Service Level Expectation Setting for Air Traffic Flow Management: Practical Challenges and Benefits Assessment

by

Michael Ball

Robert H Smith School of Business & Institute for Systems Research

U of Maryland

Joint work with: Cindy Barnhart, Mark Hansen, Lei Kang, Yi Liu, Prem Swaroop, Vikrant Vaze, Chiwei Yan

23 June 2017

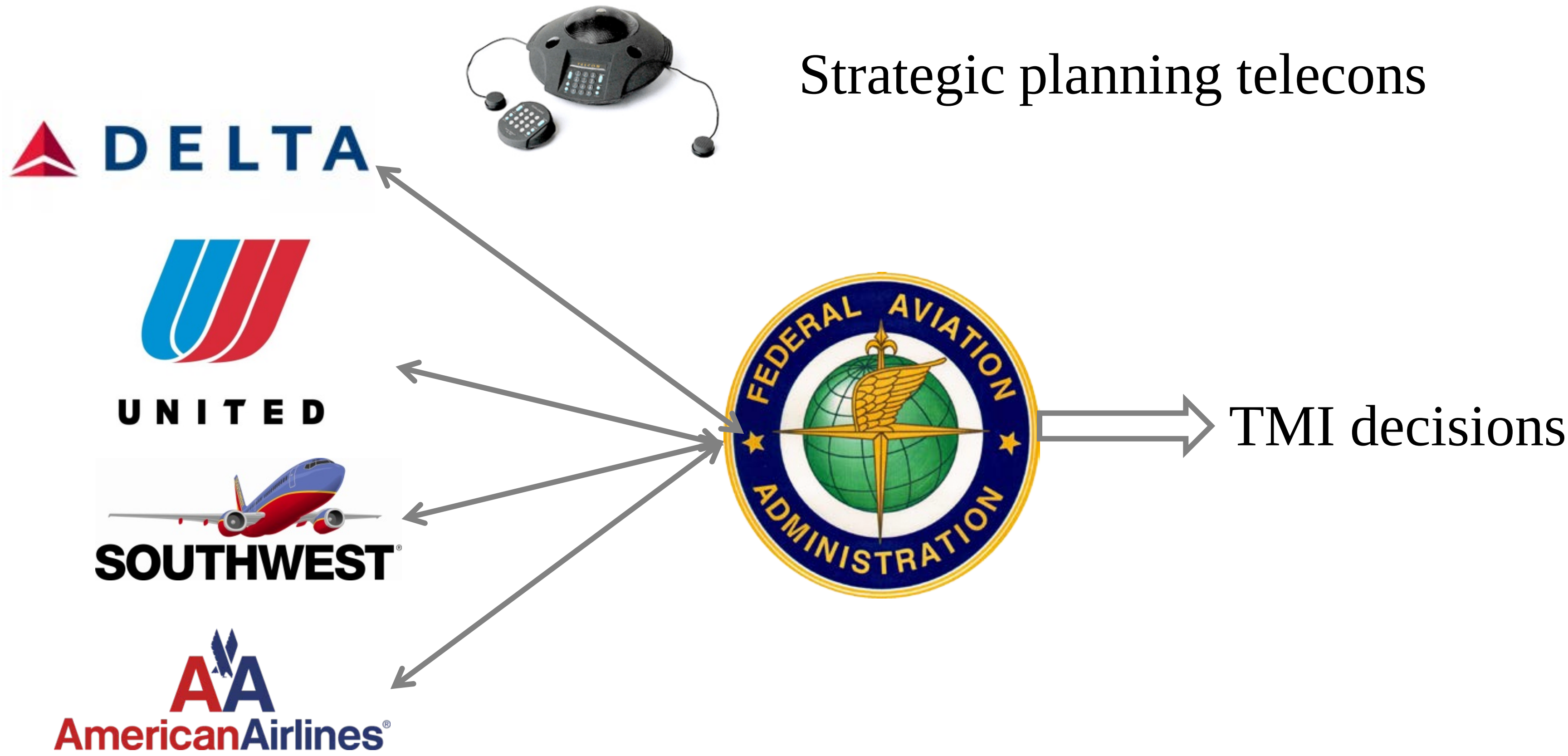
Outline

- Motivation and Project Overview
- System Description
- Benefits Assessment Using Fast-time Simulation
- Summary of Conclusions from Human-In-The-Loop (HITL) Experiment and Associated Survey

Outline

- **Motivation and Project Overview**
- System Description
- Benefits Assessment Using Fast-time Simulation
- Summary of Conclusions from Human-In-The-Loop (HITL) Experiment and Associated Survey

Current Practice on TMI Planning



Operational Challenge

- Flight operators participate in strategic TMI planning by verbal input. Operators can sometimes have a disproportionate influence on decisions that affect a broad range of others who are less vocal.
- Discussion focuses on specific parameters rather than performance goals.
- Different traffic managers may create different plans for the same situation.
- The planning process is ad-hoc and subjective.

A NextGen Vision: Performance-Based ATM

Current Practice:

NextGen Vision:

Philosophy:

- Airlines provide “consensus” service expectations
- FAA develops operational plan to meet those expectations

SLE Concept

- The SLE mechanism allows operators to submit quantitative input that represent their preferred system performance goals (capacity, predictability and efficiency).
- It then appropriately weighs and aggregates operators' inputs to determine consensus performance goals.

Underlying models, analysis and mechanisms are results of SLE project.

- These goals can then used to determine TMI parameters that are expected to best achieve the performance expectations.

"Step 2": requires additional research – performance based TMI planning.

Outline

- Motivation and Project Overview
- **System Description**
- Benefits Assessment Using Fast-time Simulation
- Summary of Conclusions from Human-In-The-Loop (HITL) Experiment and Associated Survey

COuNSEL: CONsensus Service Expectation Level Planning

Information to users:
candidate performance vectors

	Capacity	Efficiency	Predictability
V1:	0.9	0.8	0.5
V2:	0.7	0.7	0.9
V3:	0.8	0.6	0.8

Inputs from user 1:
Grades for vectors
and candidate vectors

Grades:

100%, 95%, 90%, 85% …

Consensus vector:
e.g. (.89 , .76 , .65)

Consensus Vector Chosen using *Majority Judgment (MJ)* see Balinski & Laraki

- Suppose:
  - 6 airlines (voters), voting on 3 candidates: V_1, V_2, V_3
  - grades: 100%, 99%, 98%, 97%, 96%, 95%, 94%, …
- Grades sorted after voting from worst to best:

V_1	80%	80%	90%	94%	95%	100%
V_2	75%	83%	85%	87%	88%	90%
V_3	65%	70%	88%	90%	93%	95%

Majority grades: majority would give at least that grade.
…. in this example 4th grade from right.
Vector with highest majority grade will be selected.
There is a tie-breaking rule – not discussed here.

- There is a body of research on MJ that generally indicates its superiority as a voting / grading mechanism.
- It is used to grade a variety of competitions and for some elections in Europe.

Performance Goals in SLE

- **Capacity**: maximize throughput
  - Avoid underestimating capacity and encourage quick response if weather clears early
- **Efficiency**: minimize delay cost
  - Take delay on the ground instead of in the air
- **Predictability**: provide timely, accurate, information
  - Announce GDPs well ahead of start times
  - Avoid overestimating or underestimating capacity; make program revisions unlikely

Interpretation of Performance Goals

All metrics take on values between 0 and 1

1 → perfect performance

0 → worst possible performance

The system only allows goal vectors that are “feasible”, e.g. even on a near-perfect day (1,1,1) would not be possible – perfect performance across all dimensions.

The system forces the flight operators to make tradeoffs:

(.91, .83, .85) → (.86, .89, .85)

***Reduce capacity goal*: .91 → .86**

***… in order to improve efficiency goal:* .83 → .89**

Design Tradeoffs

- SLE will enable flight operators to influence TMI design tradeoffs
- **Predictability vs. Throughput**
  - Predictability— assume lower planned rates and long planned duration so that initially assigned delays are unlikely to be extended
  - Throughput—assume higher planned rates and shorter planned duration in order to increase demand pressure
- **Efficiency vs. Throughput**
  - Efficiency—minimize airborne delay by imposing more ground delay
  - Throughput—employ higher planned arrival rates to increase demand pressure but (possibly) at the expense of more airborne delay
- **Predictability vs. Efficiency**
  - Predictability—make decisions well in advance, even though this increases the risk that they will be based on erroneous forecasts
  - Efficiency—make decisions later when better information is available, reducing the risk of airborne delay

Challenge in Application of Majority Judgment to Service Level Expectation Setting

- The basic application of MJ allows flight operators to make a consensus choice among possible goal vectors.
- **Challenge 1:** given conditions on a particular day of operations what are appropriate “possible goal vectors” that should be presented to flight operators.
  - Partial Answer: In concept there will be many (an infinite number) of vectors that represent the possible tradeoffs among the performance vectors given the weather and traffic conditions for the scenario of interest. Thus, challenge 1 becomes the problem of representing the space of performance metric tradeoffs for the TMIs under consideration.
- **Challenge 2:** given some representation of the (very large) space of possible goal vectors, what is a process for choosing among these the ones that flight operators will grade as part of the MJ process?

Solution to Challenge 1: *Set of constraints that define feasible vectors for particular day in the NAS.*

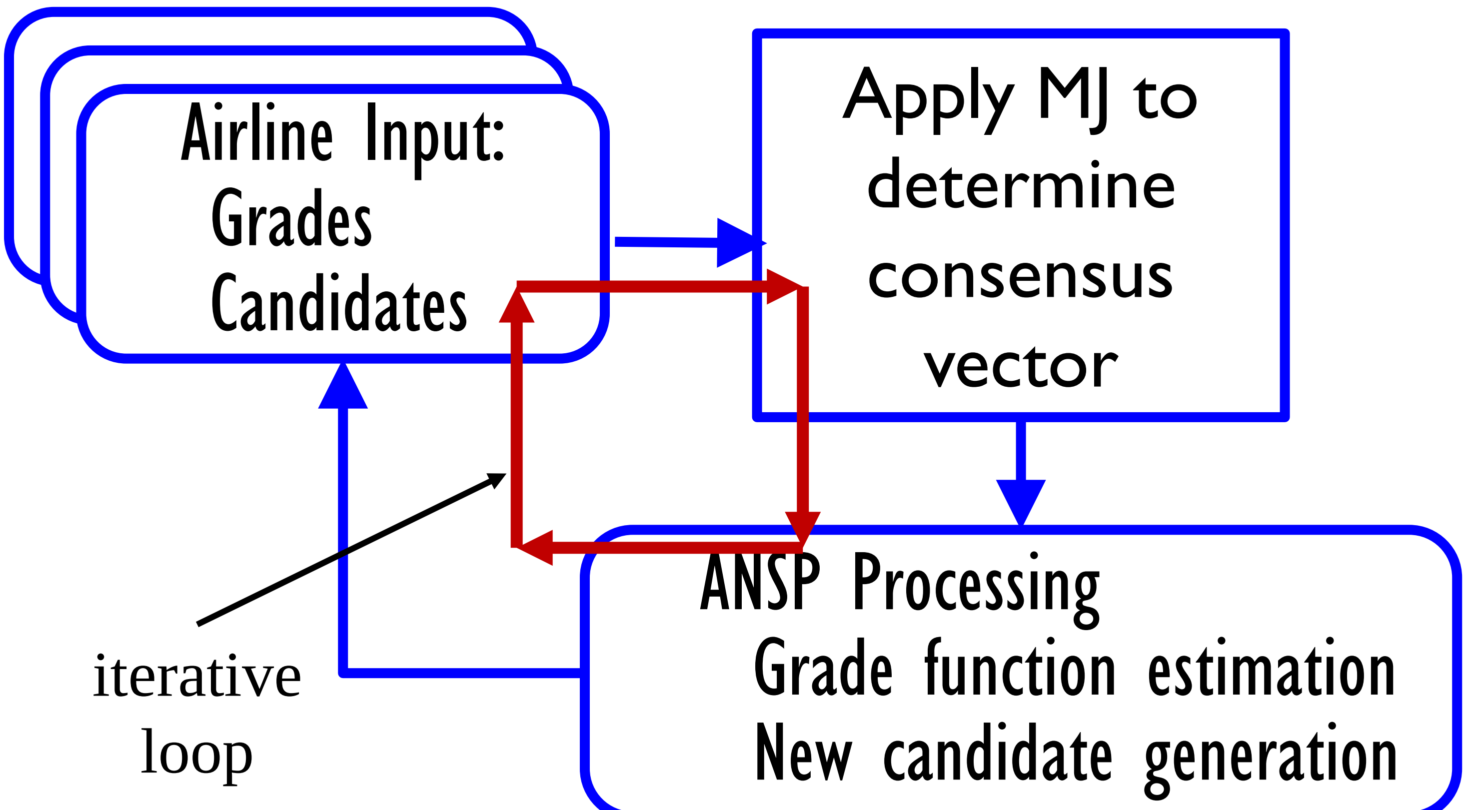
Bad weather day – sample vectors: (.90, .75, .80), (.85, .80, .83), (.85, .90, .79).

Good weather day – sample vectors: (.98, .95, .90), (.99, .92, .91), (.95, .97, .90).

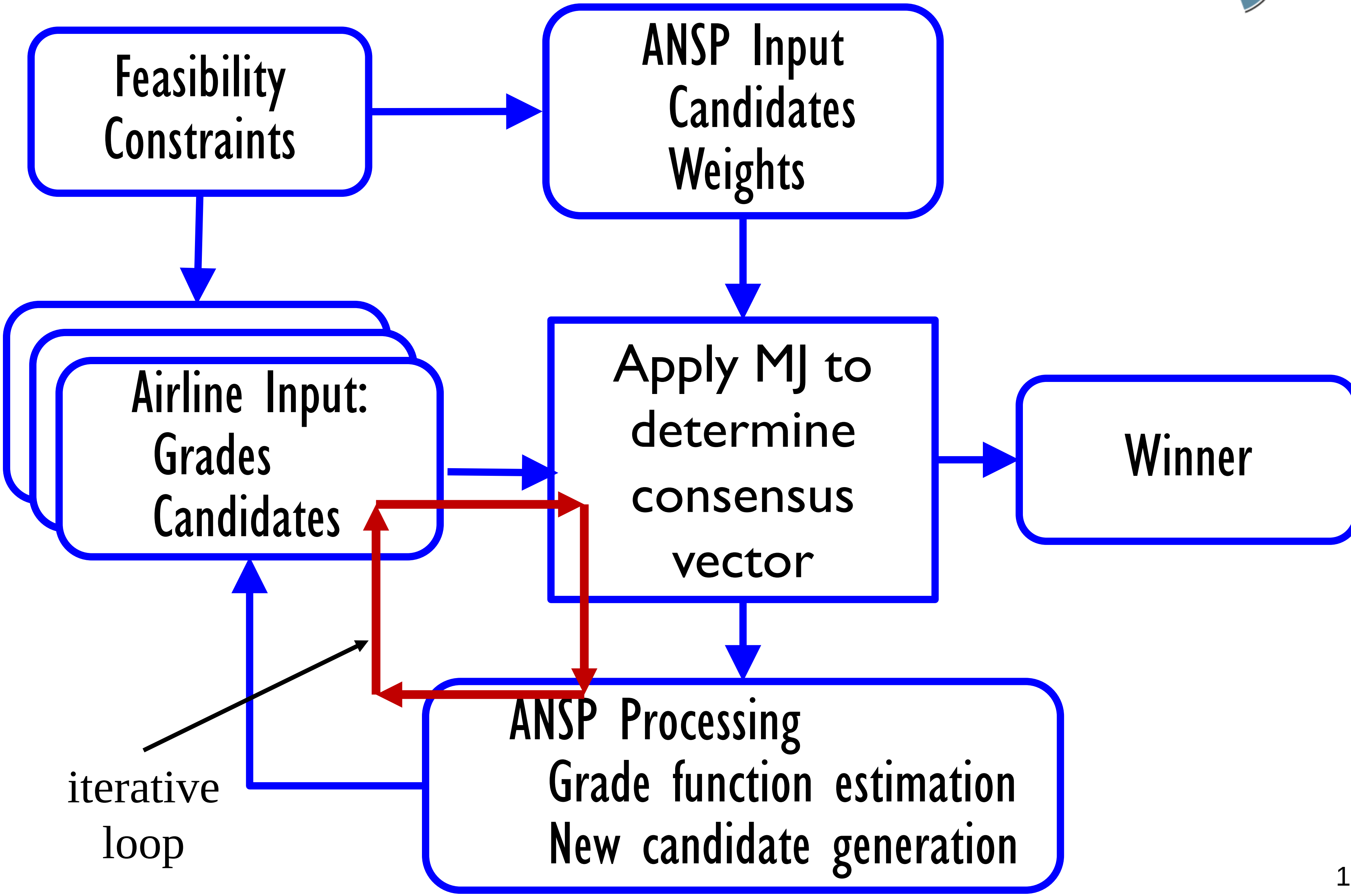
$\mathbf{m}$ is possible metric vector :

$$\mathbf{m} \in FEAS_{METRIC}$$

Solution to Challenge 2: Iteratively generate candidate vectors, applying MJ a small number of times on small sets of candidates



COuNSEL Logic Flow



SLE Particulars

- Airline votes are weighted by number of flights involved in the TMI
- Airlines may develop their own tools to assess how different candidate vectors affect their individual business objectives
- Multiple applications of COuNSEL might be used as conditions change; could be applied nationwide or to regional problem area

Significant Research Components

- **Generating candidate vectors, COuNSEL iteration mechanism:** must generate promising candidates for infinite space of possible vectors – employs optimization and statistical estimation models.
- **Definition of space of feasible candidate vectors:** analytic models of TMIs – relationship between parameter setting and performance metrics.
- **Understanding user impact and benefit mechanisms, gaining user acceptance:** outreach to flight operators; formal flight operator surveys; human-in-the-loop simulation, involving flight operators and FAA.
- **Modeling benefit mechanism and flight operator impact:** use of historical data analysis and simulation to relate flight operator performance to TMI parameter settings.
- **Modeling user voting/grading behavior:** game theory and related models to understand user payoff functions and incentives for good (and bad) voting behavior.

Outline

- Motivation and Project Overview
- System Description
- **Benefits Assessment Using Fast-time Simulation**
- Summary of Conclusions from Human-In-The-Loop (HITL) Experiment and Associated Survey

What Are the Types of Benefits?

- Reduction in time spent in unstructured Strategic Planning Teleconns (SPTs)
- Reduction in NAS-wide operating costs
- More equitable and consensus-driven TMI planning:
  - Improved performance in terms of airlines' business objectives / costs
  - Greater balance in performance across airlines

Compare to What?

Performance achieved by GDP plan produced by COuNSEL

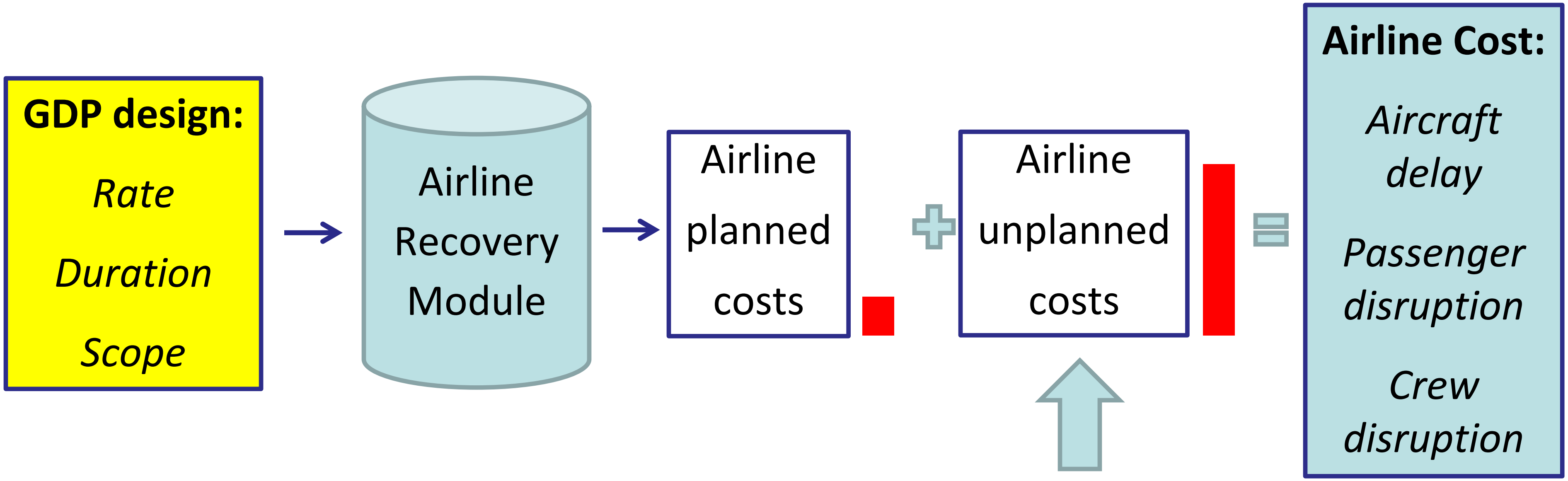
vs

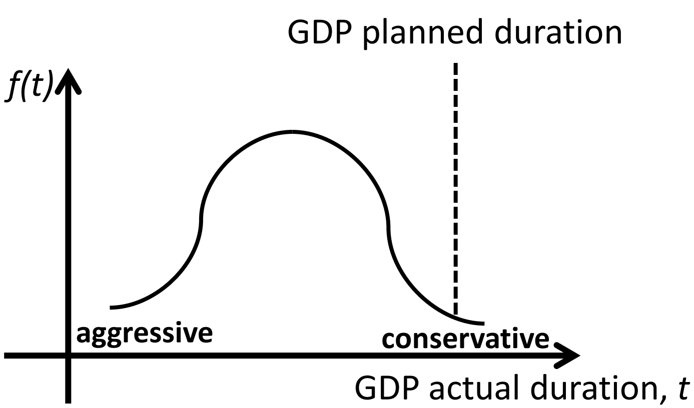
Performance achieved by GDP plan based on centralized optimization (e.g. using Richetta-Odoni 1993, Mukherjee-Hansen 2007, etc.).

How is performance measured?

Total operational costs of all airlines (simulated) with cost dispersion metrics measuring equity.

How to Assess?





Capacity is uncertain…

GDP rate may be under/over-estimated

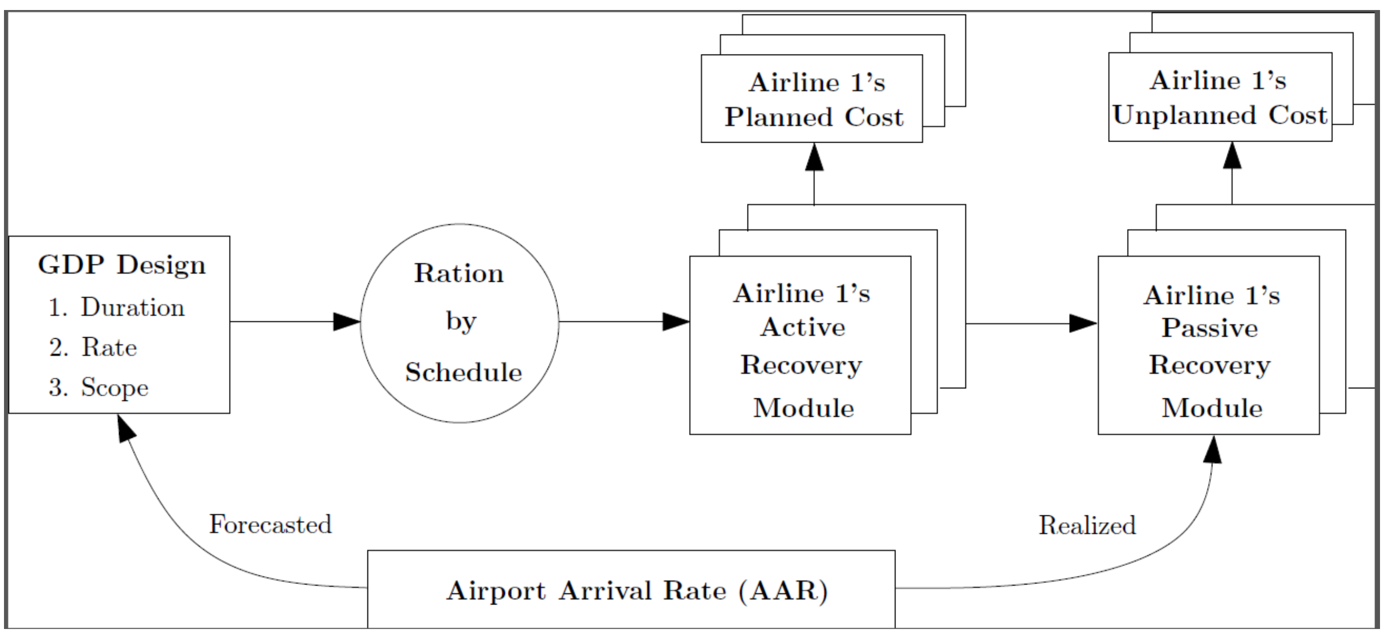
GDP duration may be too long/short

…leads to early cancellation and late extension

Unplanned cost: Additional airborne delay, passengers disruptions, fleet disruptions due to inaccurate delay information provided by FAA

How to Assess?

- Integrated FAA-Airline interaction simulator
  - 1) GDP design module, 2) FAA RBS module, 3) airline recovery module, 4) airline/NAS performance evaluation module
  - **Functionality:** given a GDP design, evaluate airline and NAS performance under capacity uncertainty



Experimental Setup

- **mm/dd/yy:** 6/16/2007
- **Airport:** SFO
- **Actual duration:** ~Uniform[3 hr, 9 hr]
- **14 candidate designs for evaluation:** planned duration: 3-9 hours, with an increment of 0.5 hours.
- **Program arrival rate:** outside of GDP duration: VFR rate; during GDP duration: IFR rate
- **Airline itinerary data source:** Generated by Barnhart et al., 2011.

- **Delay cost coefficient estimation:** BTS Form 41 financial data. Estimated separately for different airlines and different fleet types.
- **Carriers involved:**

	# Impacted Operations	# Fleet Types (# Aircraft in Each Category)	# Impacted Passengers	% Connecting Passengers
United & SkyWest	359	10 (17,4,8,3,9,1,3, 7,5,27)	24236	32.33%
American & American Eagle	70	5 (4,2,4,3,9)	7678	27.39%
US Airways	40	4 (1,4,1,4)	4007	31.57%
Continental & ExpressJet	30	5 (1,1,3,1,2)	3244	20.43%
Delta Airlines	26	4 (1,1,2,2)	3750	30.29%
Alaska Airlines	25	2 (4,3)	2461	9.47%
Northwest Airlines	23	4 (2,2,2,1)	3232	25.46%
Frontier Airlines	15	2 (2,2)	1351	31.68%
JetBlue Airways	9	1 (2)	1180	8.05%
AirTran Airways	8	1 (4)	973	32.58%

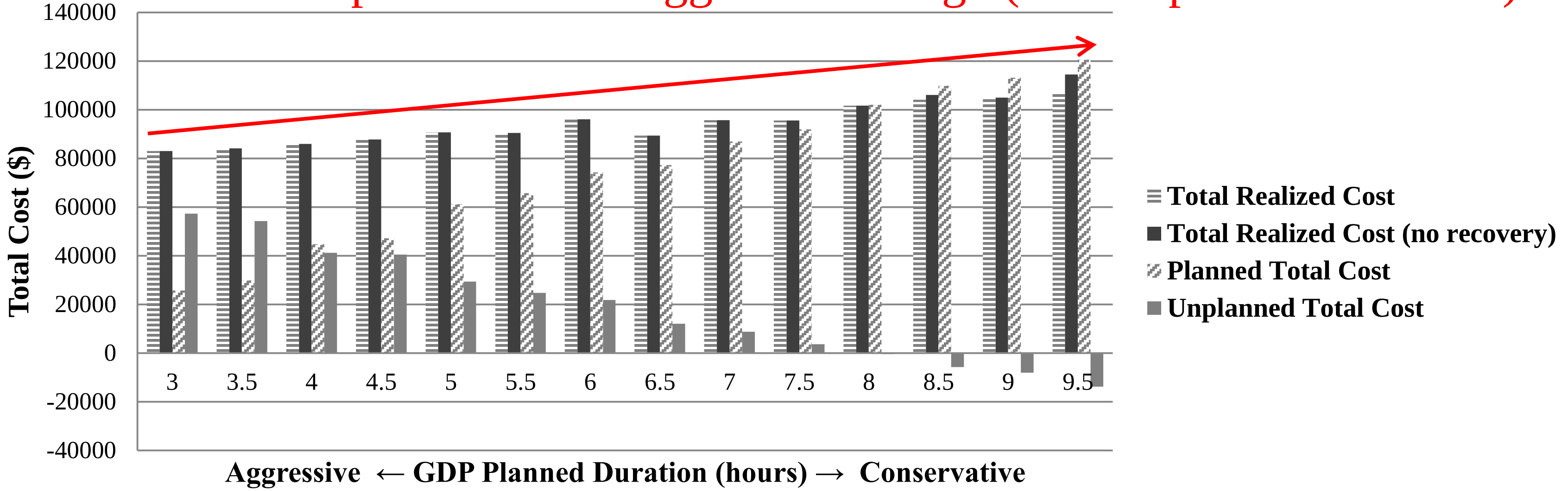
Revealing Airline's Preferences

Airline	# Impacted Operations	# Fleet Types (# Aircraft in Each Category)	# Impacted Passengers	% Connecting Passengers	Average Load Factor
US Airways	40	4 (1,4,1,4)	4007	31.57%	80.43%

small number of total operations, multiple different fleet types

little flexibility for recovery
(reduces 6.6% cost through recovery, at most)

total cost trend: preference for aggressive design (shorter planned duration)



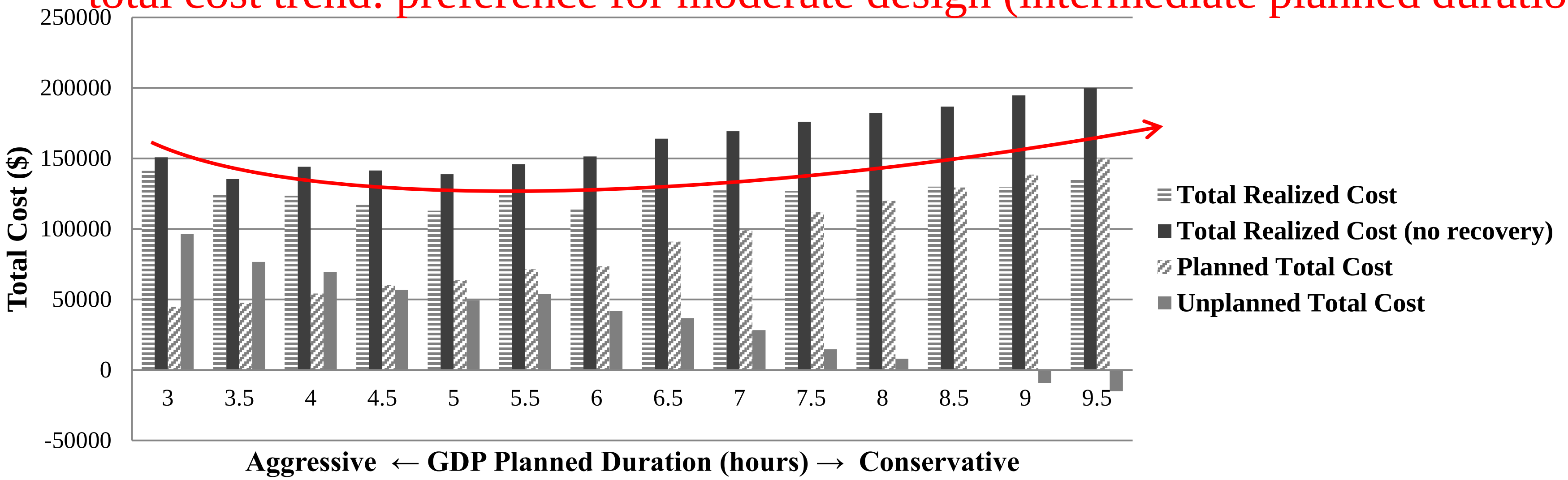
Revealing Airline's Preference

Airline	# Impacted Operations	# Fleet Types (# Aircraft in Each Category)	# Impacted Passengers	% Connecting Passengers	Average Load Factor
American & American Eagle	70	5 (4,2,4,3,9)	7678	27.39%	75.53%

medium number of total operations, multiple different fleet types

medium flexibility for recovery
(reduces 32.4% cost through recovery, at most)

total cost trend: preference for moderate design (intermediate planned duration)



Revealing Airline's Preference

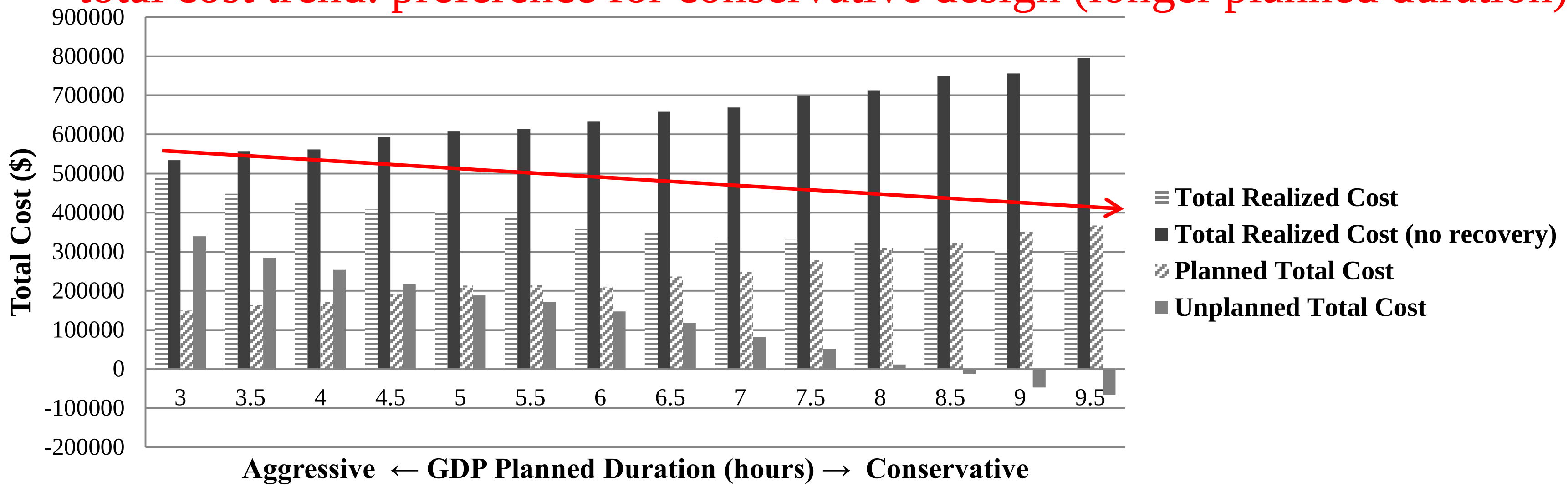
Airline	# Impacted Operations	# Fleet Types (# Aircraft in Each Category)	# Impacted Passengers	% Connecting Passengers	Average Load Factor
United & SkyWest	359	10 (17,4,8,3,9,1,3,7,5,27)	24236	32.33%	75.29%

extremely large number of total operations, multiple different fleet types

⇩

great flexibility for recovery
(reduces 62.3% cost through recovery, at most)

total cost trend: preference for conservative design (longer planned duration)



NAS-wide Benefits Assessment

Preference Category	Airline - GDP Cost Matrix	Aggressive Design ← GDP Planned Duration (hours) → Conservative Design													
		3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5
Moderate	American & American Eagle	141340	124389	123490	117142	***112998***	125420	115407	128040	127462	126762	128174	130014	129585	134946
Aggressive	Frontier	***60362***	76148	66946	81898	82396	83363	85580	88988	91783	89850	101507	94825	106871	105891
Aggressive	US Airways	***83058***	84186	85994	87735	90695	90418	96115	89400	95663	95637	101711	104089	105107	106905
Aggressive	Continental & ExpressJet	34152	37247	37844	***33511***	36526	33968	39176	37459	39935	40162	41300	43174	44005	47296
Moderate	JetBlue	9705	9849	10766	8939	8252	7983	***7577***	8367	7707	8563	9446	10863	13090	15468
Moderate	Delta	36256	35408	34897	34846	34860	***34132***	35880	34732	35531	35773	38467	39139	41918	43874
Conservative	AirTran	16600	15049	15050	13499	13363	11954	11280	11651	10338	10645	***9592***	10268	9864	12001
Aggressive	Northwest	***22247***	36705	32657	31738	31265	34185	34704	32411	36074	36831	36690	40855	40764	40228
Conservative	United & SkyWest	489250	448340	426198	408230	402122	386515	357885	354516	330232	330824	322038	309187	304852	***300218***
Moderate	Alaska	41167	35758	35713	***32724***	35337	37002	34810	36539	34573	36305	36882	37731	38215	38301
Moderate	NAS wide	934137	903079	869554	850262	847815	844941	818413	822104	***809297***	811352	825808	820144	834271	845128
Conservative	Centralized Objective	244986	235343	226604	221638	214614	210056	202624	201951	196292	189613	***188450***	195662	209204	220389

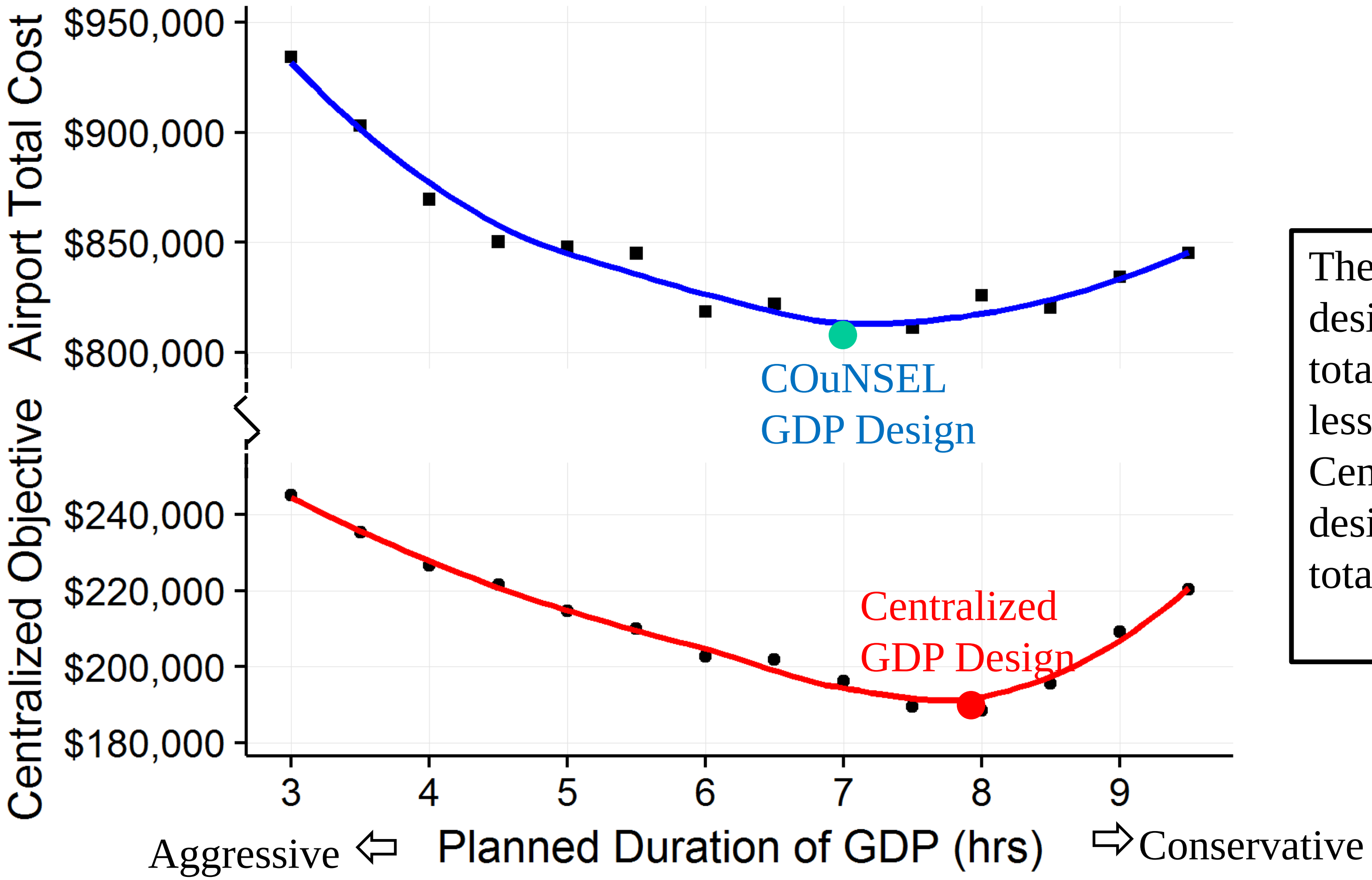
NAS-wide Benefits Assessment

Airline - GDP Grade Matrix	# impacted operation	weights	Aggressive Design ← GDP Planned Duration (hours) → Conservative Design													
			3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5
American & American Eagle	70	17.97	80	91	92	96	100	90	98	88	89	89	88	***87***	***87***	***84***
Frontier	15	6.30	100	79	90	74	73	72	71	68	66	67	59	64	56	57
US Airways	40	12.28	100	99	97	95	92	92	86	93	87	87	82	80	79	78
Continental & ExpressJet	30	10.1	98	90	89	100	92	99	86	89	84	83	81	78	76	71
JetBlue	9	4.45	78	***77***	***70***	85	92	95	100	91	98	88	80	70	58	49
Delta	26	9.16	94	96	98	98	98	100	95	98	96	95	***89***	87	81	78
AirTran	8	4.11	58	64	64	71	72	80	85	82	93	***90***	100	93	97	80
Northwest	23	8.43	100	61	68	70	71	65	64	69	62	60	61	54	55	55
United & SkyWest	359	54.63	61	67	70	***74***	***75***	***78***	***84***	***85***	***91***	91	93	97	98	100
Alaska	25	8.92	***79***	92	92	100	93	88	94	90	95	90	89	87	86	85

Majority judgment winner coincides with system-optimal design!!

NAS-wide Benefits Assessment

COuNSEL design vs centralized design



The COuNSEL design's airport total cost is 2% less than the Centralized design's airport total cost.

Measuring Equity: deviation of each airline from their preferred design

Airlines	*Centralized Design*	*COuNSEL Design*
US Airways	22.46%	15.18%
Frontier	68.16%	52.05%
Northwest	64.92%	62.15%
Continental & ExpressJet	23.24%	19.17%
Delta	12.70%	4.10%
American & American Eagle	13.43%	12.80%
Alaska	12.71%	5.65%
JetBlue	24.67%	1.72%
United & SkyWest	7.27%	10.00%
AirTran	0.00%	7.78%
Mean	24.96%	19.06%
Standard Deviation	22.00%	19.78%

Outline

- Motivation and Project Overview
- System Description
- Benefits Assessment Using Fast-time Simulation
- **Summary of Conclusions from Human-In-The-Loop (HITL) Experiment and Associated Survey**

HITL Description

- One day event held at U of Maryland
- Representatives from:
  - American Airlines, Delta Airlines, United Airlines, UPS.
  - FAA Command Center and HQ.
- Employed prototype COuNSEL software tool
- Three GDP scenarios: 2 from EWR, 1 from SFO

Principal advantages identified in HITL and survey:

- Allows each flight operator to specify a baseline both for the FAA and the operator in terms of performance expectations.
- Makes the implicit and sometimes subjectively applied performance tradeoffs in TMI planning explicit.
- Brings forth the importance of predictability as a performance goal in TMI design.
- Provides for structured operator input and enables a more systematic and consistent decision-making process.
- Includes inputs from all flight operators.

Topics for Further Investigation Identified in HITL and Survey

1) Broader role for COuNSEL should be considered:

- COuNSEL was developed to fill a specific role within the NextGen architecture ➔ it requires new models to convert its output into a TMI plan.
- However, it can be seen as a general approach to obtain consensus advice from the flight operators ➔ other potential applications.
- ***It could directly output TMI parameters or could output performance goals that are input into FAA subjective decision making; some potential application areas suggested by FAA at HITL:***
  - ***planning transcontinental routes***
  - ***developing strategies for the NY Metroplex or South Florida.***

2) Investigate practical aspects of repeated application of COuNSEL for a specific TMI:

- Seems possible that the outcome of COuNSEL for one problem class would consistently involve the same winning coalition:
  - ***e.g. in planning a GDP for ORD, United and Delta always represented the winning coalition.***
  - Such an outcome not desirable:
  - Voices of certain flight operators never heard
- System designers should consider how such a potential outcome could be avoided.
- It is also possible that participants will learn the behavior of others from repeated applications → this lead to strategic behavior / gaming on the part of participants.

3) Broadly consider and investigate wide rage of approaches for determining flight operator weights:

- General approached used: weight based on number of involved flights (NF)
  - Recognized that directly proportional to NF was not good
  - Research needed on mapping from NF to weight required
- Feedback from HITL: CDM “philosophy” ➔ FAA treats all flights equally & flight operators have flexibility to manipulate their own flights, e.g. to give priority to larger or more important flights.
  - *This does **NOT** imply that COuNSEL weights should be based only on NF.*
  - Other possible measures: total number of passengers served, total “cost or value” of impacted flights, …
  - Creative thinking might be required to determine how to best accommodate package carriers and/or general aviation flights.
  - Any weighting scheme should be transparent and considered fair; also it should not require the airlines to divulge private information.

Conclusions

- Even though this approach is a very different way of "doing business" we are encouraged that the major players understand and appreciate the approach, i.e.
  - That TMI planning can and should be viewed in terms of trading off multiple performance goals.
  - Operator input should be used to determine how these tradeoffs are made.
- While the basic approach is sound, additional research is required and features must be refined.
- User comments suggest that the basic approach (using Majority Judgment) could be appropriate for other ATM applications – this could be a good way to improve these applications and refine the basic approach.

Future Work

- Performance-Based TMI Planning (“Step 2”):
  - There is on-going work by ourselves and others in this direction.
- Better model user benefit mechanisms:
  - Aid in understanding value of approach
  - Feed into tools to support user participation
- Model dynamics of user interactions, including “gaming” opportunities
- Refine and generalize various important tool features and components, e.g. performance tradeoff space definition, practical user voting and vector input, operator weights, define specific / practical performance metrics, etc.

Questions???

contact: mball@rhsmith.umd.edu