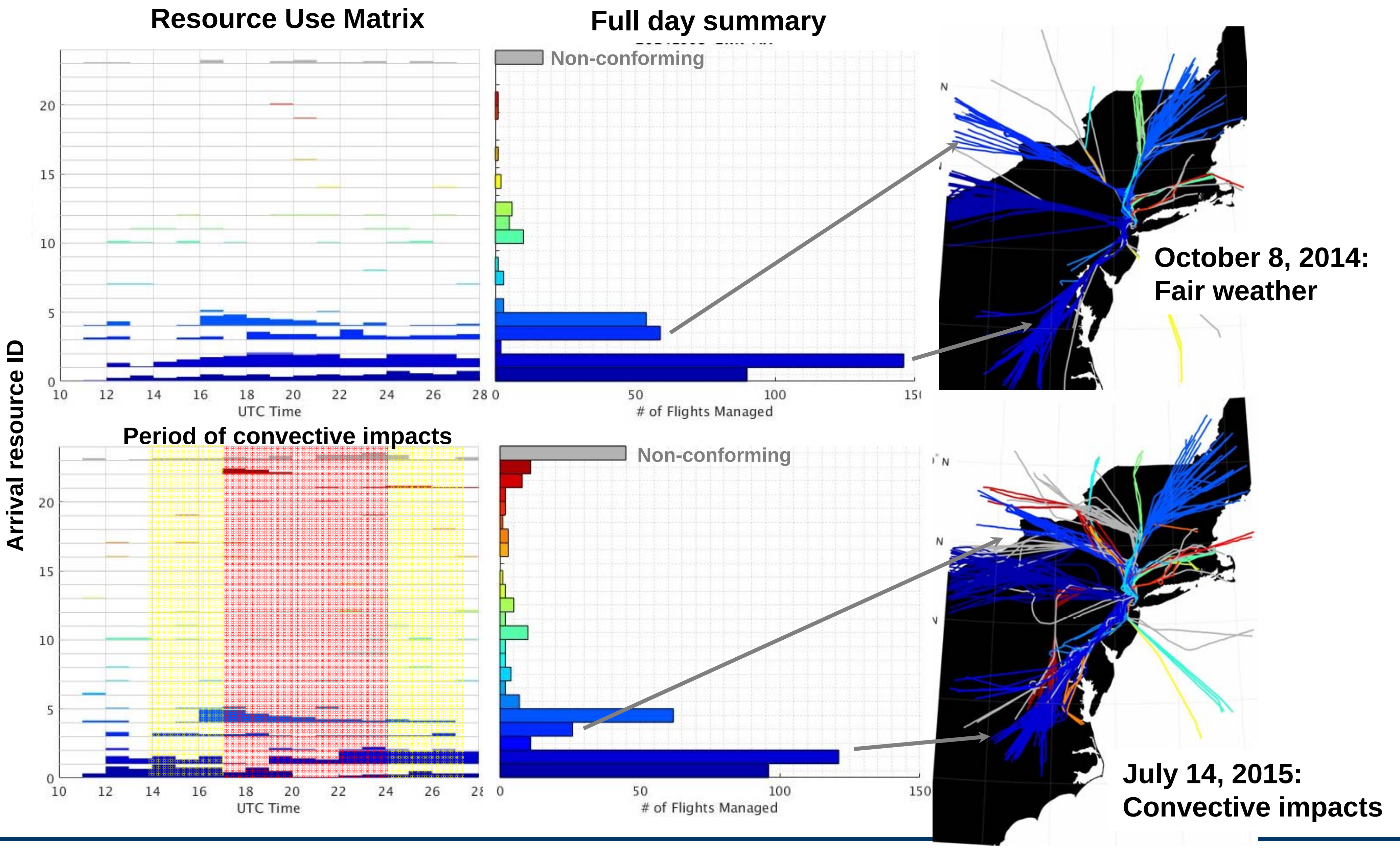


	Weather impacted days	Fair weather days
Mean	13.8%	4.1%
Standard Deviation	6.9%	2.3%

- **Trajectories assigned for dataset of 56 days including weather impacted (convection or adverse winds / ceiling / visibility) and fair weather days**
- **Significant increase in non-conforming trajectories during weather impacted days**

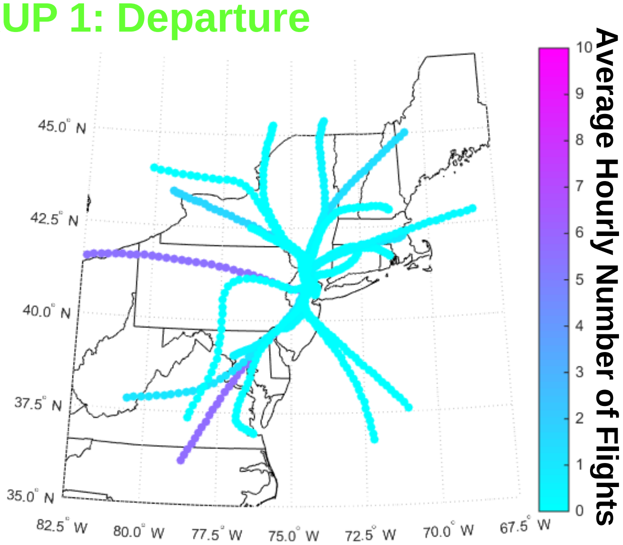
NY Metro Operational Dynamics

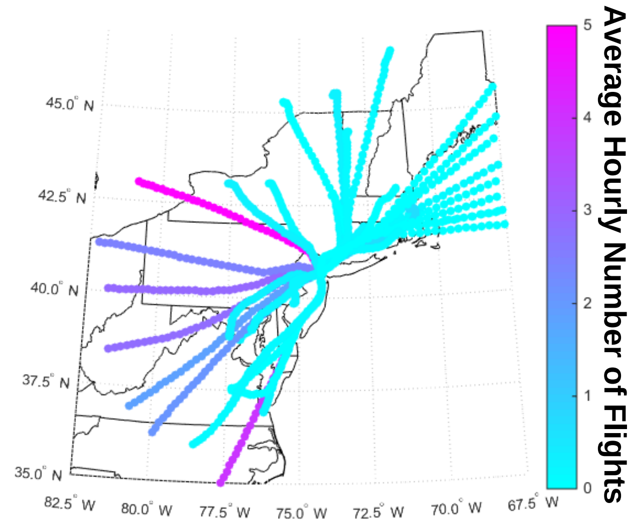
A Tale of Two Days… (EWR Arrivals)

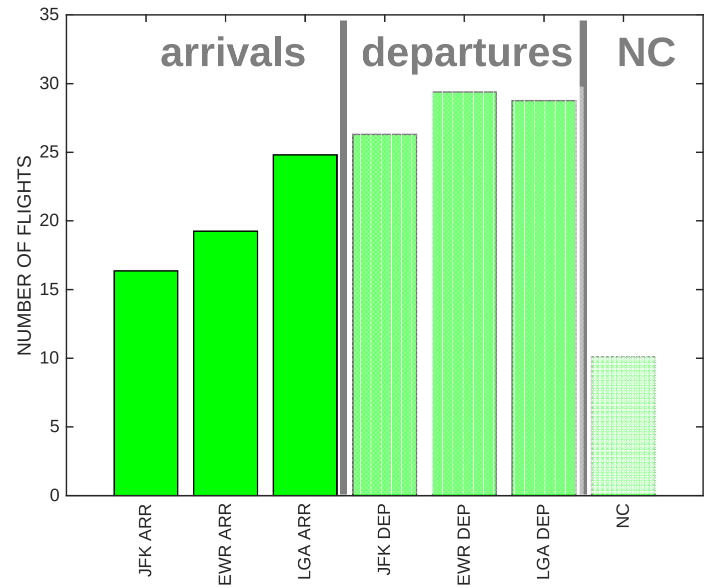


Hourly Resource Use Patterns (RUP)

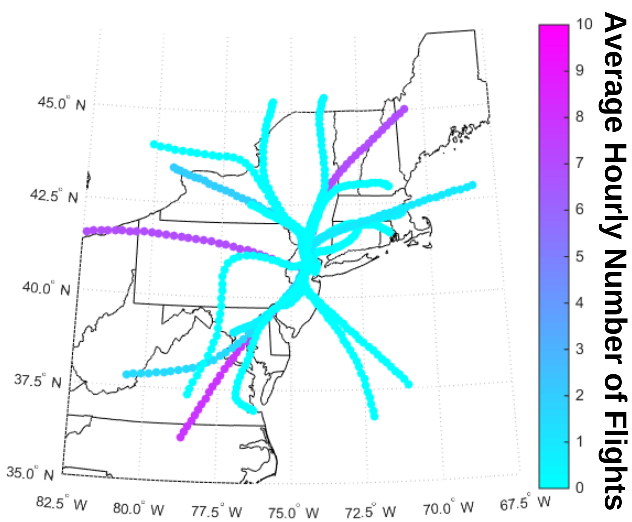
RUP 1: Departure



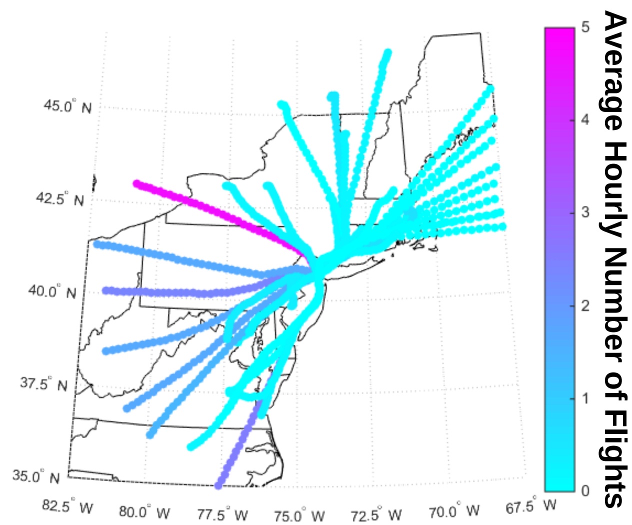




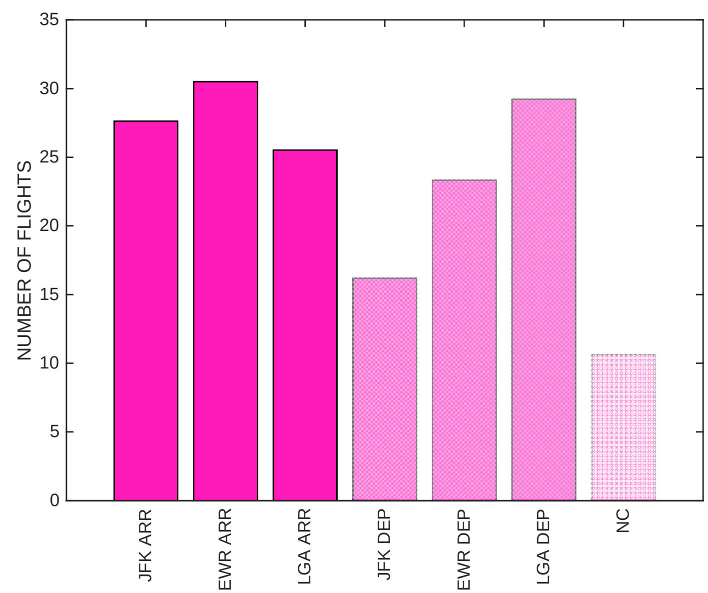
RUP 2: JFK, EWR Arrival



EWR Arrivals

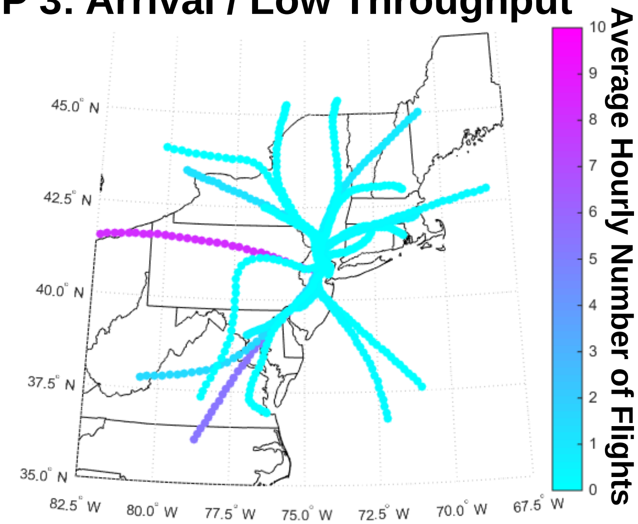


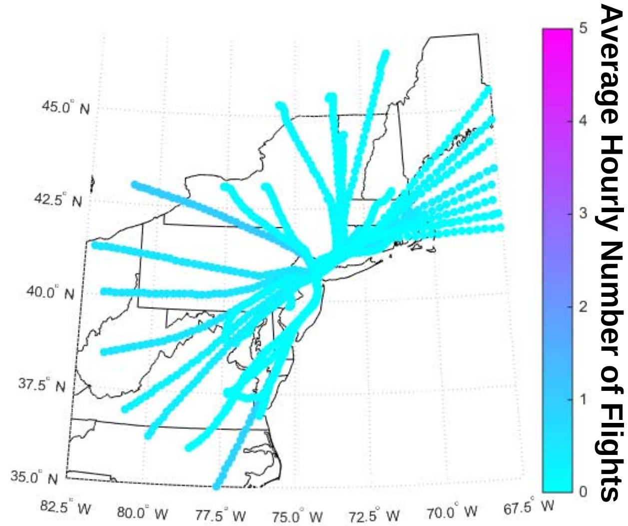
EWR Departures

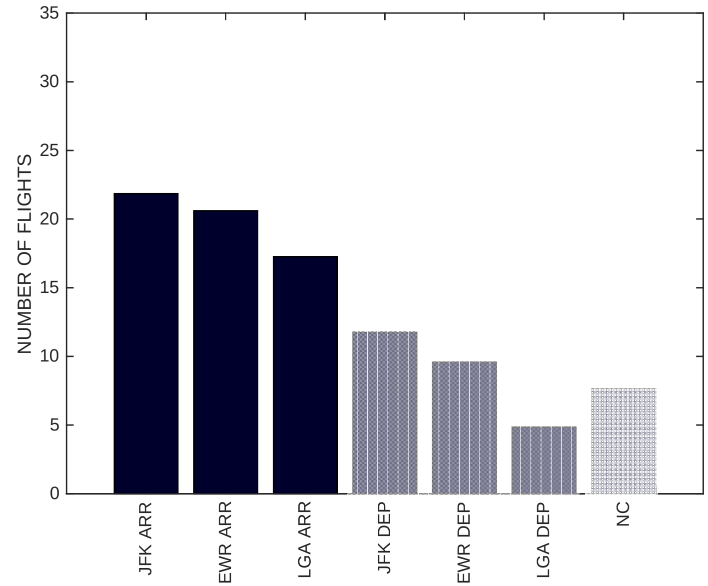


Hourly Resource Use Patterns (RUP)

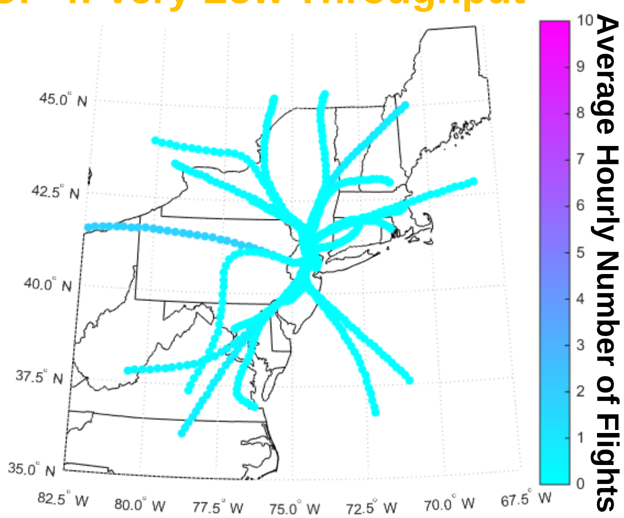
RUP 3: Arrival / Low Throughput



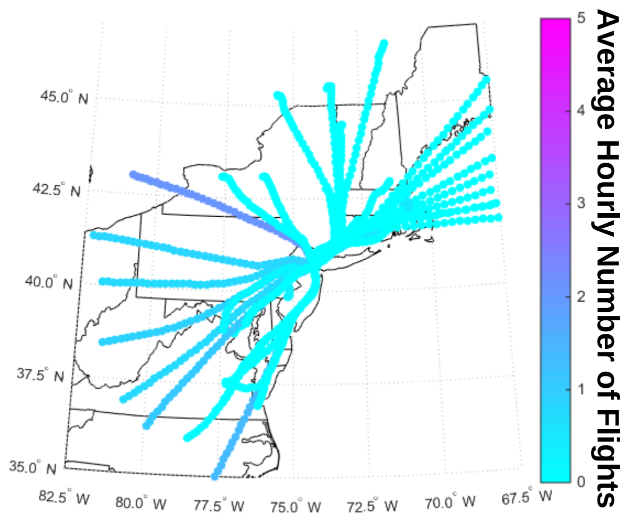




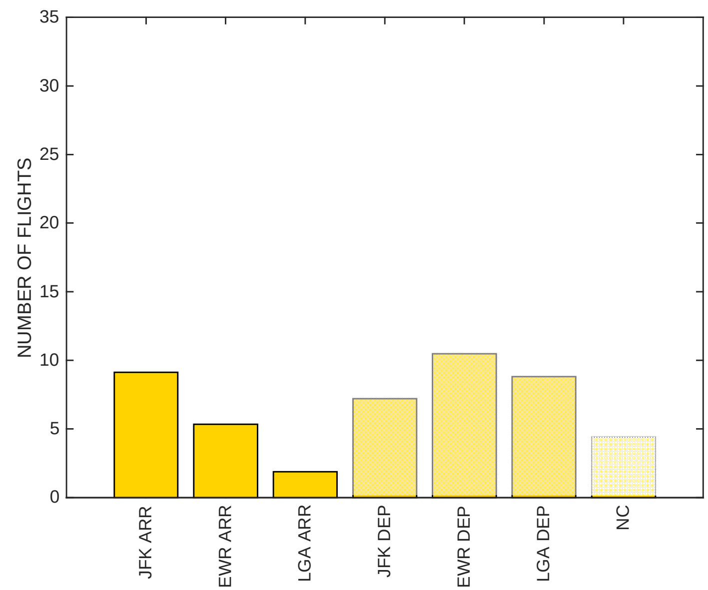
RUP 4: Very Low Throughput



EWR Arrivals

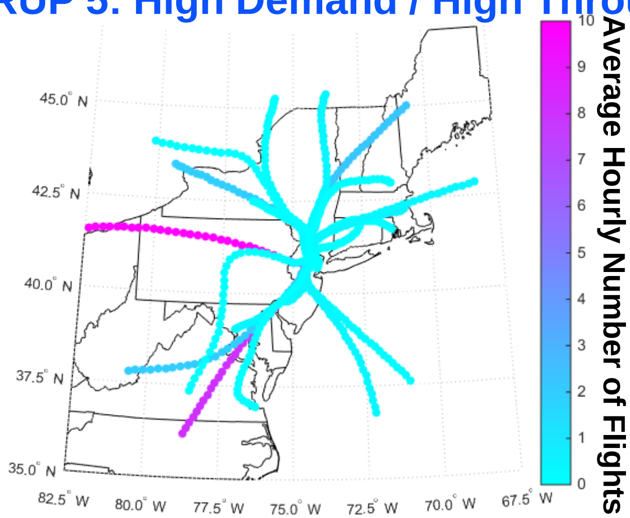


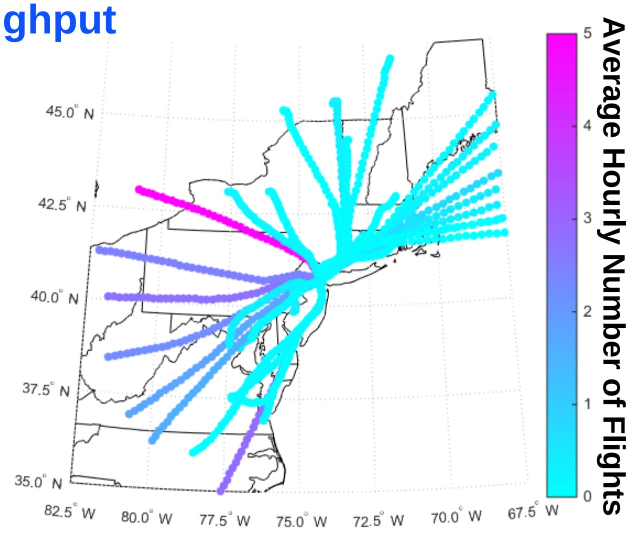
EWR Departures

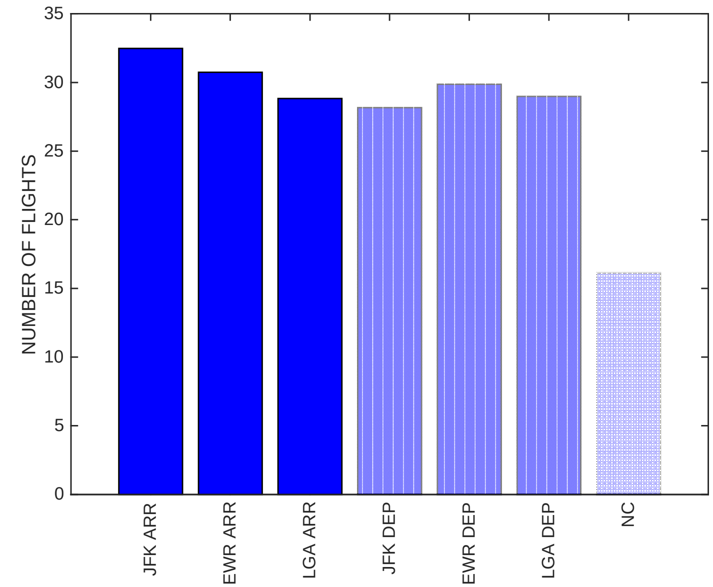


Hourly Resource Use Patterns (RUP)

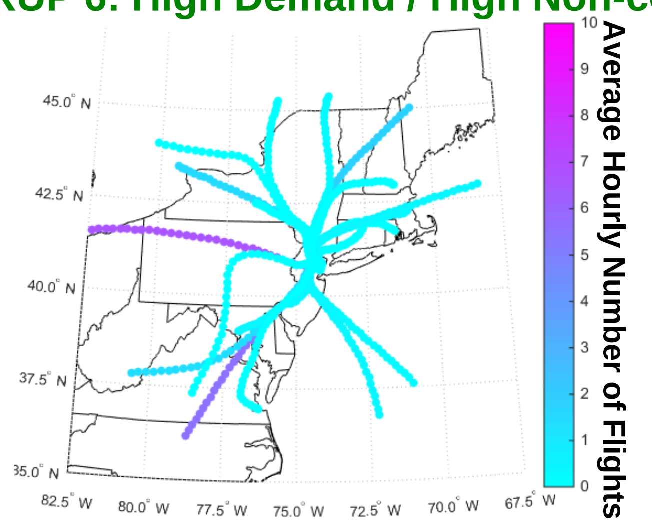
RUP 5: High Demand / High Throughput

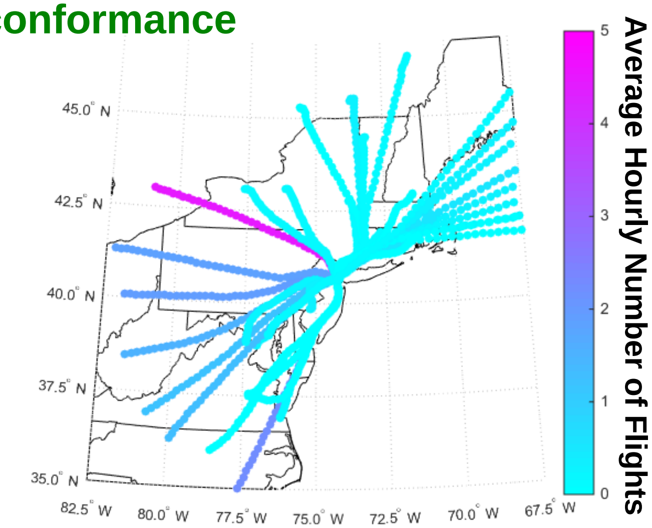


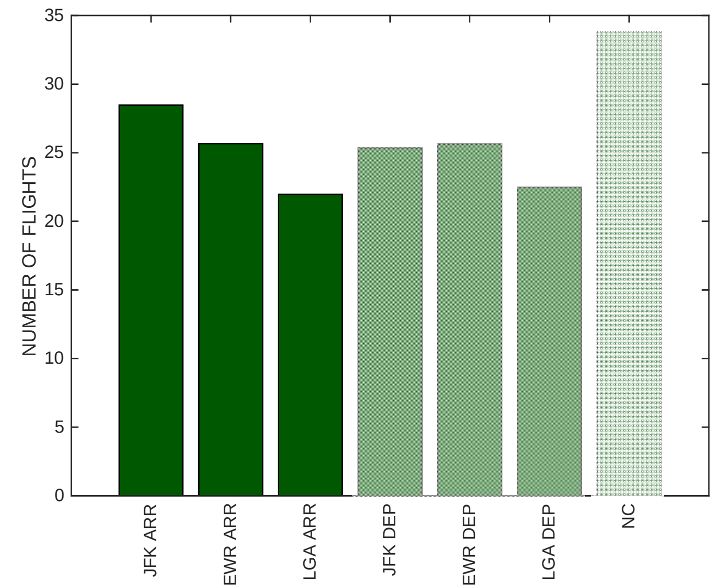




RUP 6: High Demand / High Non-conformance





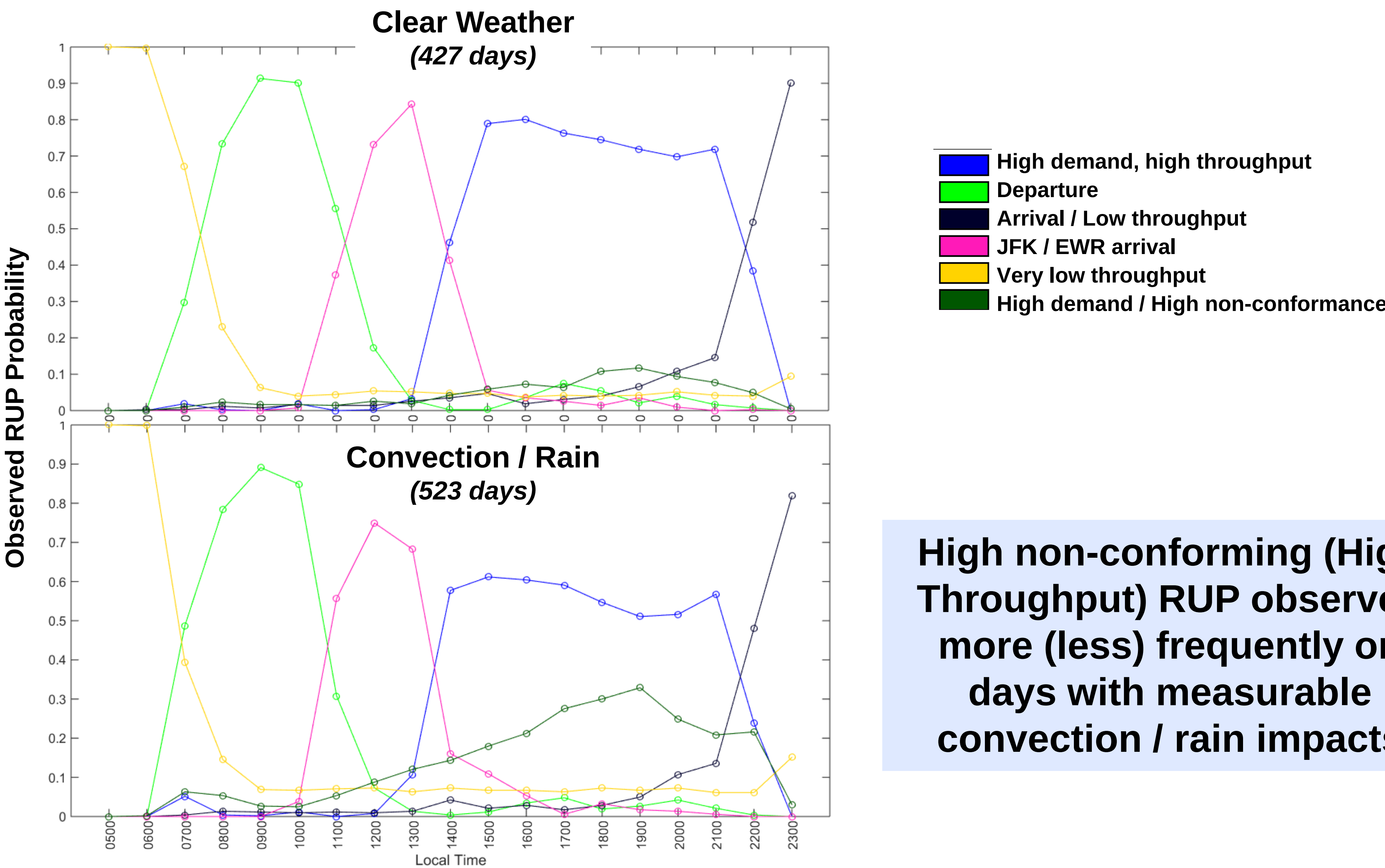


EWR Arrivals

EWR Departures

Occurrence of Resource Use Patterns

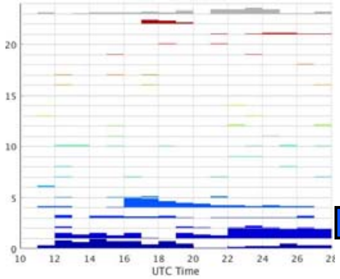
By Hour



High non-conforming (High Throughput) RUP observed more (less) frequently on days with measurable convection / rain impacts

Tactical ATC Operations: Next Steps

Resource Use Matrices



Weather impact / constraint

Clustering to identify days with similar constraints, resource use

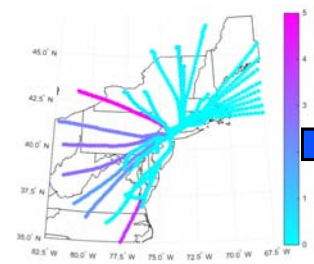
Constraint-normalized performance assessment

Case day identification / scenario generation

Daily Aggregations

Hourly Aggregations

Resource Use Patterns



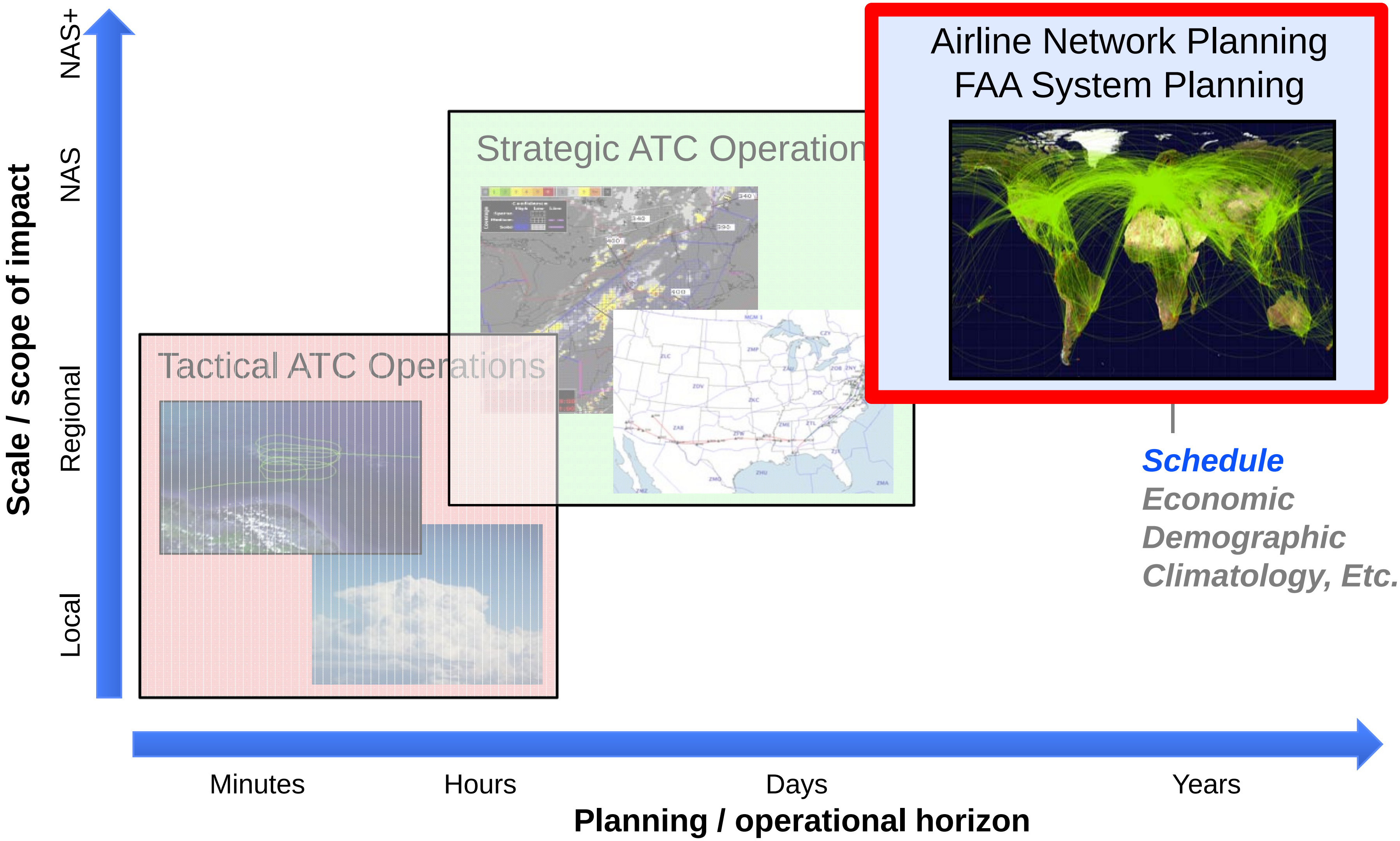
Weather impact / constraint

Correlation of Resource Use Patterns with constraints, demand

Constrained capacity modeling and prediction for decision support

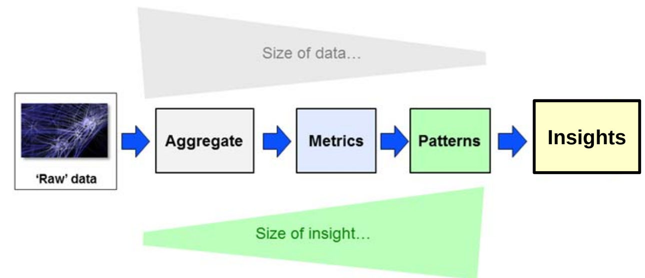
Development of best practices

Space, Time, Data, and Impacts



Air Carrier Competition: Methodology

Framework key:

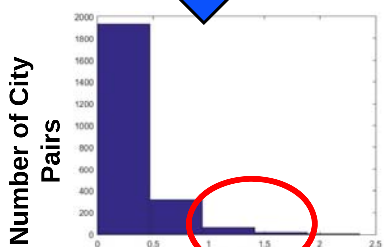


United States Department of Transportation
Office of the Assistant Secretary for Research and Technology
Bureau of Transportation Statistics

2000 - 2014

Extract all city pairs → **Identify top 40 routes Calculate # of flights, # of airlines on each** → **Define use, competition network structures** → **Annual Route Use, Competition Networks**

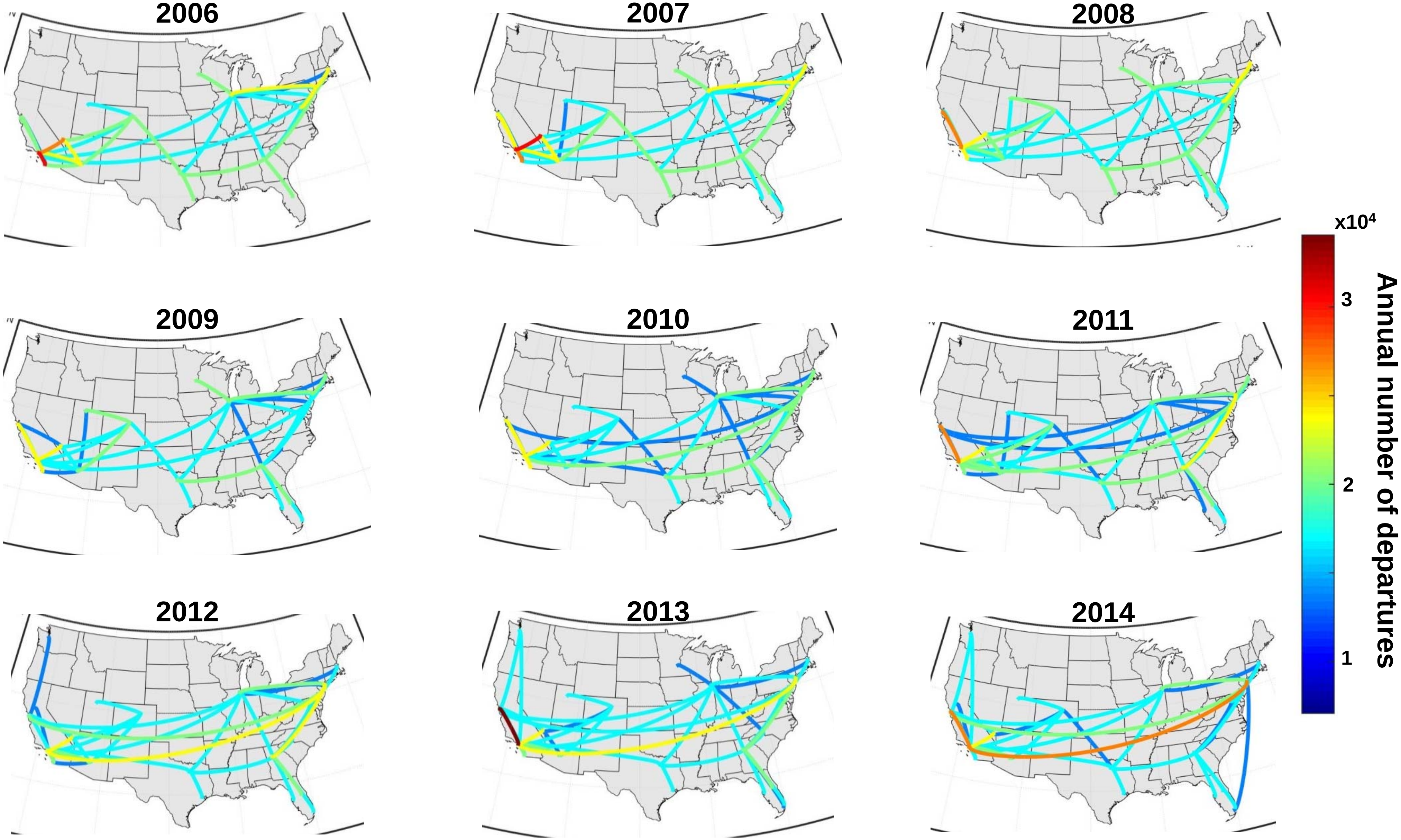
Inputs to Strategic Operations analyses
Basis for predictive models to guide capital investment



Number of Flights on route (x 10^4)

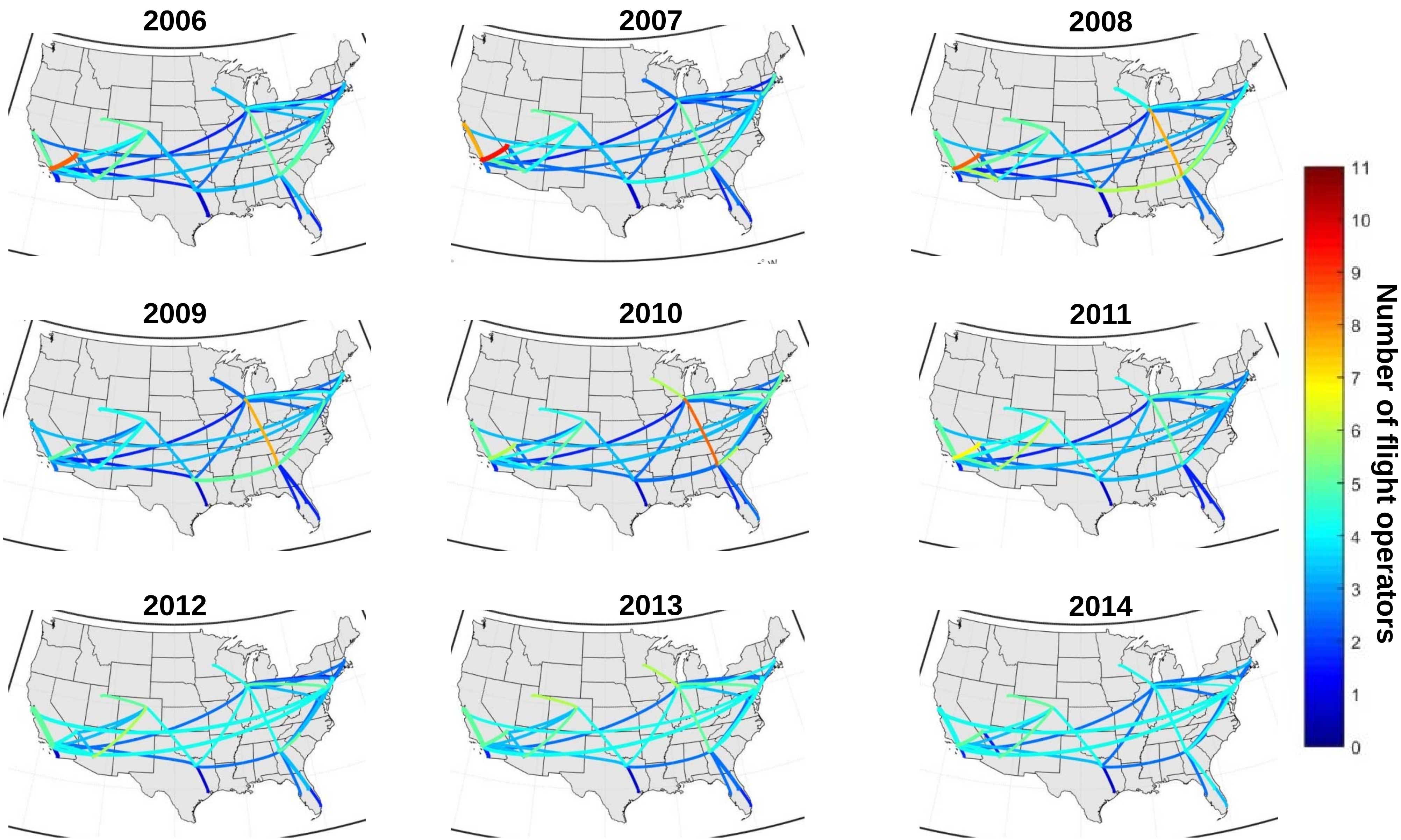
Top 40 Routes

By number of operations



Competition on Top 40 Routes

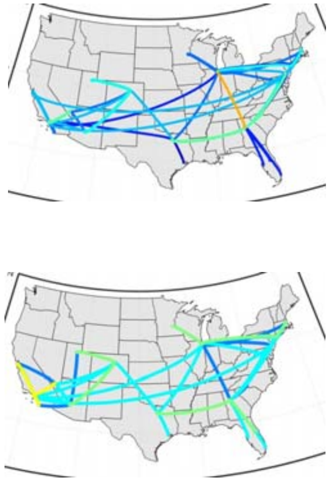
Number of airline operators



Air Carrier Competition: Next Steps

Network operations

Market competition



Characterize operational and competitive network structure as weighted connectivity matrices

Effect of structure on annual NAS performance measured by delay, cancellation

Correlation to observed frequency of Delay, Cancellation States

Outline

- Motivation: Air transportation system challenges and Big Data opportunities
- Technical approach & Selected results:
  – Strategic ATC Operations
  – Tactical ATC Operations
  – Airline Network Planning
- Summary of innovations, Potential impacts and Next step recommendations
- Distribution / Dissemination & Acknowledgements

Phase 1 Innovation Summary

- **Developed Big Data analysis framework using novel metrics & analytics to provide new insight across a range of fundamental scales in air transport:**

	Aggregate	Metrics	Patterns	Insights
Tactical ATC Operations	• Terminal area trajectory clustering under range of operating conditions	• Assignment of trajectories to standard resources • Determination of non-conforming flights	• Identification of small number of key resource use patterns	• Resource use pattern dynamics across airport locations and operating conditions
Strategic ATC Operations	• Airport-pair delay and cancellation weighted directional connectivity matrices	• NAS network hub and authority scores at range of temporal scales • Assessed over multi-years	• Identification of small number of key NAS-wide delay and cancellation states	• System-wide delay and cancellation dynamics across operating conditions
Airline/FAA Planning	• Airline network definitions across decades	• Top route and competition evolutions over decades	• Identification of dominant scheduled routes • Competition dynamics	• Network structural evolution over time • Initial correlations of network structure with external influences

- **Insights provide foundation for performance evaluation and predictive models**

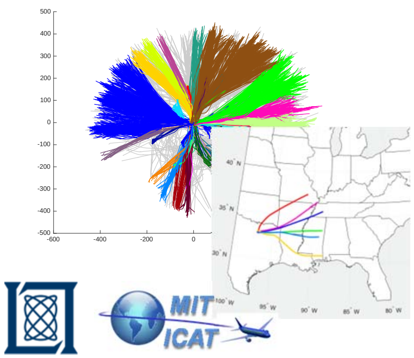
Phase 1 Innovation & Impact Summary => Phase 2 Recommendations

	Phase 1	Phase 2
Data Layer	• Flight trajectories • Flight delay • Weather • Cancellations • Schedules	• Traffic Management Initiatives • Emerging data types (FAA SWIM, other?) • Database structure & technology
Analytics Layer	Tactical ATC Operations Analysis Strategic ATC Operations Analysis Airline/ FAA Planning Analysis	• Refinements across areas • Extensions where appropriate
Application Layer	• *Diagnostic* system characterization • Baseline, anomaly, scenario identification	• *Predictive* modeling • Control action analysis • Tool building (visualization & analysis)
Impact & Tech Transfer Layer	• NASA: technical interchange meetings • Other: Publications	• NASA: tools for integration into existing programs • FAA / Industry: performance analysis

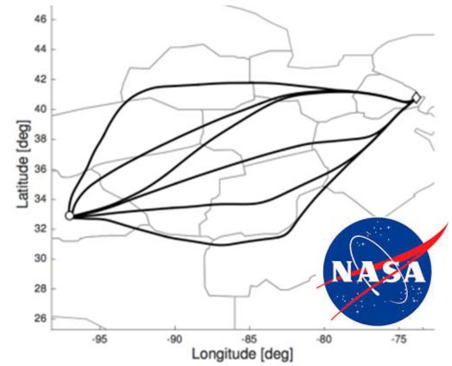
Tech transfer opportunities inform research needs

Current & Potential Future Connections to NASA Efforts

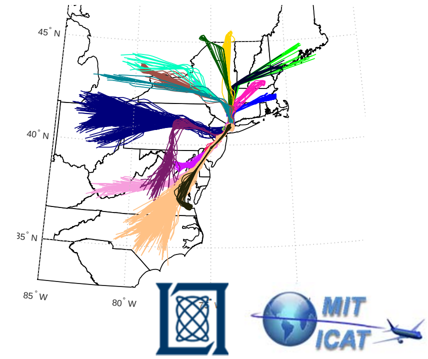
LEARN Phase 1 / 2
DFW departure resources



NASA ARC
DFW-LGA trajectory prediction



LEARN Phase 1 / 2
LGA arrival resources



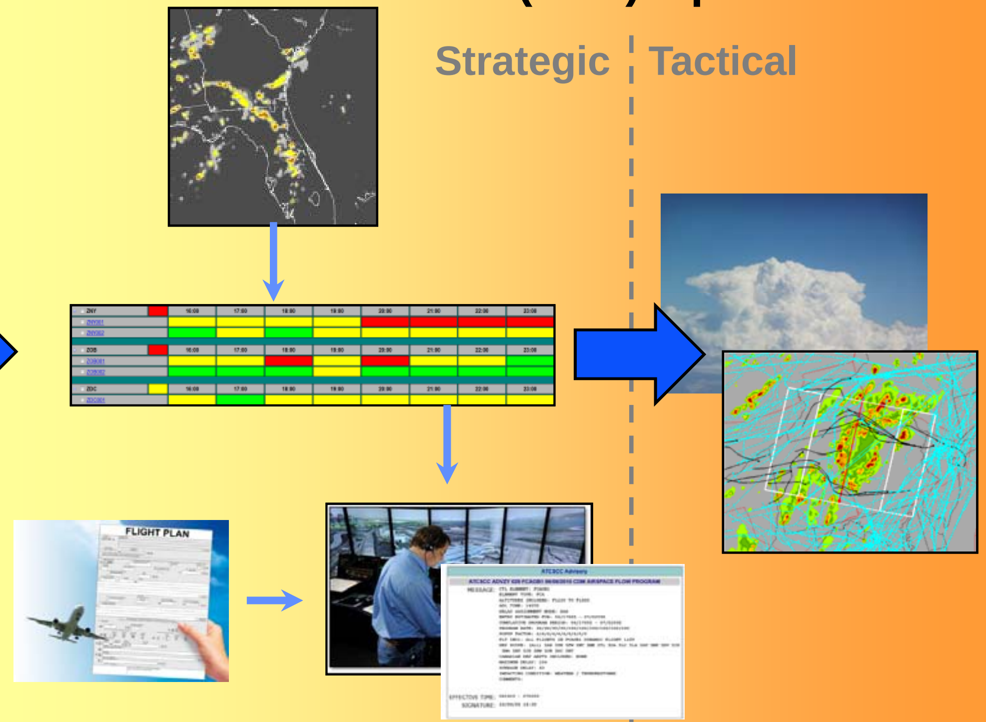
- **Tactical Operations / 4D-TBO: end-to-end modeling of TBO-based traffic management (illustrated)**
- **Strategic, Tactical Operations / SMART-NAS Testbed: real-time analytics and visualization tools**
  - **Simulation modules**
  - **Review of archives to identify case studies and define scenarios**
- **All / Sherlock Data Warehouse: information models for analytic products**

4D-TBO = 4 Dimensional-Trajectory Based Operations
SMART-NAS = Shadow Mode Assessment Using Realistic Technologies for the National Airspace System

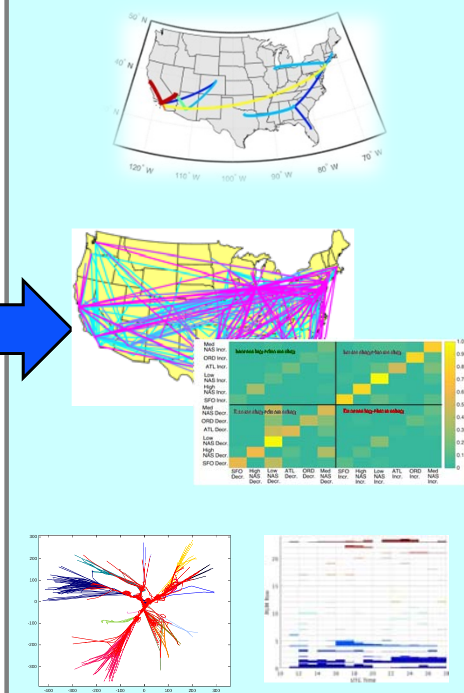
Ultimate Impact: Influencing Future National Airspace System Operations

System Planning

Air Traffic Control (ATC) Operations



Analytics



Structural inefficiencies
Capital needs projection

Performance-driven best practices
(post-event analysis)
Operational decision support
(real-time predictive models)

Outline

- **Motivation: Air transportation system challenges and Big Data opportunities**
- **Technical approach & Selected results:**
  - **Strategic ATC Operations**
  - **Tactical ATC Operations**
  - **Airline Network Planning**
- **Summary of innovations, Potential impacts and Next step recommendations**
- **Distribution / Dissemination & Acknowledgements**

Distribution/Dissemination

- **Papers**
  - **"Multi-Scale Data Mining for Air Transportation System Diagnostics", accepted to *16th AIAA Aviation Technology, Integration, and Operations Conference*, 13-17 June 2016, Washington DC.**
  - **"Clusters and Communities in Air Traffic Delay Networks", accepted to *2016 IEEE American Control Conference*, 6-8 July 2016, Boston, MA.**
  - **"A Visual Analytic Platform for Air Traffic System Strategic and Tactical Operational Evaluation and Control", accepted to *2016 Integrated Communications Navigation and Surveillance (ICNS) Conference*, 19-21 April 2016, Herndon, VA.**
  - **"Airline Network & Competition Characterization using Big Data Approaches", to be submitted to *35th Digital Aviation Systems Conference*, 25-29 September 2016, Sacramento, CA.**
- **Presentations**
  - **"Big Aviation Data Mining for Robust, Ultra-Efficient Air Transportation", Kick-off Meeting & Overview for NASA ARC Aviation Systems Division researchers, NASA Ames Research Center, 4 April 2015.**
  - **"Big Aviation Data Mining for Robust, Ultra-Efficient Air Transportation", Status report & Technical Interchange Meeting for specific NASA ARC ASD programs, NASA Ames Research Center, 18-19 November 2015.**
- **Other**
  - **Numerous telcons with NASA researchers to discuss potential mutual value from collaboration (including SMART-NAS, 4D-TBO, Sherlock data warehouse programs)**

Acknowledgments

- **Many thanks to the following:**
  - **NARI** for supporting the project and promoting collaboration
  - **Sarah D'Souza and Michael Bloem,** NASA ARC for providing excellent technical oversight and helping connect us to relevant NASA researchers
  - **NASA ARC program researchers** for their invaluable technical discussions, feedback on our approach and identification of relevant problem areas
    - **4D-TBO (Paul Lee, Heather Arneson, Tony Evans, …)**
    - **SMART-NAS (John Robinson, Kee Palopo, Gano Chatterji, …)**
    - **Sherlock data warehouse team (Michelle Eshow, Rich Keller, Ron Reisman, …)**
    - **William Chan (Branch Chief)**
    - **Sandy Lozito (Division Chief)**

Thank you!