Improving Air Traffic Operations through Data Analysis and Modeling

Hamsa Balakrishnan

Aeronautics and Astronautics

Massachusetts Institute of Technology

Joint work with Diana M. Pfeil , Ioannis Simaiakis and Harshad Khadilkar

NASA Conference on Intelligent Data Understanding

October 20, 2011

Air Transportation Systems

- Air transport is a key factor in global travel and commerce
  - 600M passengers/year in the US
  - 35,000+ commercial flights/day in the US
  - US traffic expected to grow ~2-3x by 2025 (rel. to 2004)
- NextGen is the “Next Generation Air Transportation System”
- Objectives:
  - Expanding capacity
  - Ensuring safety
  - Protecting the environment
  - Retaining US leadership in global aviation
  - Ensuring national defense and securing the nation

Practical algorithms for air transportation systems

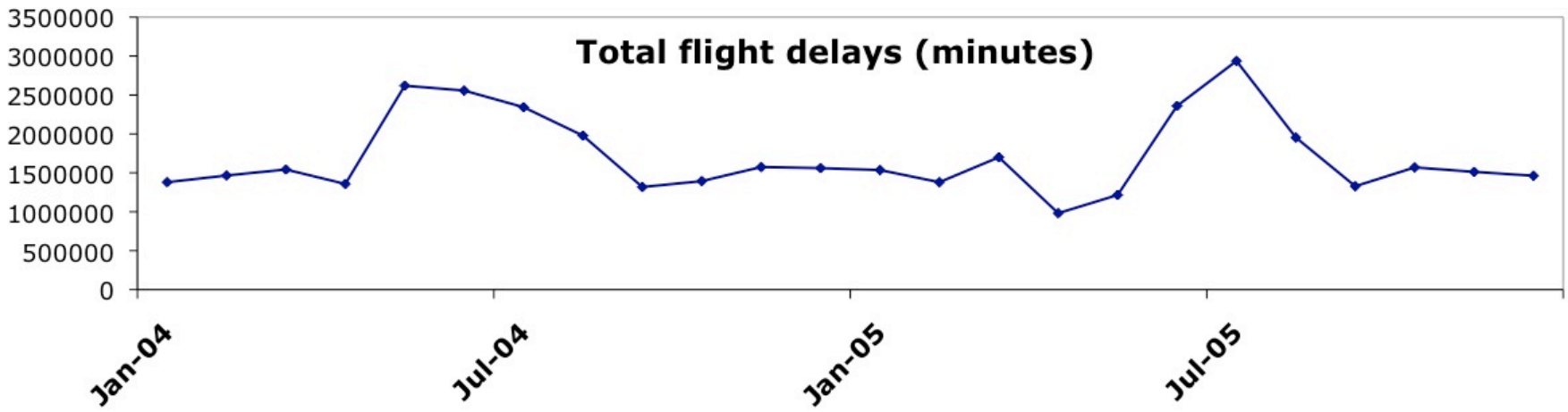
- Goal
  - Develop algorithms that increase efficiency and robustness, and ensure safety…
  - … while coping with uncertainty, **competing interests and** environmental concerns
- Our approach: Leverage large amounts of available data (e.g., weather, trajectories, operations) to
  - Build simple models for desired objectives and operational constraints
  - Develop scalable control and optimization algorithms
- Practical algorithms and air traffic controller decision-support automation are vital to meet demand increase

Two example problems

- Identifying robust routes through convective weather (with Diana Pfeil)
- Increasing the efficiency of airport surface operations (with Ioannis Simaiakis and Harshad Khadilkar)

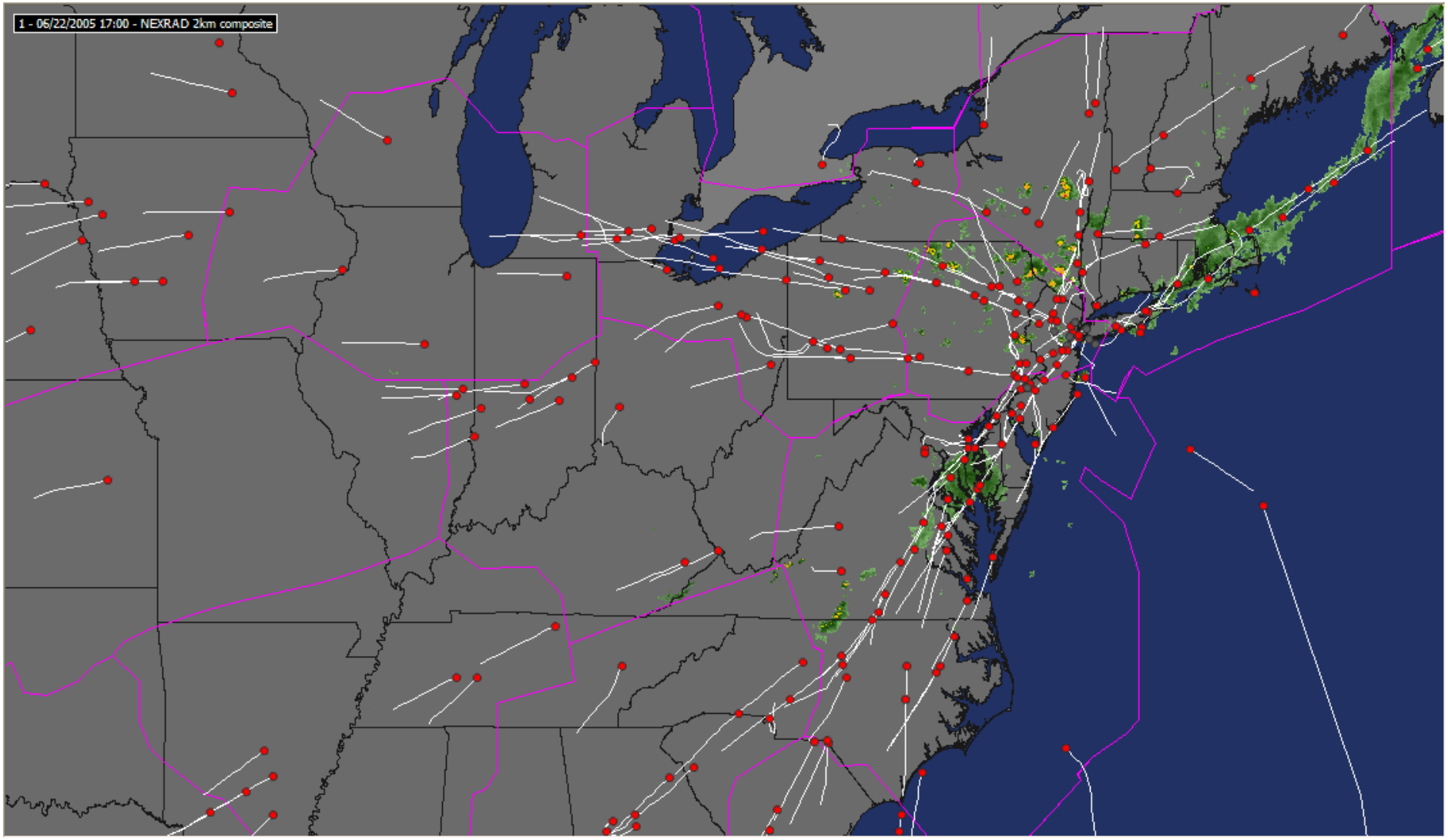
Problem: Bad weather leads to air traffic delays

- Weather is responsible for
  - 66% of total delay minutes in the National Airspace System (2009)
  - 42% of all delayed flights [BTS]
- Convective weather is responsible for a significant fraction of flight delays in the NAS
  - Particular concern in summer months



[FAA OPSNET data]

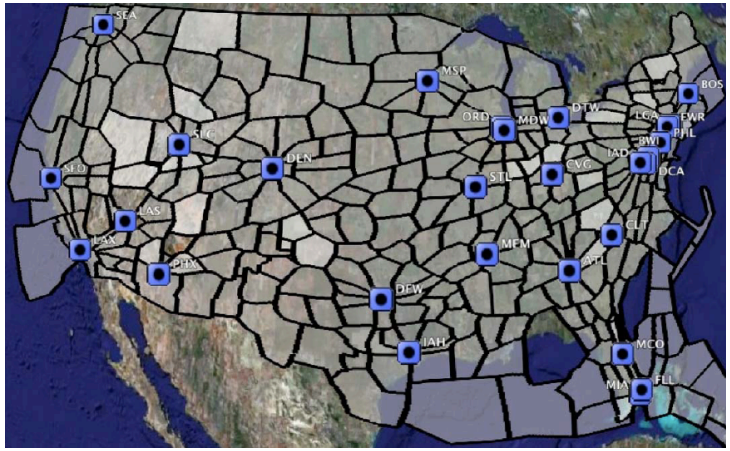
Impact of weather on air traffic flows



[Courtesy MIT Lincoln Labs]

Air Traffic Flow Management

- Make strategic decisions a few hours ahead of operations to balance supply (capacity) and demand of NAS
  - Requires dealing with hazardous weather
  - Typically assumes as input:
    - Expected sector capacity, or
    - Capacity scenarios and probabilities

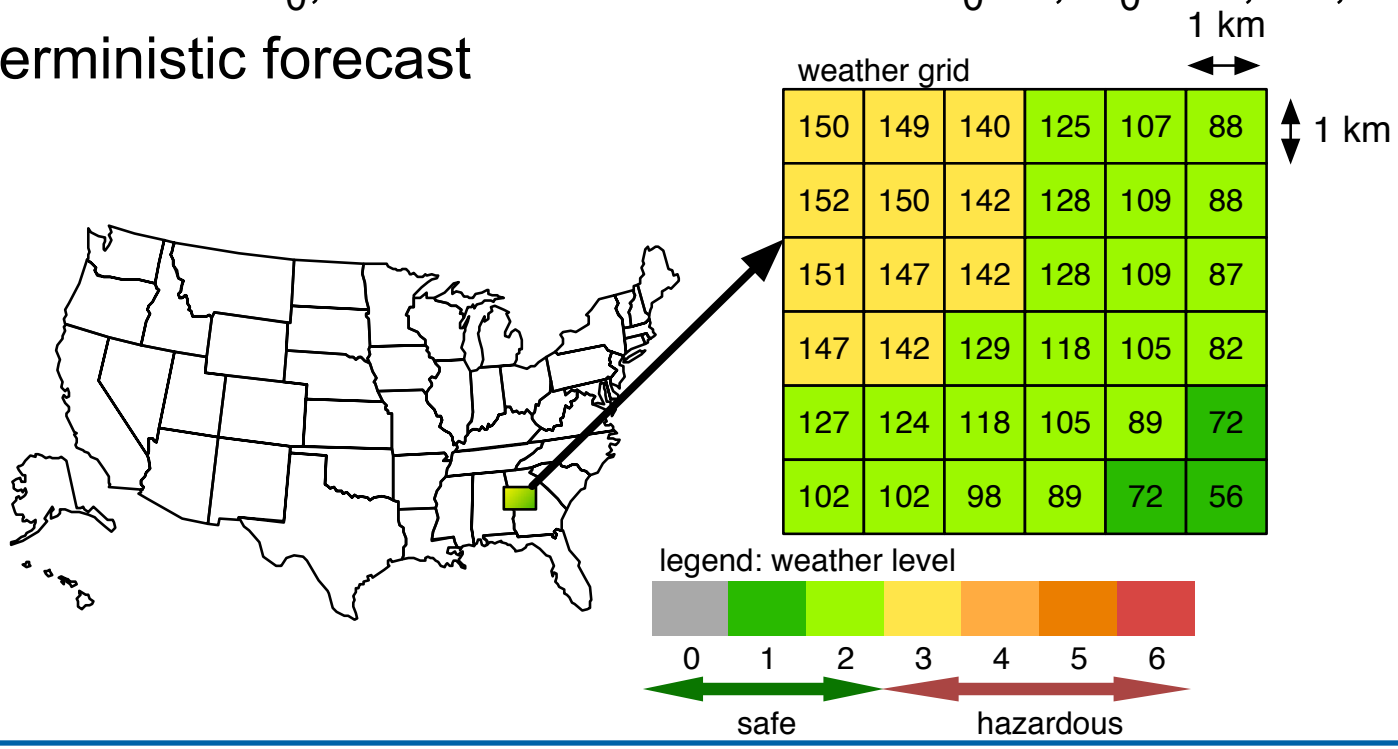


- Our approach:

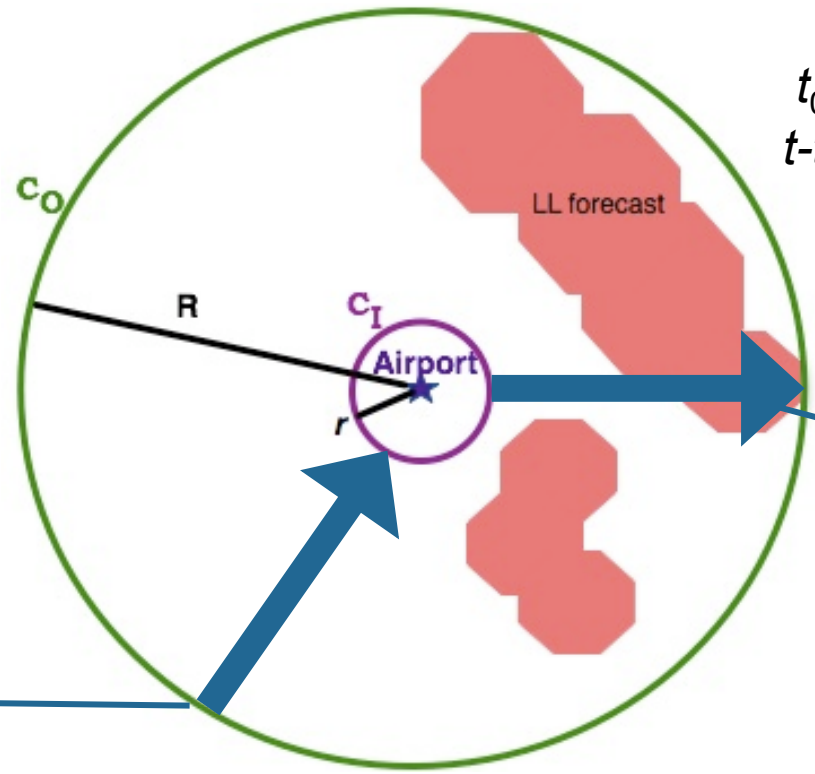
Meteorology → Aviation weather forecasts → Stochastic or robust optimization

Lincoln Lab Convective Weather Forecast (CWF)

- Data set: Each 1km × 1km airspace pixel contains value of Vertically Integrated Liquid (VIL) between 0-255
  - Translated into a “Level”; pilots typically avoid level 3+
- At each time T_0, forecasts available for T_0+5, T_0+10, …, T_0+120
- Deterministic forecast



Solution: Identify open routes from weather forecasts



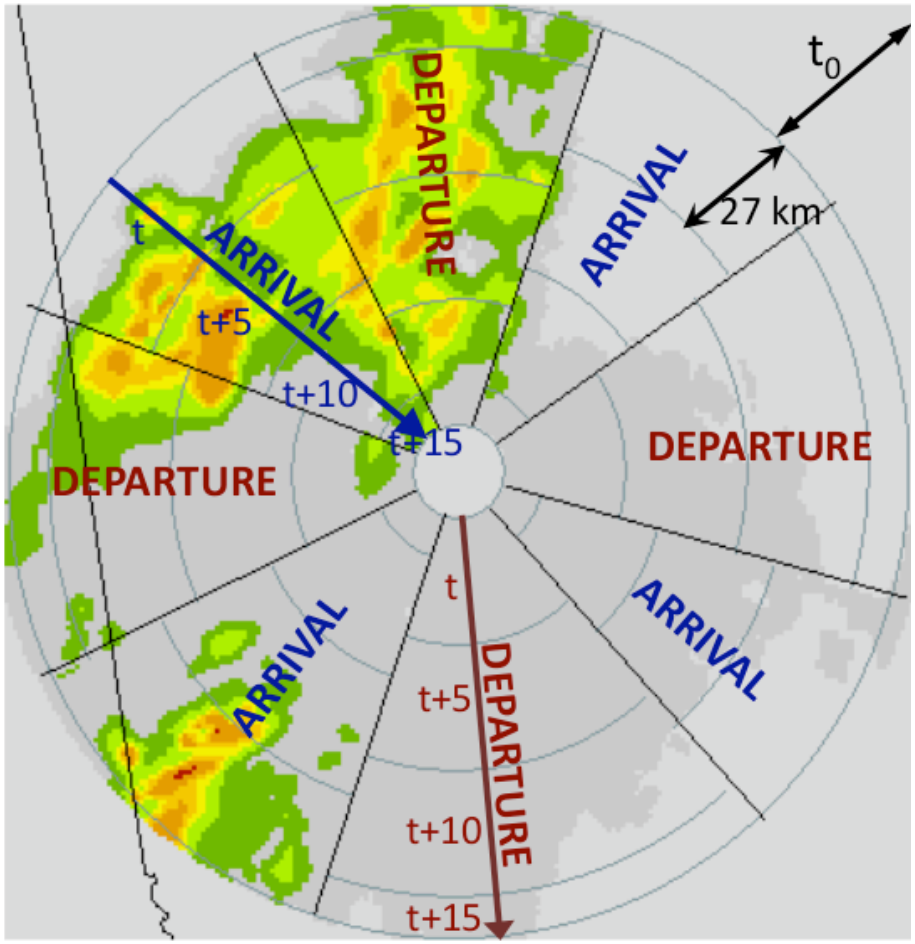
t_0 - planning time horizon
t-t_0 - current time

Potential departure route crosses C_I headed outbound at time t

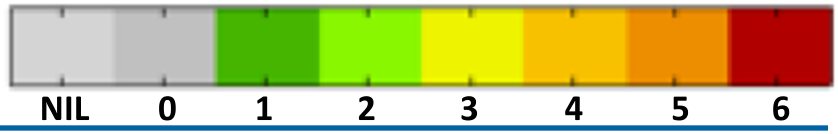
Potential arrival route crosses C_O headed inbound at time t

Given a weather forecast for some time in the future and a set of predetermined potential routes, which routes can be flown in the actual weather that materializes, and with what probability?

Dynamic weather grid and route flexibility



- Limited flexibility to modify routes in practice
- Route *P* is **open** in the observed weather if there exists a route that
  - Does not pass through any node containing Level 3+ weather
  - Is within *B* km of *P* (allows for wiggle room)

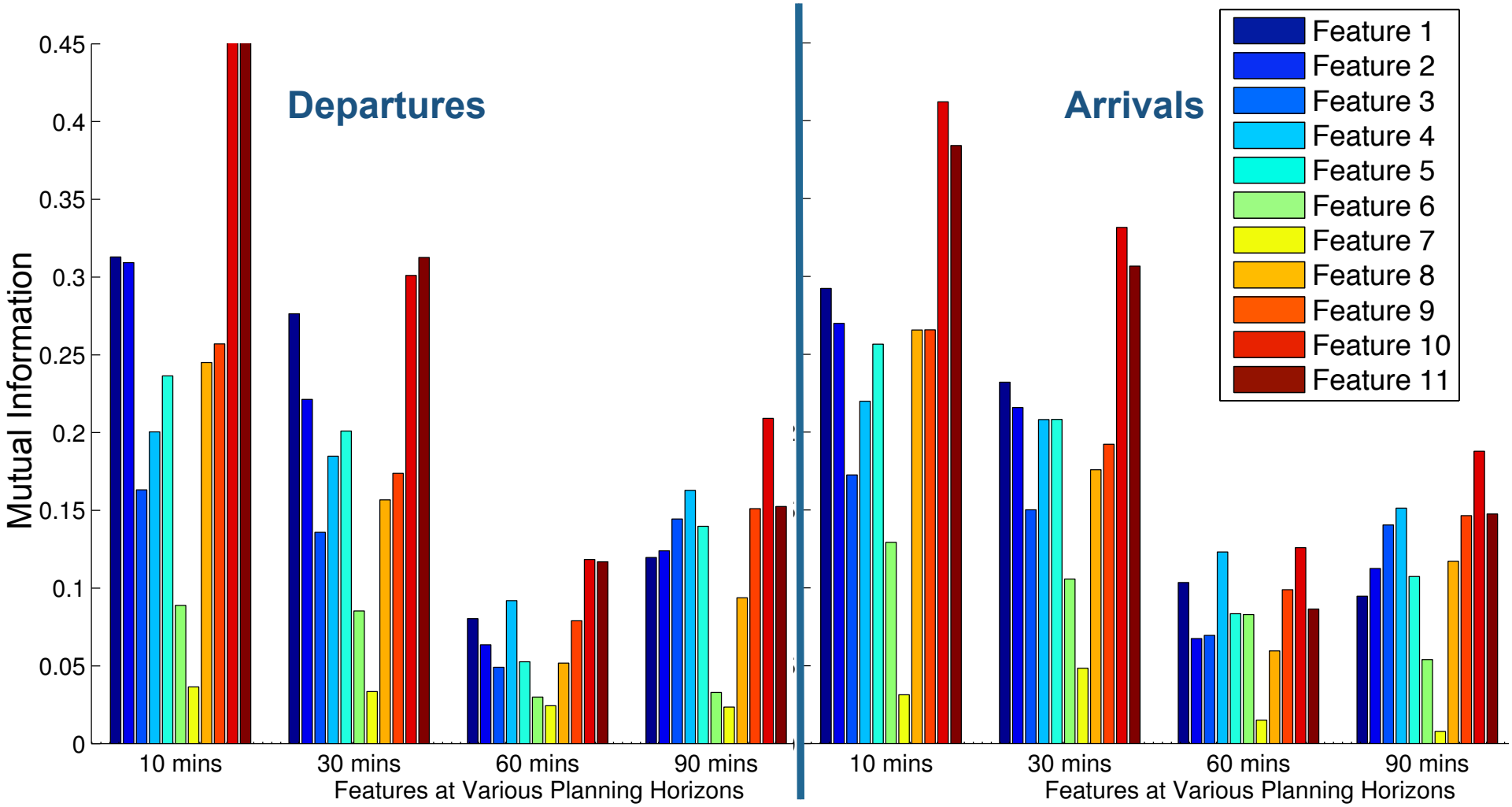


Feature selection: What factors are good predictors of route availability?

- Potential features:
  1. Mean VIL along path
  2. Standard Deviation of VIL along path
  3. Minimum distance to level 3+ weather along path
  4. Mean distance to level 3+ weather along path
  5. Maximum VIL in neighborhood of path
  6. Theoretical capacity for weather scenario
  7. Number of segments in the minimum cut
  8. Length of paths' minimum cut segment
  9. Minimal bottleneck
  10. Maximum density of level 3+ weather along path
  11. Maximum VIL density along path
- Evaluated using Mutual Information:

$$I[X;Y] = \sum_x \sum_y p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

Mutual Information values for features, compared across time horizons



Michalek and Balakrishnan, *Transportation Science* 2011

Step 1: Classification algorithm to predict blockage

- Incorporate feature values to predict and determine the likelihood of route blockage
- Standard two-class confusion matrix:

	Predicted Open	**Predicted Blocked**
Actual Open	True Positive (TP)	False Negative (FN)
Actual Blocked	False Positive (FP)	True Negative (TN)

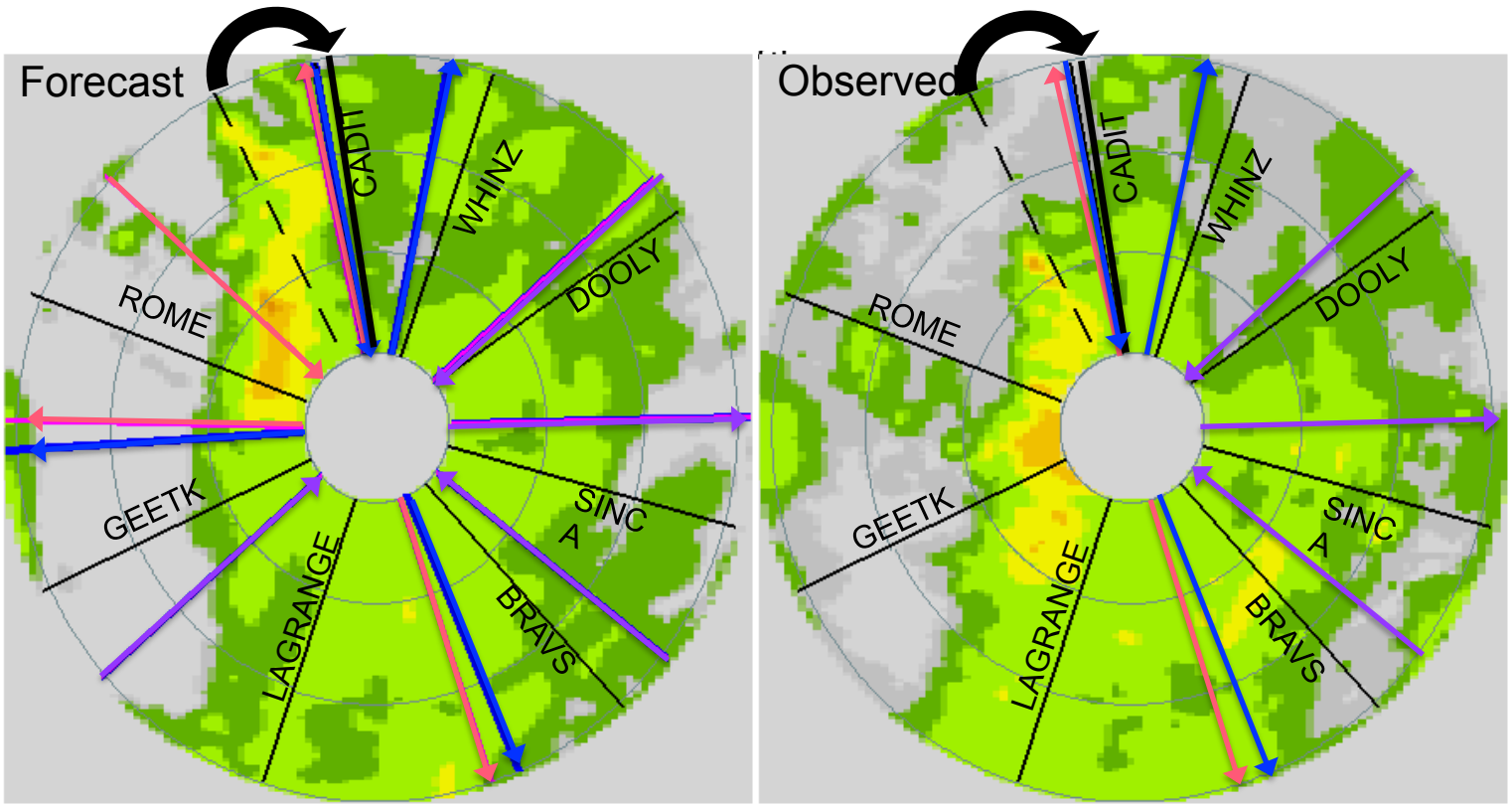
- For our application, we want to minimize scenarios in which we predict that the route is open, but in reality it is blocked (FP)
  - In other words, we seek to maximize the recall:

$$a^{-} = \frac{TN}{TN+FP}$$

  - Train an Ensemble of Support Vector Machines (EnsSVM)
  - Extend to probabilistic prediction of route blockage

Step 2: Optimization of airspace structure

Dynamic fix/route relocation, resectorization



Legend

- New Sector boundary
- Old sector boundary
- New Fix
- Old Fix

Fix	Probability	
	Old	New
WHINZ	0.975	0.975
CADIT	0.275	0.975
ROME	0.025	0.875
GEETK	0.825	0.975
LAGRA	0.975	0.975
BRAVS	0.825	0.925
SINCA	0.975	0.975
DOOLY	0.975	0.975

Michalek and Balakrishnan, IEEE Conference on Decision and Control 2010

Summary statistics of fix/route relocation

Empirical percentage of optimal fixes that are open given that they were assigned probability between 0.95 and 1

	Horizon (min)	% open | $p \in (0.95, 1.00]$	% open | $p \in (0.75, 0.95]$	% open | $p \in (0.50, 0.75]$
Arrivals	10	97.92	-	-
	30	94.38	-	-
	60	-	91.46	-
	90	96.30	88.90	-
	100	-	-	95.35
Departures	10	96.00	-	-
	30	97.53	-	-
	60	98.11	87.50	-
	90	91.05	88.46	-
	100	-	92.31	88.24

Blank entries correspond to cells with fewer than 10 data points (sample standard deviation 0.10)

Empirical percentages of open routes tend to stay within predictions

Michalek and Balakrishnan, *Transportation Science* 2011

Benefits of identifying robust routes

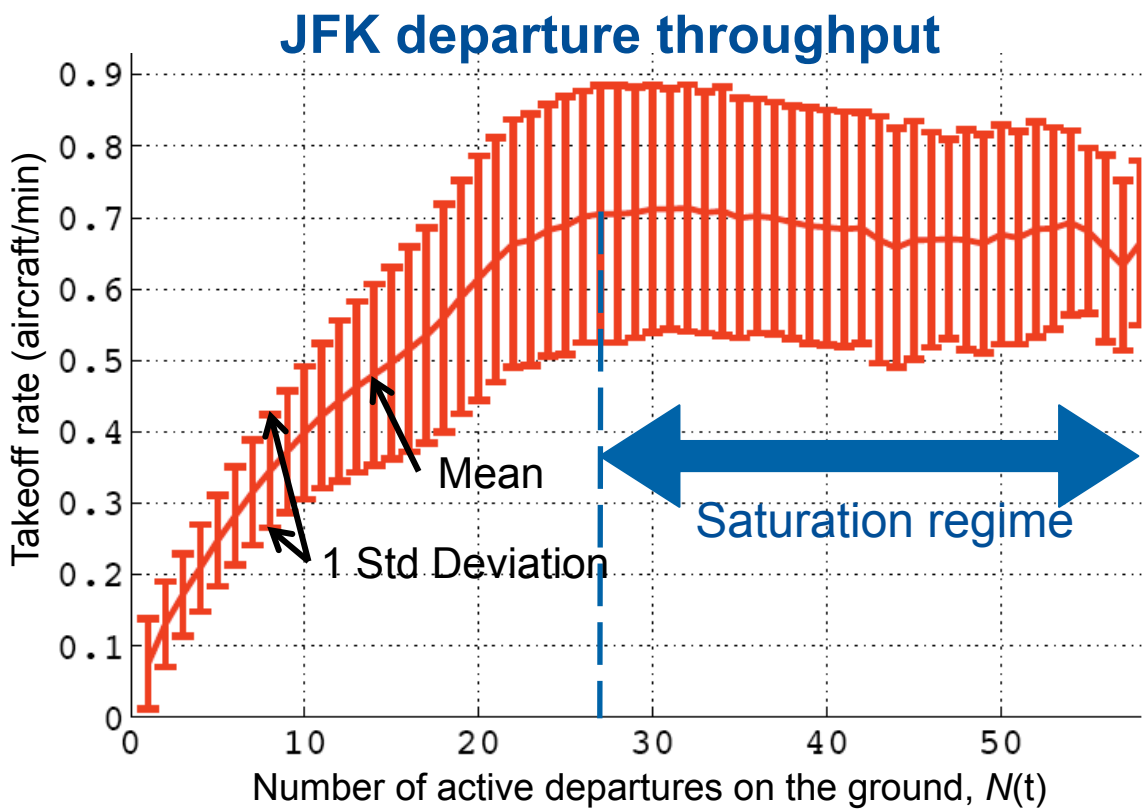
- Classifiers + optimization algorithms for determining optimal fix location, terminal-area routes and sectorization
  - Fix and associated route movements occur less than 40% of the time, sector movements less than 7% of the time
  - 13% more available routes initially forecast to be closed are opened up
  - Error rates are low (only 5% of modified routes are closed in actual weather)
- Can be extended to probabilistic forecasts of airspace capacity
- Promising approach for integration with ATFM algorithms under uncertain weather

Problem: Taxi-out fuel burn and emissions

- Airport capacity is a limiting constraint to air traffic operations
  - Major airports are already operating close to capacity
  - Terminal-area volume was responsible for 19% of delays at the top 35 airports in 2009
- Congestion leads to increased taxi times, fuel burn and emissions
- Annually, at major airports in the United States
  - Over 32M minutes taxi-out delays (over unimpeded times)
  - Over 13M minutes taxi-in delays (over unimpeded times)
  - 600M gallons of jet fuel expended in taxi-out process
- Taxiing aircraft contribute to noise and surface emissions, e.g. CO_2, NOx, SOx, CO, HC, and Particulate Matter
  - 6M metric tons of CO_2 emissions due to taxi-out processes

Main culprit: Surface congestion

- Major airports are frequently severely congested, resulting in large taxi-out delays and inefficient operations



When JFK is saturated:

In Visual Meteorological Conditions, the average taxi-out time is **56 min**

In Instrument Met. Conditions, the average taxi-out time is **69 min**

Unimpeded taxi-out time: **16-19 min**

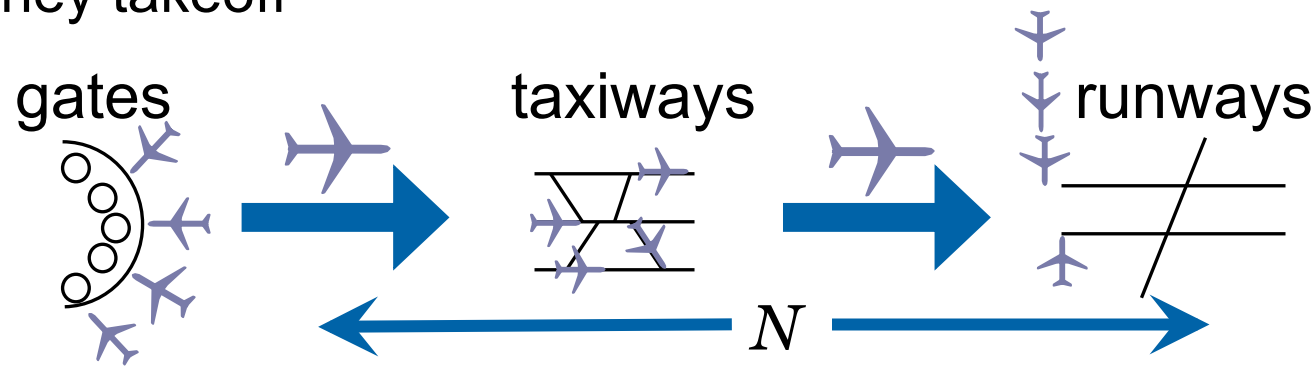
JFK is saturated 18% of the time in VMC, and 24% of the time in IMC

32% of departures at JFK takeoff during saturated periods

Simaiakis and Balakrishnan, *Transportation Research Record: Journal of the Transportation Research Board*, 2010
(Confirms Pujet, Delcaire and Feron, BOS 1999).

Our solution: Pushback Rate Control

- Aircraft pushback from gates, start their engines, and then taxi until they takeoff

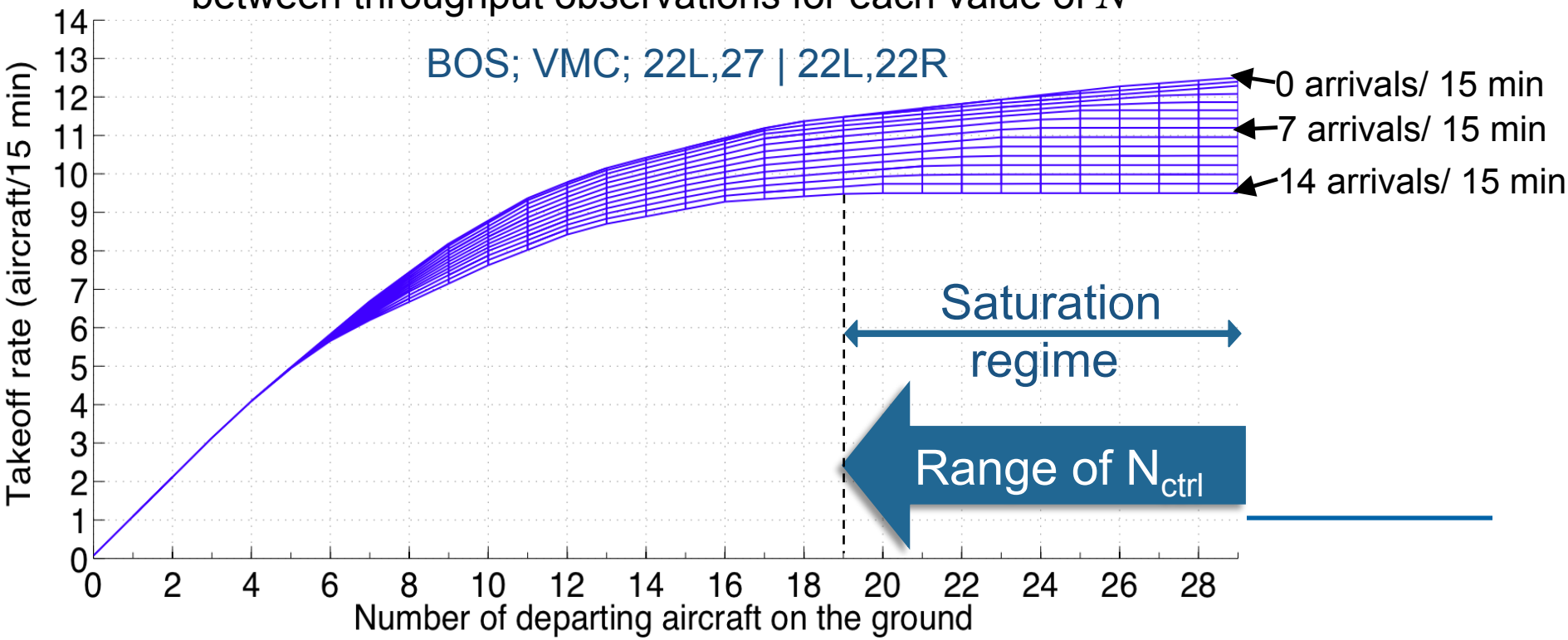


- Want to control pushbacks to keep number of departures on the surface (N) close to target value, N_{ctrl}
- Challenges:
  - How do we choose N_{ctrl}?
  - How do we design and implement the control strategy?
  - How do we interface with the human controllers?

Simaiakis et al., USA/Europe Air Traffic Management R&D Seminar 2011.

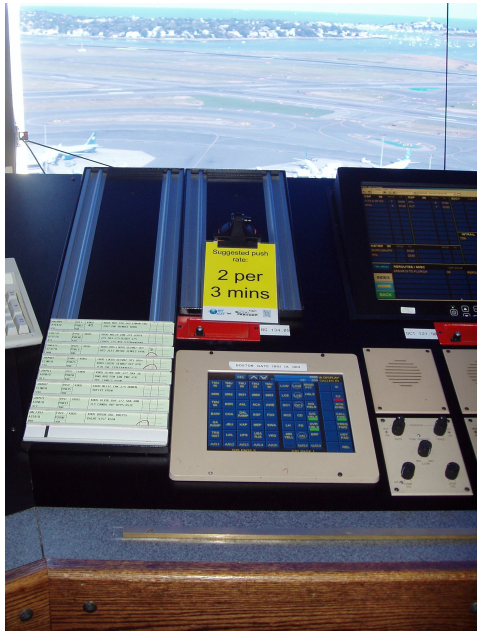
Step 1: Choosing N_{ctrl}

- Data set: Historical pushback and takeoff times (departures), airport configuration, weather conditions, landing and gate-in times (arrivals)
- Estimation of range of N_{ctrl}
  - Convex optimization formulation for estimating departure throughput curves parameterized by number of arrivals
  - Estimate saturation value (N^*) by testing for significant differences between throughput observations for each value of N



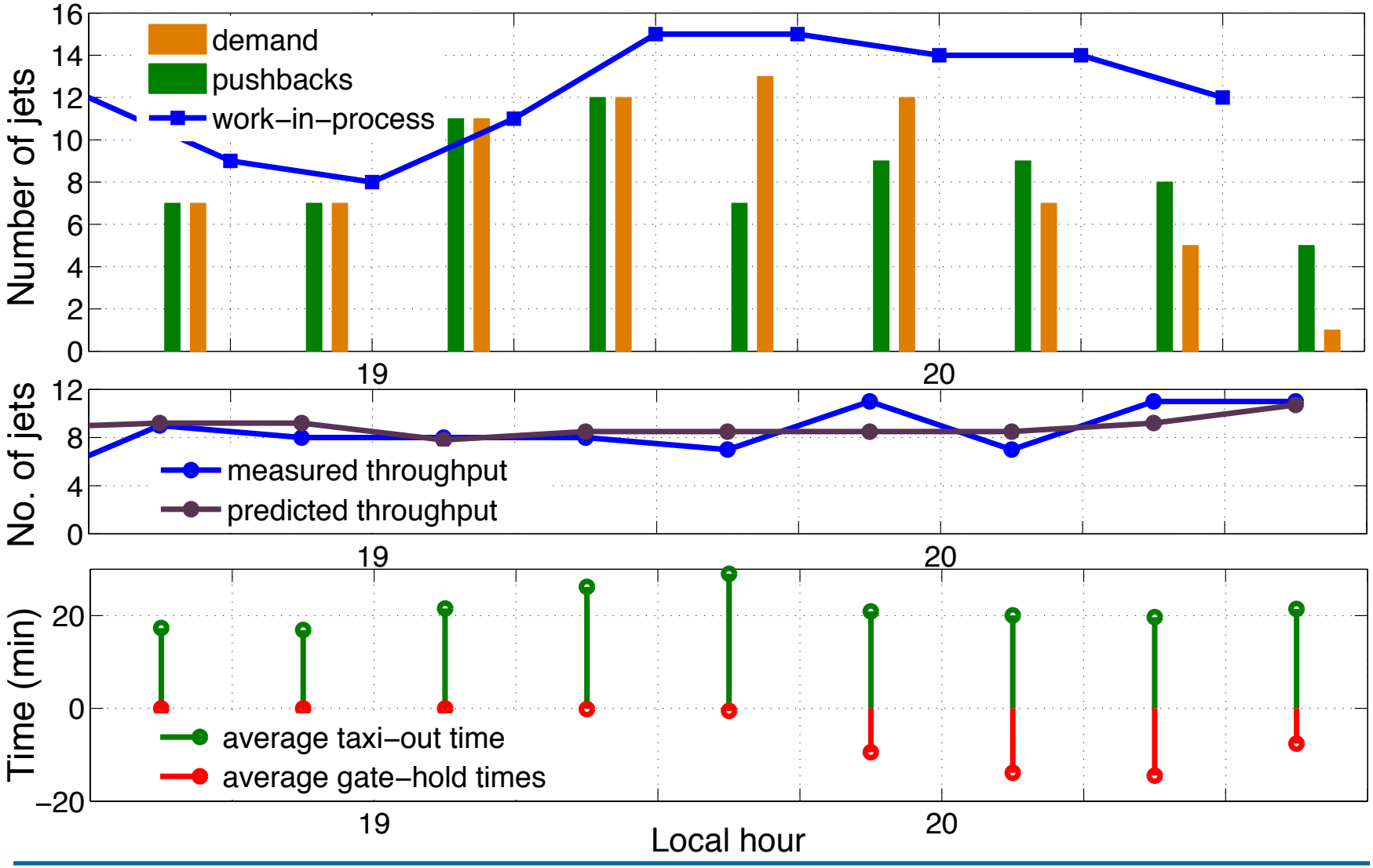
Step 2: Designing and implementing control strategy

- On-off control does not work
  - Rather than release an aircraft every time that a flight takes off, controllers prefer a rate at which to let aircraft pushback from gates
  - Rate is updated periodically
- Need to predict departure throughput in order to determine recommended pushback rate
  - Regression tree to predict departure throughput
  - Determine optimal rate using dynamic programming
- Field-tests at Boston Logan airport (BOS)
  - Communicate recommended pushback rate using color-coded cards
    - No verbal communications with controllers



Simaiakis et al., USA/Europe Air Traffic Management R&D Seminar 2011.

Sample test results: 07/21/2011



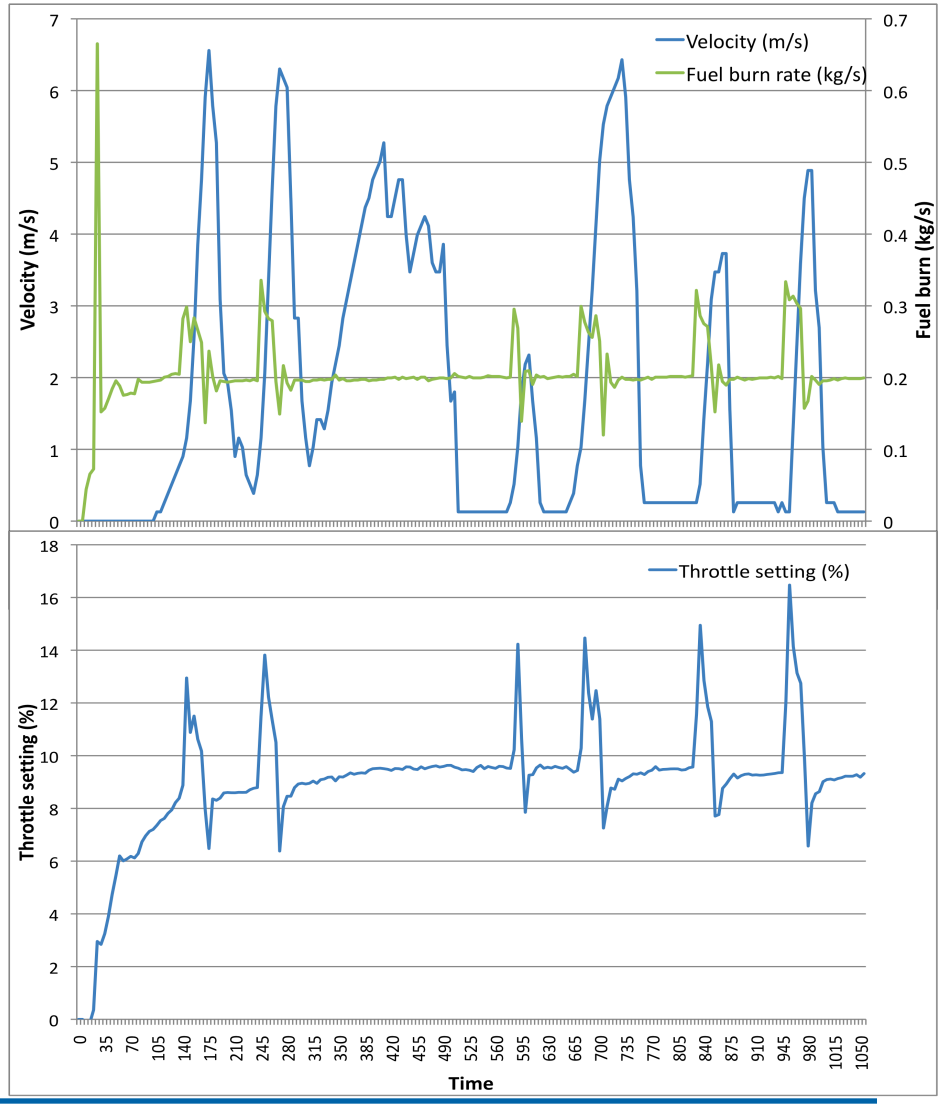
Simaiakis and Balakrishnan, submitted, 2011.

Data visualization (7/21/2011)

Google Earth visualization

Estimating fuel burn impact of taxi trajectories

- ICAO provides baseline fuel flow rates for aircraft taxi
  - Assume 7% constant thrust
  - Measured on static test-beds
  - Neglect pilot behavior
  - Operational fleet may have different baseline fuel flow rates and thrust settings
- Flight Data Recorder archives used to build fuel burn models
  - Acceleration events prove to be significant
- Surface surveillance data (ASDE-X) used to find relation between congestion and acceleration events



Khadilkar and Balakrishnan, AIAA Guidance Navigation and Control Conference, 2011.

Estimated benefits of pushback rate control at BOS

- Significant benefits shown across two phases of testing
  - Average 4.7 min decrease in taxi-out times of gate-held flights
  - 23-28 tons (6,800-8,000 US gal) reduction in fuel burn across 14 congested evening periods of 3 hours each
  - 57-67 kg decrease / gate-held flight
- Control strategy shown to be capable of maintaining runway utilization
- Promising stakeholder feedback
  - Traffic managers in tower noted improved surface “flows”
  - Anecdotal positive feedback from pilots, controllers and airlines

Summary

- Possible to enhance performance of air traffic system by leveraging large amounts of operational data
  - Weather forecasts and observations for robust routing
  - On, Off, Out, In data (FAA, ASPM) and surface surveillance data for predicting taxi-out times and airport performance, and for the optimal control of departure processes
  - Flight Data Recorder archives to estimate fuel burn impacts
- Some other applications
  - Estimation of maximum-likelihood discrete-choice models of the runway configuration selection process
  - Evaluation of infrastructure investments (e.g., new runways)
  - Network congestion control models and algorithms for departure management

Ramanujam and Balakrishnan, American Control Conference, 2011
Simaiakis and Balakrishnan, INFORMS 2011
Khadilkar and Balakrishnan, submitted, 2012