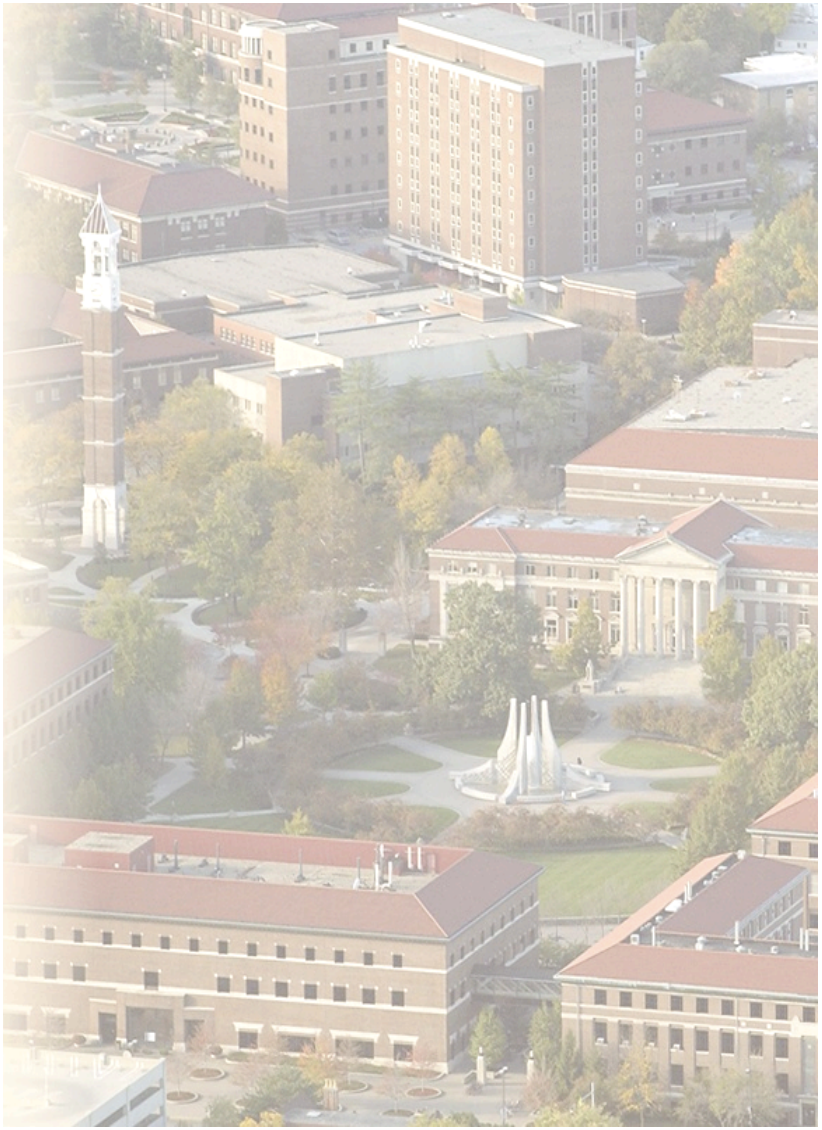
Crafting Useful Connectivity Measures

Dan DeLaurentis
Lauren Bowers
Linas Mockus
(Shashank Tamaskar)
(Zhemei Fang)

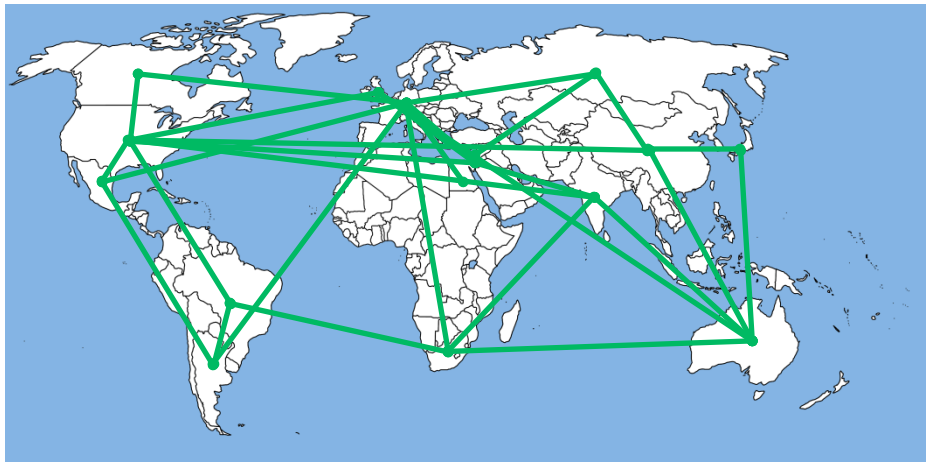
And… Thea Graham ! (and Dan Murphy!)

School of Aeronautics & Astronautics
Center for Integrated Systems in Aerospace

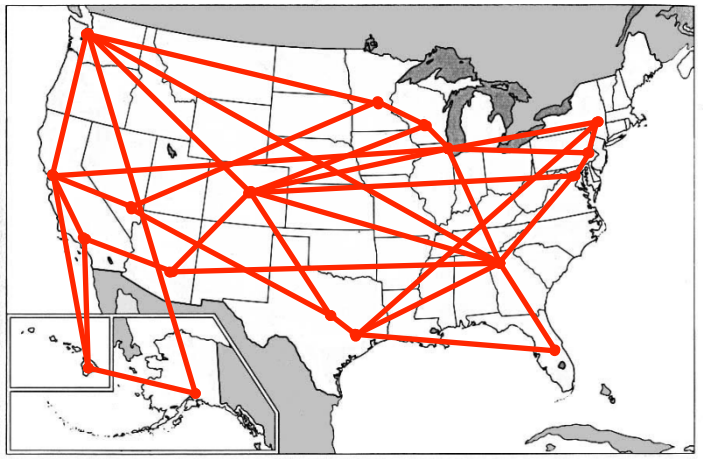
Purdue University
ddelaure@purdue.edu
765-494-0694
https://engineering.purdue.edu/people/daniel.a.delaurentis.1/



Connectivity: Where, what, and why



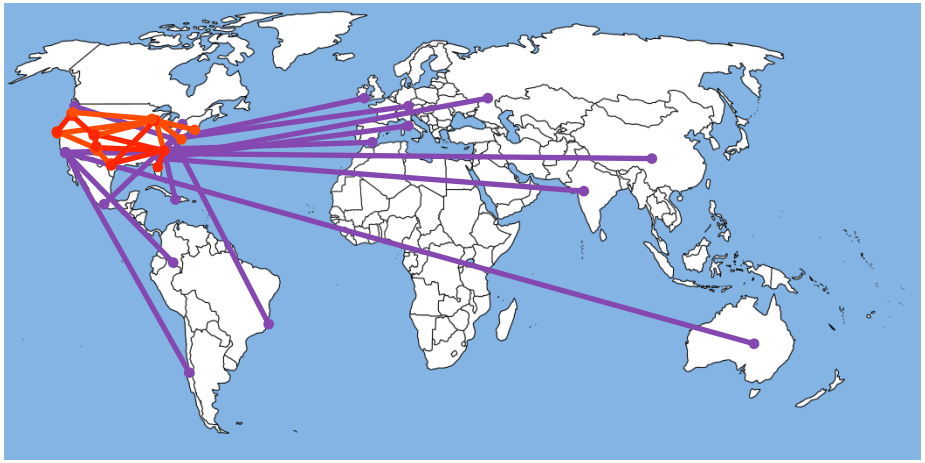
Global: country linked to country (introduce WB ACI)



US Domestic: metro linked to metro in US

Connectivity metric is merely a "reflection" of what is going on, usefulness depends:

- Network Definition
  - Data set and abstraction/granularity level
  - Flow variable (capacity, traffic, freq, etc.)
- Purpose
  - Correlative or Predictive
  - Local (nodal importance) vs Global (resilience, capacity)
  - …



US Domestic and International: US metros linked to US metros and US metros linked to foreign metros (also airport-to-airport)

What constitutes a good metric?

- **Stability** (might have to sacrifice accuracy)
- **Accuracy** (explanatory power at needed depth of detail… trade-off with stability and simplicity)
- **Sensitive and Intuitive:** Correlates with something a stakeholder cares about
- **Simple, easy, but rigorous**- (as few arbitrary parameters as possible)
- **Robust and Extensible**
  - When applied to various abstractions and contexts
  - In network hierarchy (intl' linked to regional linked to national)
- (There are more!: see, e.g., WB ACI Working Paper

The Art and Science of Modeling !

Calculating the ACI Metric*

$$\textbf{ACI} = \textbf{Geometric Average}\left(\frac{\textbf{pull of all partners}}{\textbf{maximum possible pull}}, \frac{\textbf{push of all partners}}{\textbf{maximum possible push}}\right)$$

$$\overline{ACI}_i = \sqrt{\frac{\sum_j X_{ij}}{\frac{A_i}{\sum_j B_j} + B_i}} \times \sqrt{\frac{\sum_j X_{ji}}{\frac{B_i}{\sum_j A_j} + A_i}}$$

X_{ij} : Flow (passenger, flight, etc) between nodes i, j
A_i : Repulsive potential of node i
B_i : Attractive potential of node i

Basis:
Generalized Gravity model

$$\hat{X}_{ij} = A_i B_j K_{ij} = A_i B_j \exp(-\beta f(d_{ij}));$$

$$f(d_{ij}) = \alpha \ln(1 + \frac{d_{ij}}{\alpha})$$

$\hat{X}_{ij}$: Estimated flow (passenger,flight,etc)
K_{ij} : Trade cost (impedance) from i to j
d_{ij} : Distance between nodes i, j
a: scale parameter
β : Regression parameter

Parameter Estimation (potentials A_i, B_i):
Poisson pseudo-maximum likelihood estimator

$$\Pr(X_{ij} = \lambda \mid \hat{X}_{ij}) = \frac{e^{-\hat{X}_{ij}} (e^{-\hat{X}_{ij}})^{\lambda}}{\lambda!}, \lambda = 0, 1, 2, \ldots$$

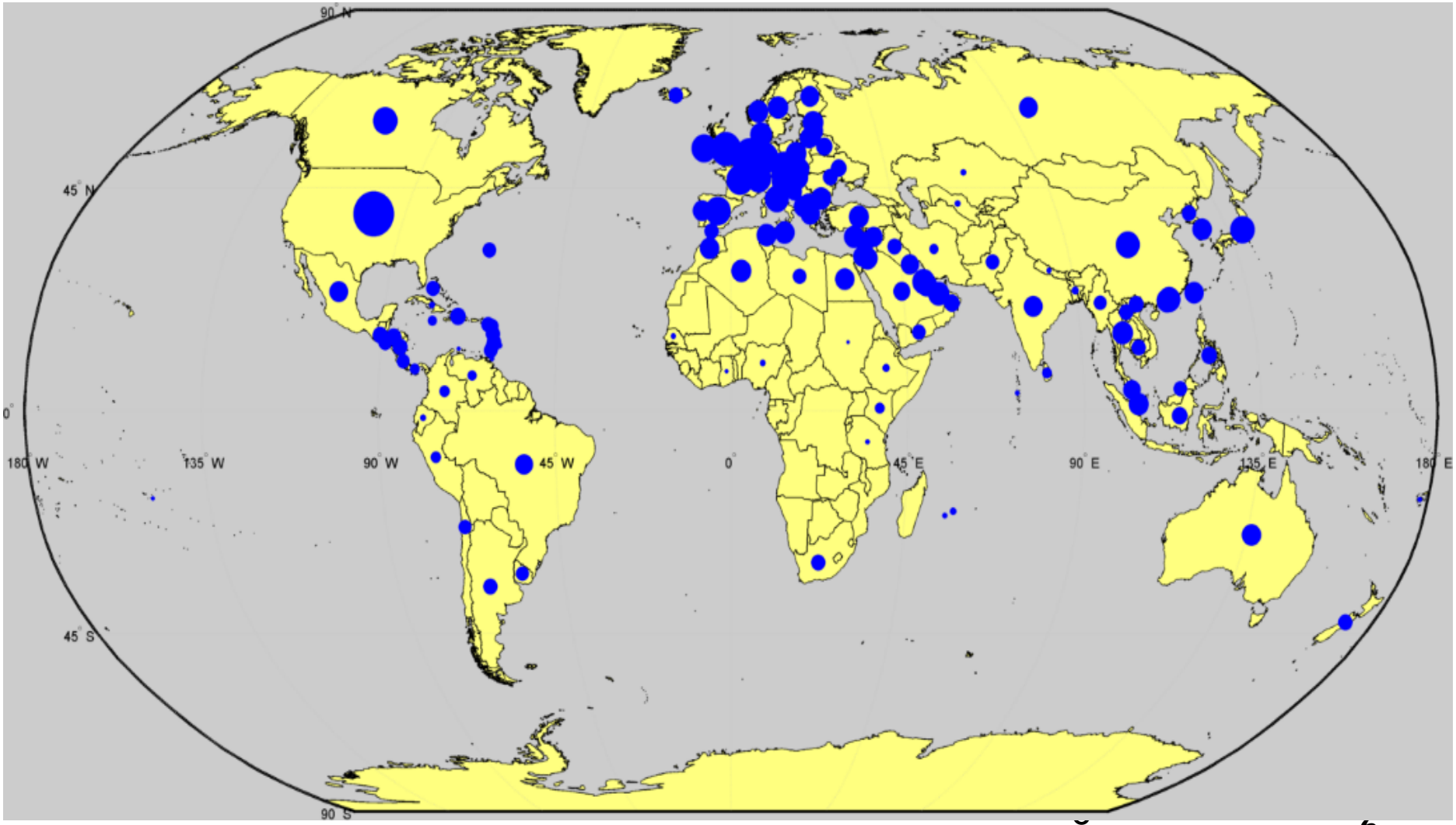
(Able to preserve equality between actual and estimated total trade flows)

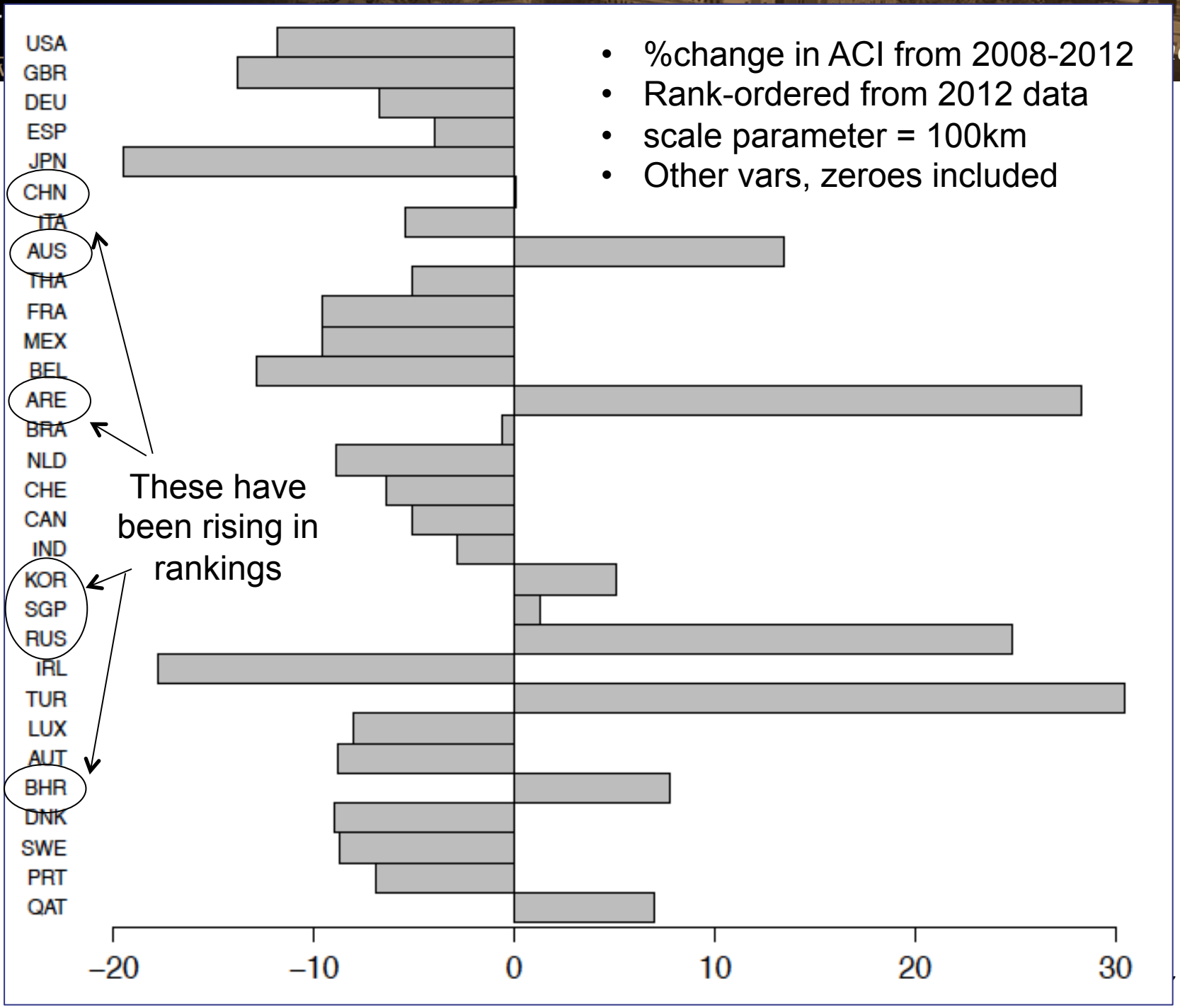
Summary of Modifications to Original WB ACI Formulation

- Flow variable: Using Seats (Capacity) vice Flights
  - Similar fit, less penalty on distance (less frequent) pairs
- Link definition: Exclude Layovers (same as WB report)
- Scale Parameter: Use value (distance) near the minimum threshold in the network
  - Performed extensive sensitivity studies
  - Balance the effect of impedance (distance) and flow
- Include a small number of "other explanatory variables"
  - e.g., two countries have had a common colonizer after 1945, are contiguous, share a common language
- Data dictates methodology…but there are constraints

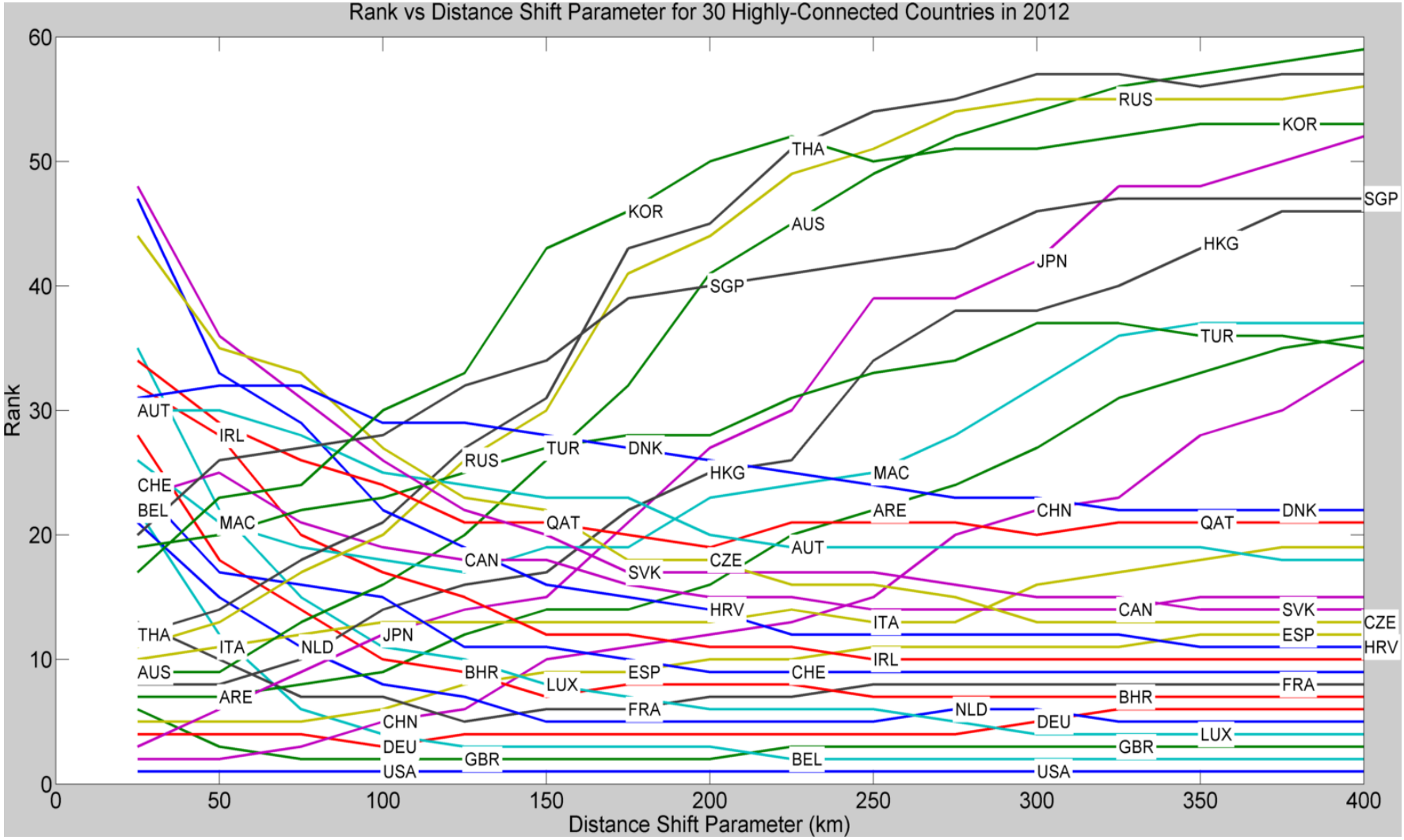
Quick View

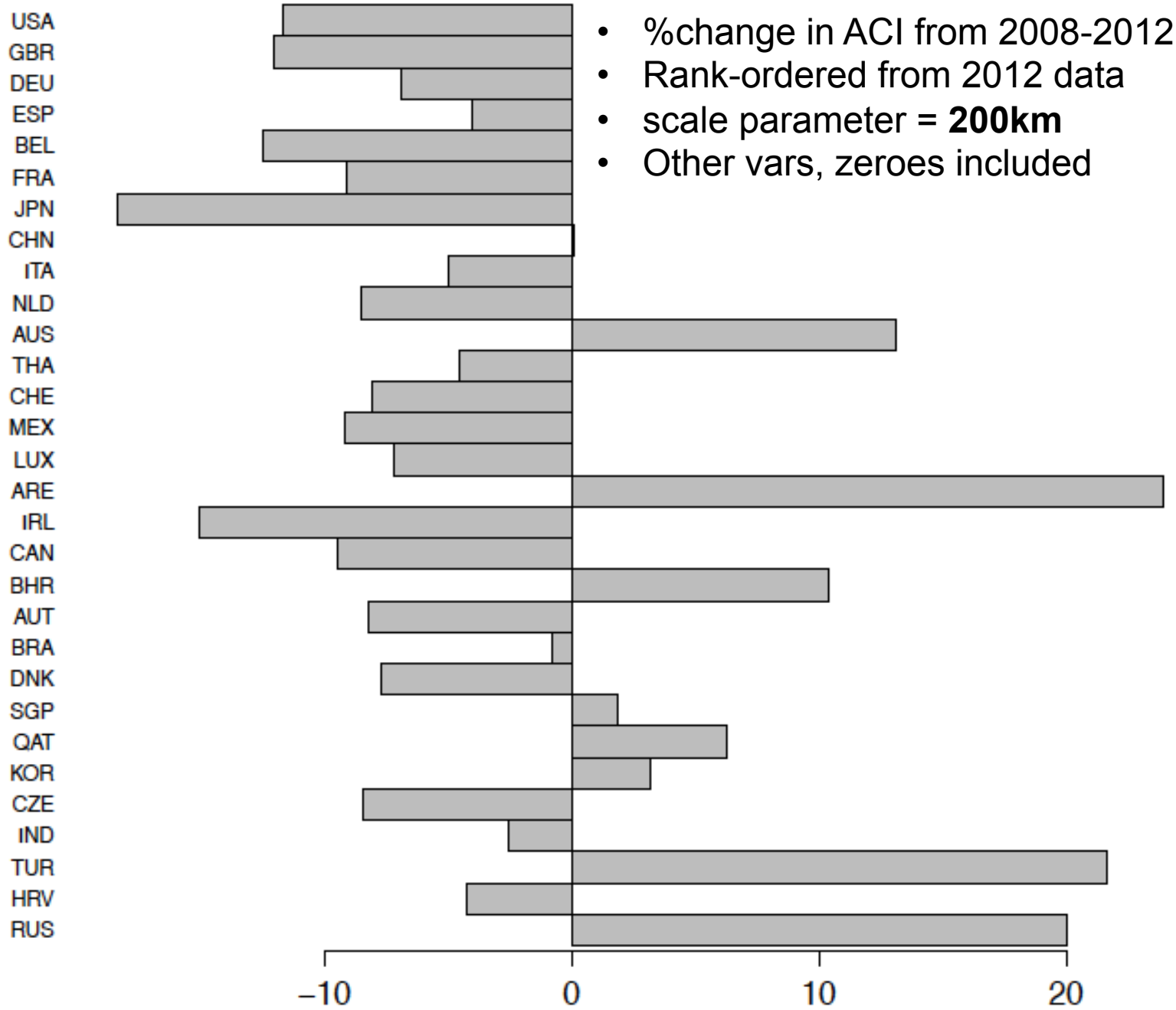
ACI 2008




USA
GBR
DEU
ESP
JPN
CHN
ITA
AUS
THA
FRA
MEX
BEL
ARE
BRA
NLD
CHE
CAN
IND
KOR
SGP
RUS
IRL
TUR
LUX
AUT
BHR
DNK
SWE
PRT
QAT
−20
−10
0
10
20
30
These have been rising in rankings
%change in ACI from 2008-2012
Rank-ordered from 2012 data
scale parameter = 100km
Other vars, zeroes included

Sensitivity to Scale Parameter





- %change in ACI from 2008-2012
- Rank-ordered from 2012 data
- scale parameter = **200km**
- Other vars, zeroes included

What about use of ACI within a large country? Region?

- **Objective**: Report on the viability of developing a US ACI
  - Methodological viability: does the formulation need to change? What is the most informative network definition?
  - Use viability: is there some correlation between ACI and SoS/ network behavior to inform policy and infrastructure planning
- **Findings**: Extensive methodological and study results that characterize capability and insights with US ACI
  - Time series and explanatory models; Reason about trends
  - Compare with related connectivity measures
- **Ongoing Activity**: Explore ACI on Regional Basis
  - North Atlantic countries and also Asian countries
  - Globally (e.g., compare North Atlantic countries with all others)
  - Examine correlations with policy and economic factors

Data Collection (years 2011 results shown)

- Only Scheduled Passenger/ Cargo Service routes are selected
- T100 Domestic Segment (US Carriers) has become our standard data source and is used for all studies in this presentation
- We have occasionally used DB1B for studies involving fare or origin-destination data
  - DB1B data sets are difficult to analyze because of size and are only used when necessary
- For airport-level studies, we used Airport ID to identify a unique airport
- For metro-level studies, use CityMarketID to consolidate airports serving the same city market
- For state-level studies, use OriginStateFips code to identify a state

ACI for different nodal abstractions

Rank	State	ACI
1	Georgia	0.2403
2	Illinois	0.2396
3	Florida	0.2315
4	California	0.2295
5	Virginia	0.2158
6	Texas	0.2152
7	New York	0.2089
8	Maryland	0.2055
9	North Carolina	0.2027
10	Pennsylvania	0.1990

Rank	Metros	ACI
1	Atlanta GA	0.3299
2	Chicago IL	0.3227
3	Charlotte NC	0.3060
4	Washington DC	0.3050
5	Detroit MI	0.2998
6	Dallas TX	0.2950
7	Philadelphia PA	0.2817
8	Denver CO	0.2754
9	Houston TX	0.2752
10	Minneapolis MN	0.2676

Top 10 Connected Airports

Rank	State	ACI
1	Hartsfield-Jackson Atlanta International	0.3284
2	Chicago O'Hare International	0.3046
3	Detroit Metro Wayne County	0.2868
4	Charlotte Douglas International	0.2868
5	Dallas/Fort Worth International	0.2868
6	Denver International	0.2831
7	Philadelphia International	0.2745
8	George Bush Intercontinental/Houston	0.2626
9	Minneapolis-St Paul International	0.2621
10	McCarran International	0.2601

Correlation of ACI with Facility and Econometric Variables

- **Operations_Commerical**: $R^2 = 0.6407$
- **LandAreaCoveredByAirport**: FAA Airport Data "Amount of land owned by the airport in acres." $R^2 = 0.2766$
- **Number of Runways**: Based on listings in Airport Runways Data. Includes helipads. $R^2 = 0.4076$
- **Herfindahl-Hirschman Index** using T100 Domestic Segment Data in 2012. $R^2 = 0.3343$
  - Increases in the HHI (ranges 0:1) generally indicate a decrease in competition and an increase of market power
  - For each airport, determined the market share of an airline by percentage of outgoing passengers
  - s_i is the market share of firm *i* in the market; *N* is the number of firms

$$HHI = \sum_{i=1}^{N} s_i^{\,2}$$

Airport-Level Prediction Statistics

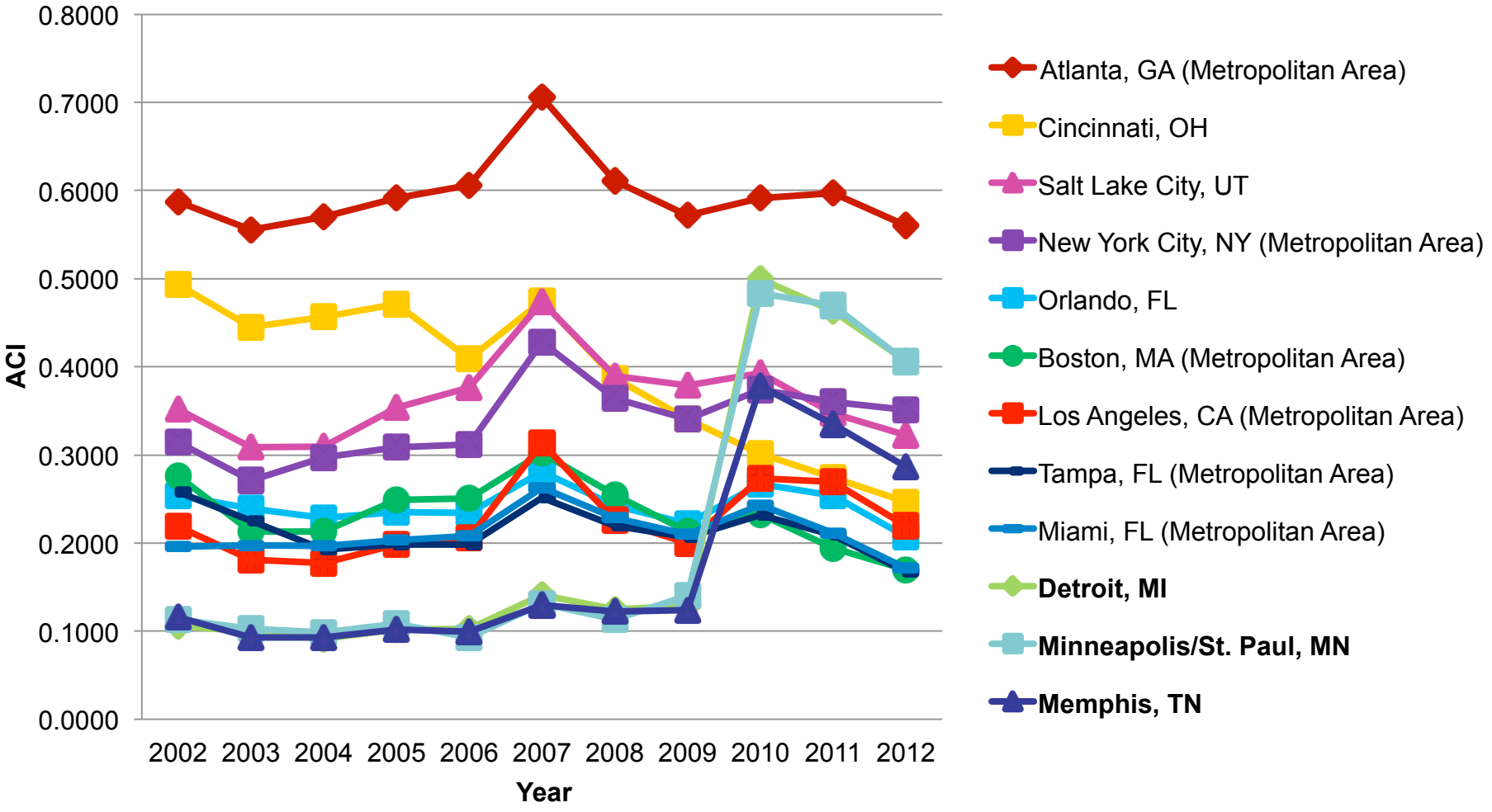
- ACI = 6.43E-02 + 4.57E 07*OperationsCommercial+ 9.58E-03*NumberRunways+ 1.78E-06*LandAreaCoveredByAirport- 8.42E-02*HHI
- R Square = 0.80
- Adjusted R Square = 0.80

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	6.43E-02	5.73E-03	11.22	0
OperationsCommercial	4.57E-07	2.68E-08	17.05	0
Number Runways	9.58E-03	1.97E-03	4.87	0
LandAreaCoveredByAirport	1.78E-06	7.59E-07	2.35	1.95E-02
HHI	-8.42E-02	5.77E-03	-14.60	0

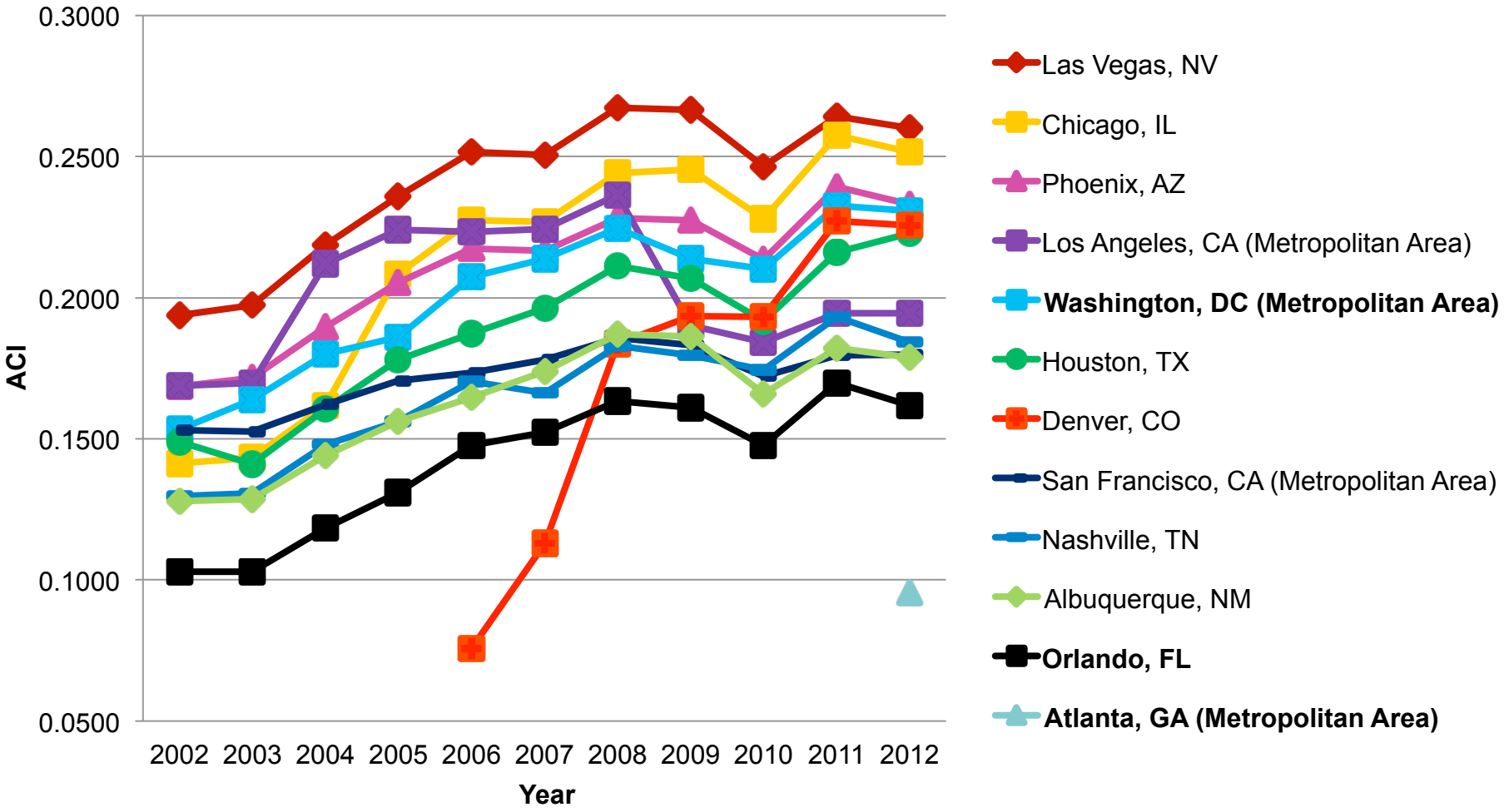
Airline Time Series Study

- Motivation: Is the ACI sensitive to airline behavior, such as mergers?
- US Domestic Network
  - Data originates from T100 Domestic Segment (U.S. Carriers) covering 2002-2011
  - ACI averaged over 10 years
- Consolidated into metropolitan areas by City Market ID
- Only Scheduled Passenger/ Cargo Service routes
- Airlines consolidated by UniqueCarrier

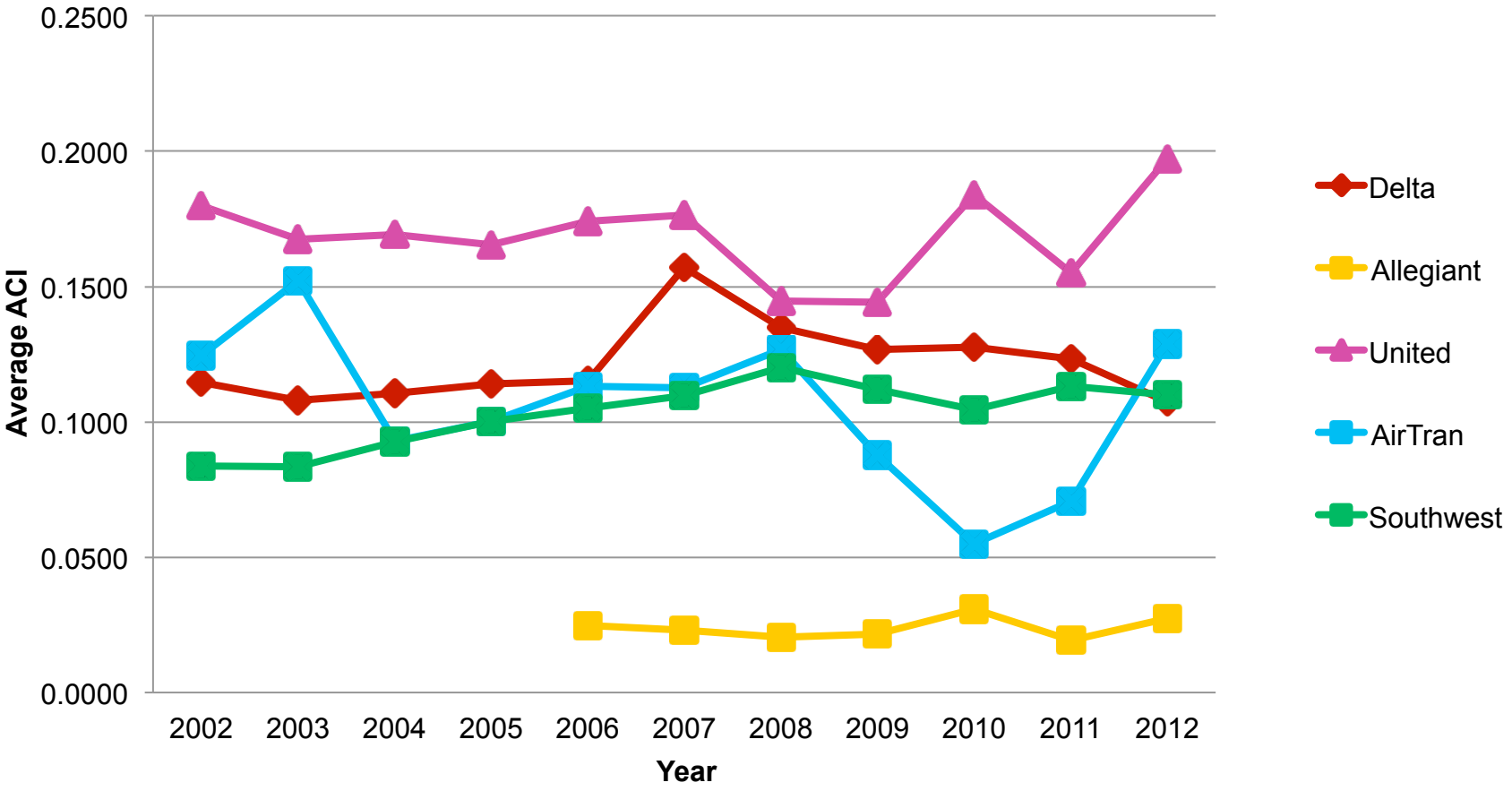
ACI Evolution For Delta
-- Top 10 Average ACI Cities



ACI Evolution For Southwest -- Top 10 Average ACI Cities



Airline Yearly ACI



Airports with Large Net Change

Airport Name	Code	ACI Change 2006-2012	Notes
Cincinnati/Northern Kentucky International	CVG	-0.0601	Cut from Delta
Northwest Florida Beaches International	ECP	0.0460	Opened for commercial flights in 2010
Charles M. Schulz - Sonoma County	STS	0.0431	Horizon Air added flights steadily from 2007 to 2012
Pittsburgh International	PIT	-0.0366	Cut from US Airways
Bozeman Yellowstone International	BZN	0.0348	Served by Allegiant Air
Bellingham International	BLI	0.0345	Served by Allegiant Air starting 2008 and later by Alaska and Frontier
Denver International	DEN	0.0328	Fastest-growing market for Southwest
Missoula International	MSO	0.0319	Served by Allegiant Air
Greenville-Spartanburg International	GSP	0.0314	Southwest started service in 2011

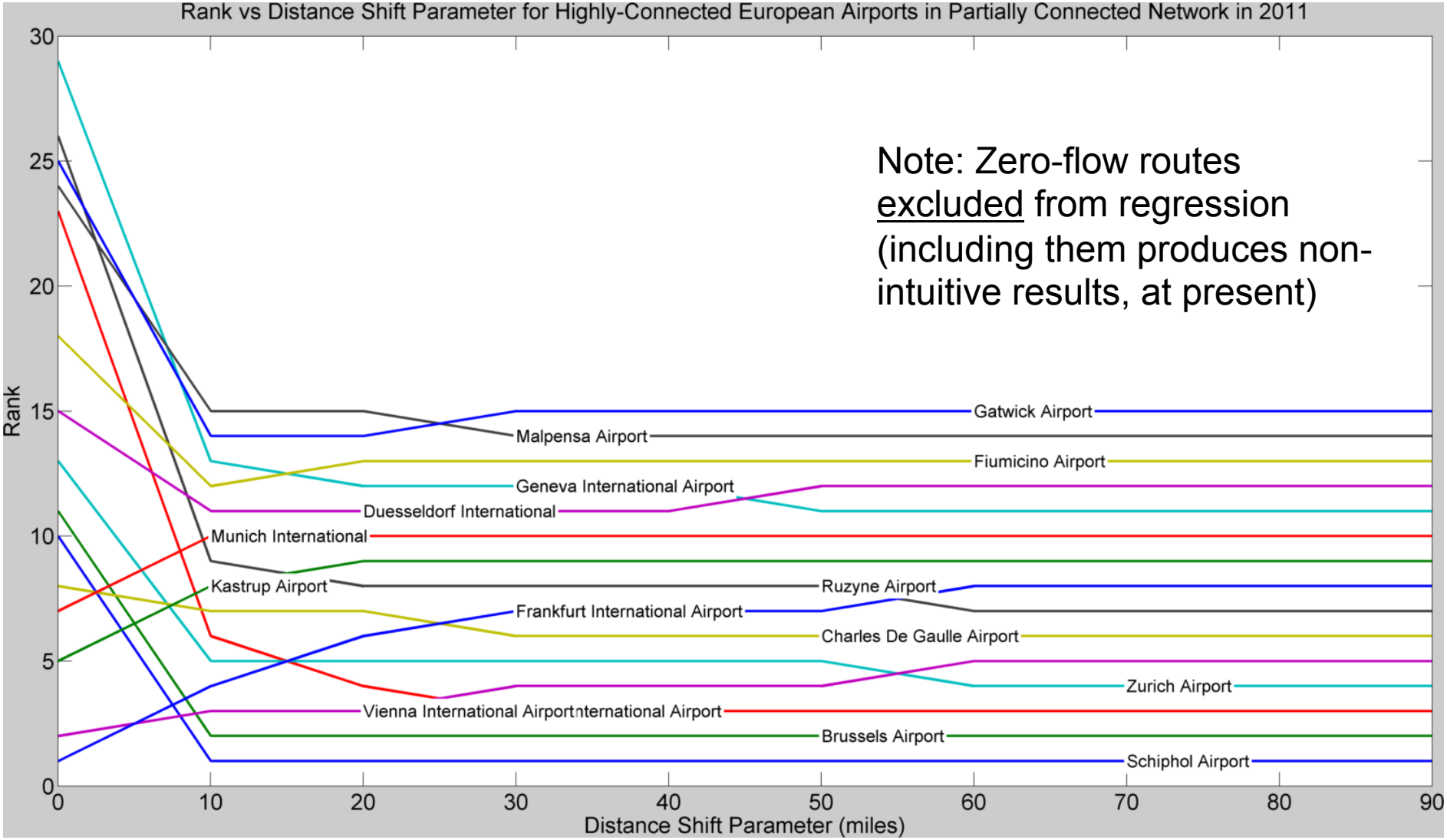
High Level Summary on US domestic Application

- ACI reveals useful information, especially metropolitan and airport levels
  - Which level to choose depends on application
- ACI is sensitive to airline behavior at airport and metropolitan level
  - Has not yet revealed anything non-intuitive
- Airport ACI can be predicted with facility-related variables (number of runways, land area, number of operations, and competitive index)
- Metro ACI can be predicted with facility and econometric variables (GDP, fare, and competitive index)

Preliminary Look: Europe

- Data on passenger flow between airports from Innovata (thanks to FAA for help)
- Only considered non-stop routes
- Data from 2011 only
- European network defined as all airports recorded in “Europe” global region
- Distance computed by calculating the great circle distance between airports using their latitude and longitude

ACI Rank Sensitivity as Scale Parameter Varies



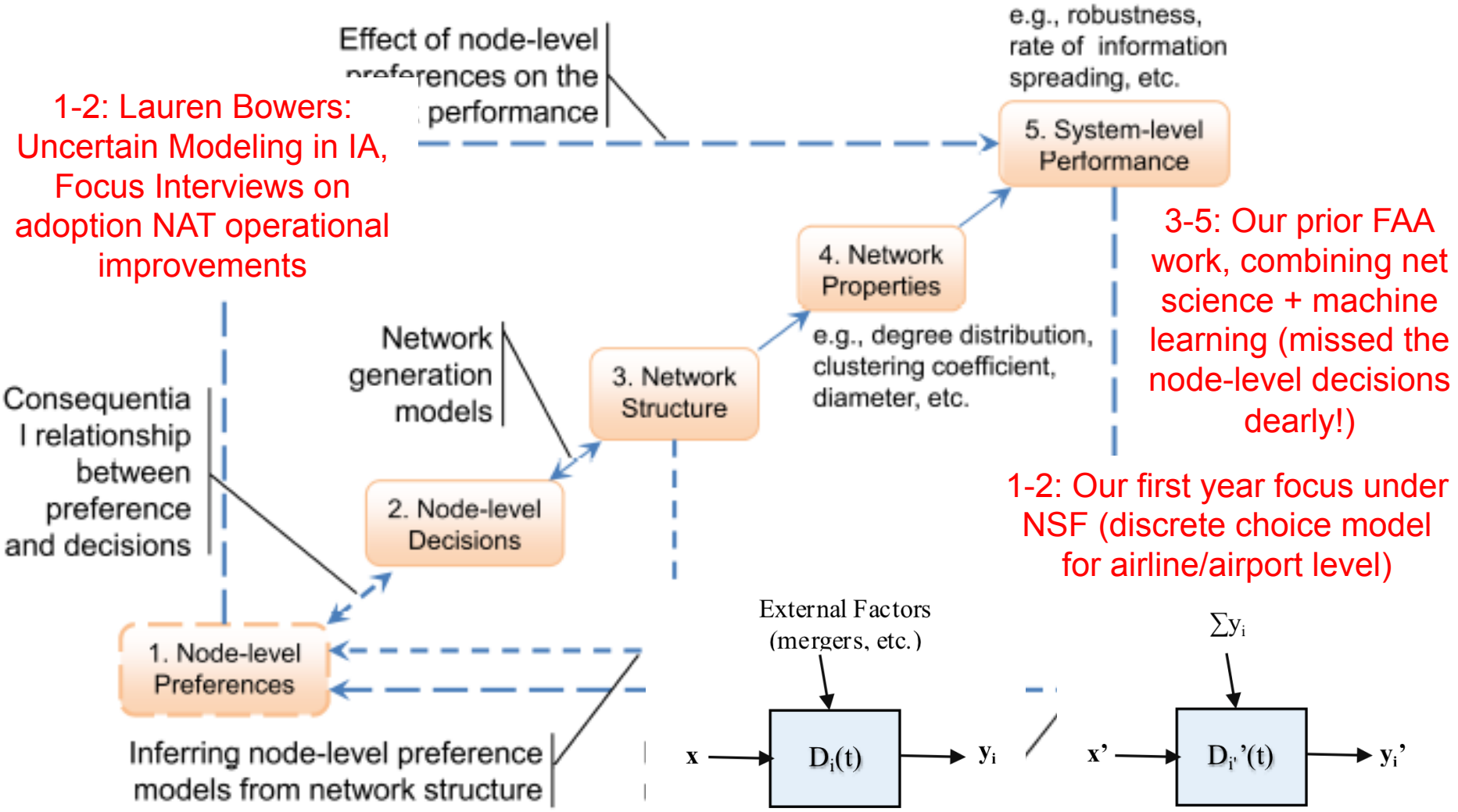
Note: Zero-flow routes excluded from regression (including them produces non-intuitive results, at present)

Challenges Remain

- **Sparsity**: Airport-to-airport networks are very sparse compared to country-to-country and state-to-state networks, creating challenges for the regression
- **Computational expense**: increases exponentially with number of nodes. There are 196 recognized countries, but 1520 airports
- **Consistency**: conflict between maintaining consistency with gravity/trade literature and tailoring the methodology to produce the most robust, sensitive, accurate, and intuitive results
  - For example, should we include routes with zero flow in the regression?
- **Heteroscedasticity:** variability of a variable is unequal across the range of values

Research Manifesto: Do you agree?

(NSF Grant 1360361; J. Panchal, D. DeLaurentis- Purdue, 2014-2017)



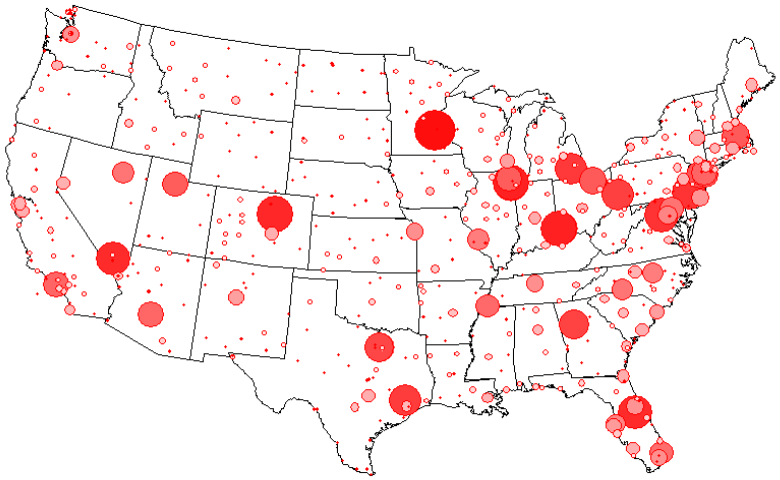
1-2: Lauren Bowers: Uncertain Modeling in IA, Focus Interviews on adoption NAT operational improvements

3-5: Our prior FAA work, combining net science + machine learning (missed the node-level decisions dearly!)

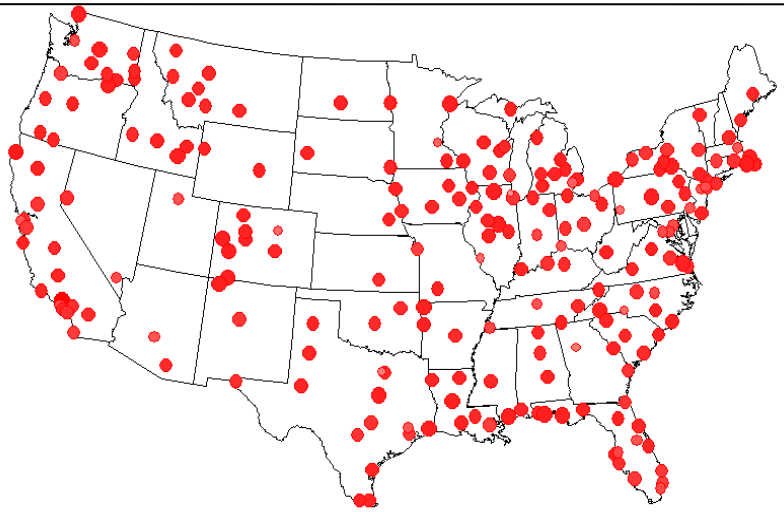
1-2: Our first year focus under NSF (discrete choice model for airline/airport level)

Forecasting Future Air Traffic (Addition of New Service Routes) (Sponsor: FAA)

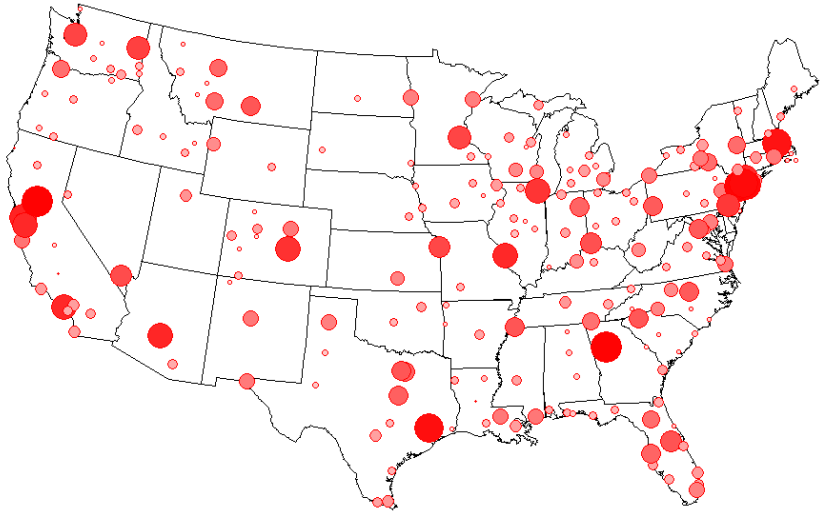
Historical Data



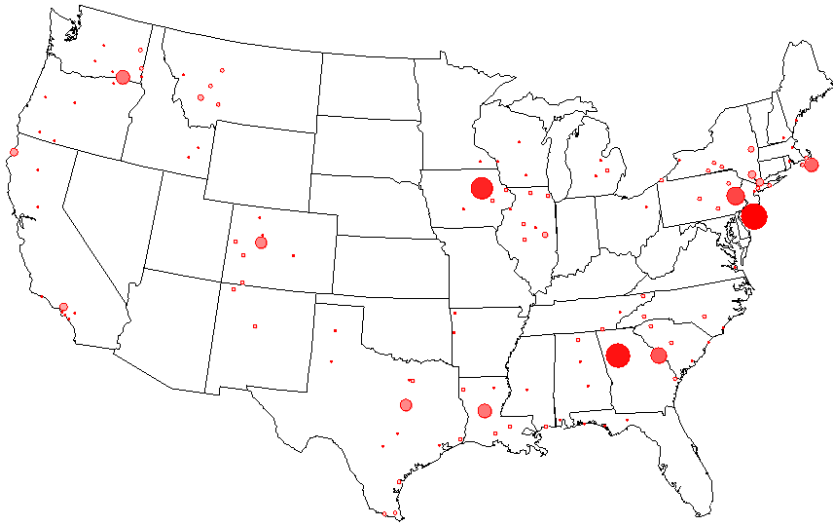
Artificial Neural Network



Fitness Function



Logistic Regression



What constitutes a good metric?

- **Stability** (might have to sacrifice accuracy)
- **Accuracy** (explanatory power at needed d… trade-off with stability and simplicity)
- **Sensitive and Intuitive:** Correlates with so…thing a stakeholder cares about
- **Simple, easy, but rigorous**- (as few arbitrary parameters as possible)
- **Robust and Extensible**
  - When applied to various abstractions and contexts
  - In network hierarchy (intl' linked to regional linked to national)

Thanks

The Art and Science of Modeling !