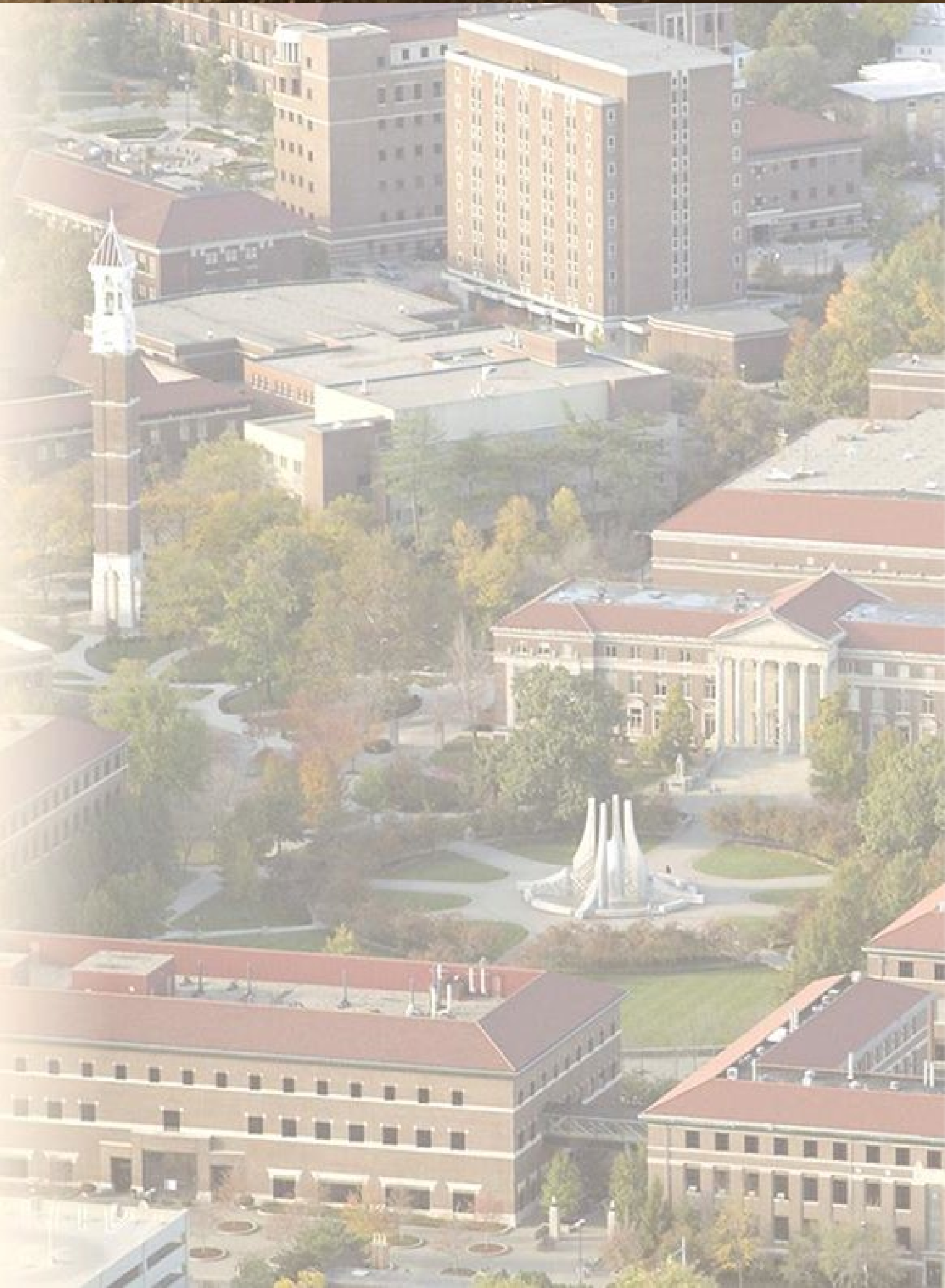
A System-of-Systems Approach to Enterprise Analytics Design: Acquisition Support in the Age of Machine Learning and Artificial Intelligence

19^{th} Annual Acquisition Research Symposium
Naval Postgraduate School

Cesare Guariniello
Prajwal Balasubramani
Daniel DeLaurentis

cguarini@purdue.edu
School of Aeronautics & Astronautics
Purdue University



Acquisition and System-of-Systems: how to control what is outside our control

Characterizing features of SoS*

- Managerial independence
- Operational independence

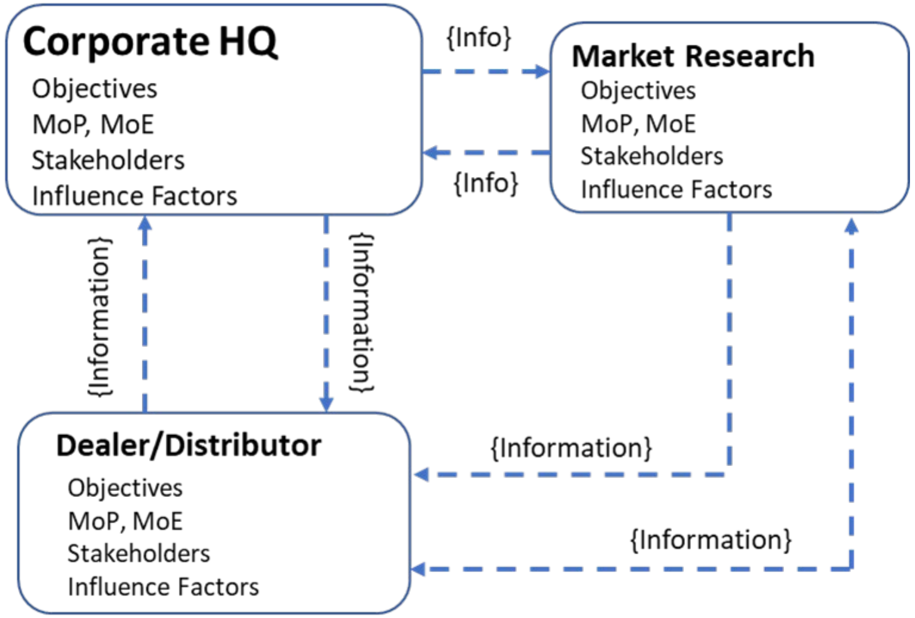
Consequence → additional uncertainty due to unavailable data

The acquisition problem

- Uncertainty on future status
- Uncertainty on impact of choices
- Uncertainty on desired objectives (SoS)
- Uncertainty on decision of other stakeholders (SoS)

Many question. Here we address some first steps

- If we can influence data flows, what info do we need and how to use it
- If we cannot, what kind of assumptions we can make?
  - How to use predictive and prescriptive data analytics, and Machine Learning to inform and support decision-making?



Conceptual problem and information flow in market decision making

* Mark W. Maier. Architecting principles for Systems-of-Systems. *Systems Engineering: The Journal of the International Council on Systems Engineering*, Vol. *1* no. 4, 1998.

Literature review: data analytics and Machine Learning methodologies

Two areas and use of methods in DoD:

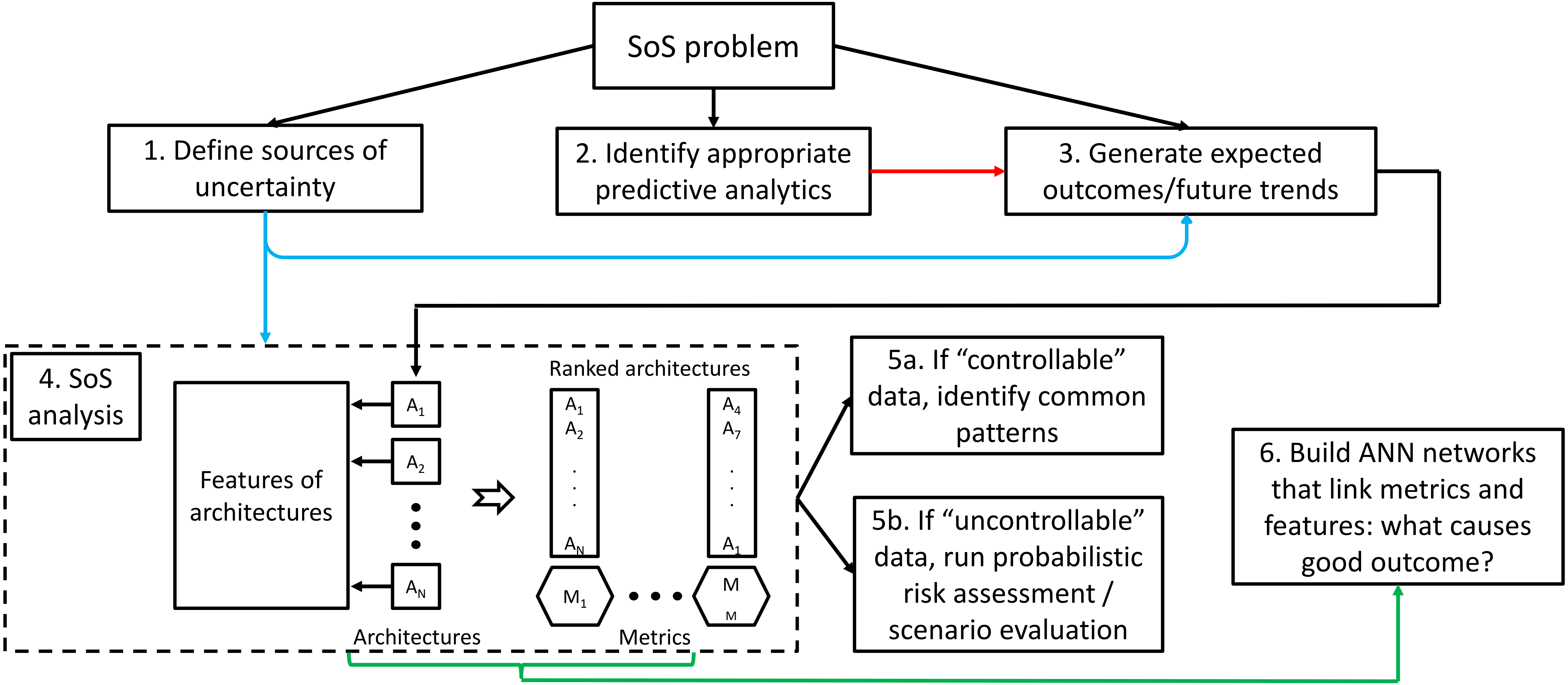
- Predictive data analytics and Machine Learning
  - Regression models and other methods to predict future
  - Artificial Neural Networks to identify patterns
- Prescriptive data analytics
  - Methods for decision-making: what actions will be better?

Results:

- Useful approach to SoS uncertainty
- Predictive methods specific to problem at hand, choice is very important
- Prescriptive methods fall a little short due to complexity of acquisition problem. Probabilistic risk assessment and/or some form of sensitivity analysis seem best choice

Method	Key Features	Notes	DoD Reference
Linear Regression	Fits quantitative/categorical predictors and continuous response to regression line using OLS	…	Moore and White III (2005)
Ridge Regression	Modification of linear regression that uses L2 norm when multi-collinearity assumption in linear regression is broken	…	Huang and Mintz (1990)
Lasso Regression	Used as a variable reduction or feature selection technique that shrinks some predictor coefficients to exactly zero to reduce overfitting from the linear regression model	…	Wang and Yang (2016)
Binary Logistic Regression	Models the log odds (using logit link) of a categorical binary outcome variable as a linear combination of quantitative/categorical predictors	…	Apte et al. (2016)
Support Vector Machine	Uses a linearly separable hyperplane to classify data into two classes	…	Wei et al. (2006)
Artificial Neural Networks	Model consisting of interconnected nodes that receive inputs and return outputs based on an activation function	…	Brotherton and Johnson (2001)
K-Nearest Neighbors	Used to classify data points based on class that appears the most among neighboring points (classification) or average of classes (regression)	…	Xiao et al. (2006)
Naive Bayes Classifier	Uses Bayes theorem to calculate probabilities of a class response and selects the class with highest probability as the output	…	Freeman (2013)
Decision Tree	Algorithm that recursively and iteratively partitions the data into homogeneous subsets to identify a target outcome	…	Apte et al. (2016)
K-means	Use to identify homogeneous clusters in a data set	…	Zainol et al. (2017)

A methodology for decision support in SoS acquisition



Example 1 - Internal uncertainty in centralized SoS: stakeholders with different objectives in Multi-domain Battle Scenario

Problem setup

- Multi-domain battle scenario
- Systems provide system capabilities and support capabilities
- Systems capabilities are combined into SoS capabilities
- Support capabilities are required to satisfy operational constraints
- Systems associated with cost and uncertainty on capabilities
- Optimization of selected portfolio of systems based on constraints, cost/budget, and risk aversion, using Robust Portfolio Optimization*
- **Different stakeholders give different relative importance to SoS capabilities**

General appearance of input to RPO in Excel format

			Support Input Requirement					Support Output Requirement		
No.	System Type	System Name	Transport Range	Transport Capacity	Refuel	Communication Relay	Operator	Transport Range	Transport Capacity	Refuel
-	-	-	Range (mi)	Capacity (lb)	Fuel capacity (lb)	Comm Relay Rating (n.d.)	Number of Operators	Range (mi)	Capacity (lb)	Fuel capacity (lb)
1	Air Systems	F-15	0	23000	36200	0	1	0	0	0
2		B-17 Flying Fortress	0	8000	14500	0	10	0	0	0
3		C-47	0	0	5369	0	4	150	3500	0
4		B-52H Stratofortress	0	15000	80000	1	12	0	0	0
5		B-2 Spirit	0	5000	32000	1	3	0	0	0
6	Ground Systems	Infantry Platoon	15	2600	0	0	40	0	0	0
7		M114 155mm Howitzer	200	3480	0	0	4	0	0	0
8		M-4 Sherman	150	1251	1015	0	5	0	0	0
9		M8 Greyhound	55	460	353	0	4	0	0	0
10		Jeep Willis	0	0	420	0	1	180	400	0
11		"Deuce and a half" (supply truck)	0	0	950	0	2	180	3800	0
12		Advanced Targeting Pod	5	200	0	0	1	0	0	0
13		TARDEC Chassis	0	0	878	0	1	100	4000	0
14		TARDEC Anti Air Module	150	1879	0	0	4	0	0	0
15		TARDEC Artillery Module	150	1900	0	0	4	0	0	0
16		TARDEC Personal Module	150	0	0	0	0	0	3000	0
17		Bofors 40 mm gun (L60)	80	1200	0	0	2	0	0	0
18		Refuel Depot	0	0	0	0	2	0	0	100000
19		Resupply Depot	0	0	0	0	2	0	15000	0
20	Naval Systems	Allen M. Sumner Destroyer	0	0	0	0	126	100	500	0
21		Higgins Boat (LCVP)	0	0	0	0	3	50	4100	0
22		Landing Ship, Tank (LST)	0	0	0	0	140	300	6000	0
23		Battleship	0	0	0	0	1,800	0	0	0
24	Space Systems	Ultrahigh Frequency Follow-on (UFO) Communication Satellite	0	0	0	0	50	0	0	0
25		Wideband Global Satellite Communication Satellite (WGS)	0	0	0	0	100	0	0	0
26	Human Systems	General Personnel	0	0	0	0	0	0	0	0

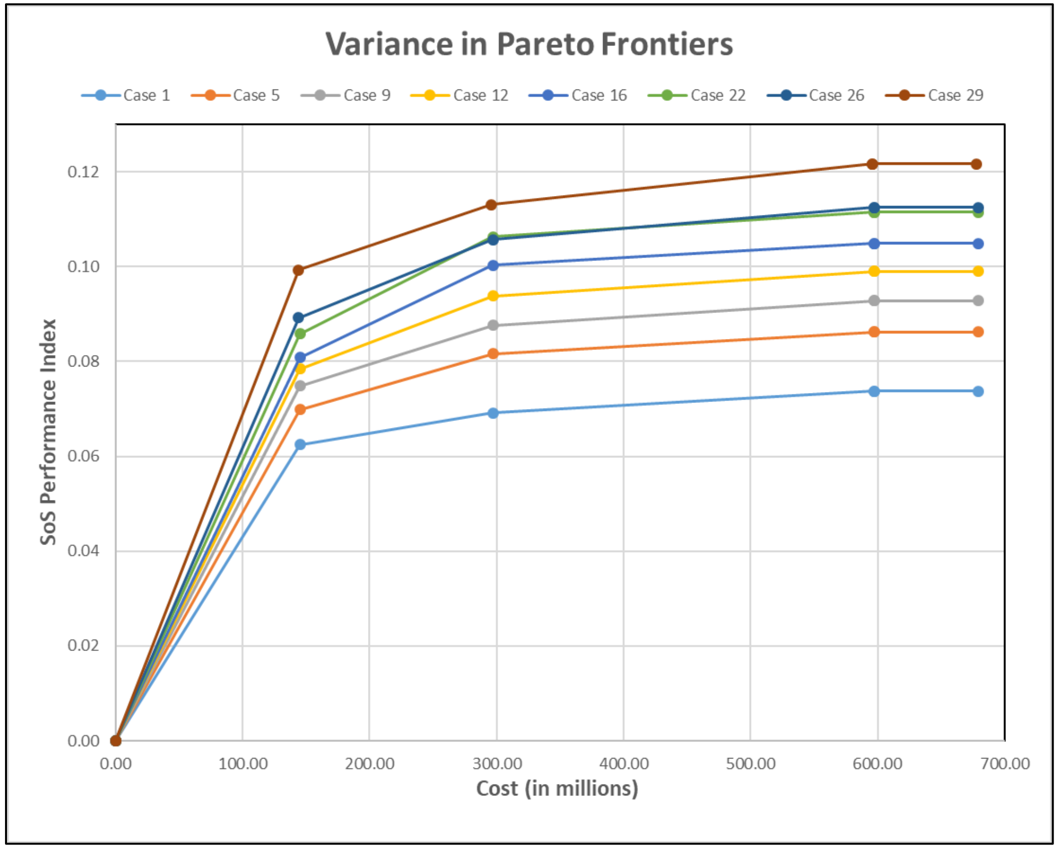
* Davendralingam, N. and DeLaurentis, D., 2013. A robust optimization framework to architecting system of systems. *Procedia Computer Science*, *16*, pp.255-264

Example 1 - Internal uncertainty in centralized SoS: stakeholders with different objectives in Multi-domain Battle Scenario

- Each different combination of relative importance of SoS capabilities result in a separate Pareto front of cost vs. performance for a given level of risk aversion
- Each point on the curve is an optimal portfolio of systems

- Two uses to guide stakeholder decisions:
  - Identify commonalities among optimal portfolios
  - Evaluate "losses" in desired performance if a specific weight combination is chosen

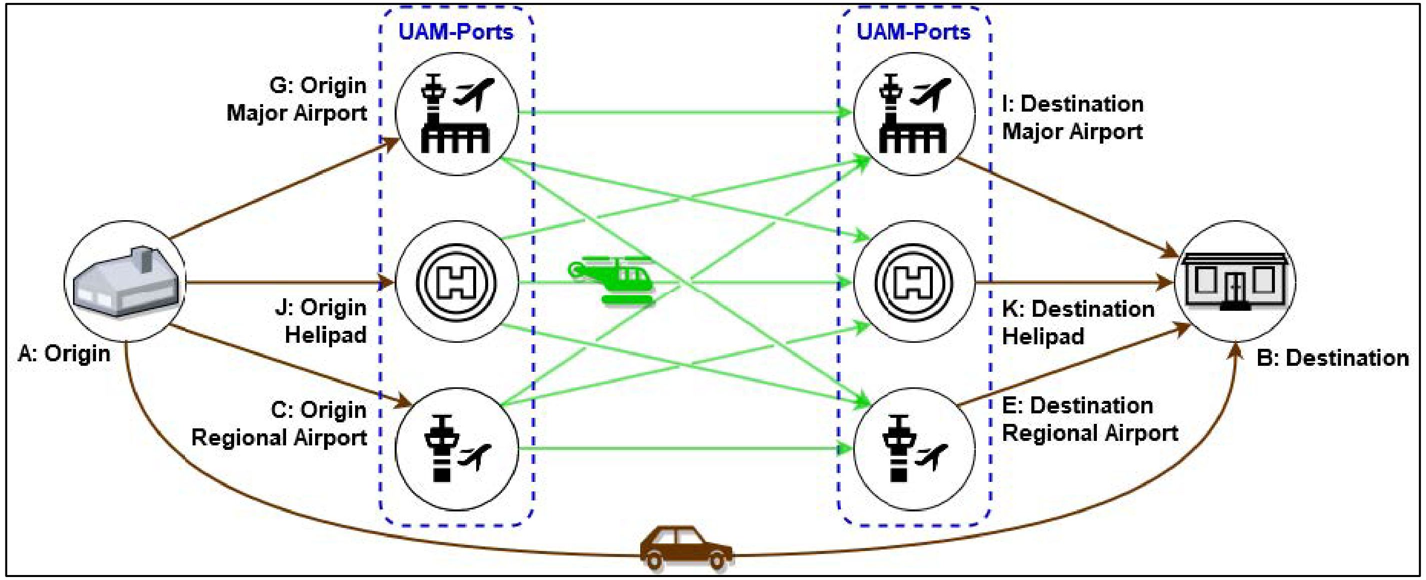
	Weights		
Cases	Air Superiority	Naval Superiority	Reconnaissance
1	0.8	0.1	0.1
2	0.7	0.2	0.1
3	0.7	0.1	0.2
4	0.6	0.2	0.2
5	0.6	0.3	0.1
6	0.6	0.1	0.3
7	0.5	0.1	0.4
8	0.5	0.2	0.3
9	0.5	0.3	0.2
10	0.5	0.4	0.1
11	0.4	0.5	0.1
12	0.4	0.4	0.2
13	0.4	0.3	0.3
14	0.4	0.2	0.4
15	0.4	0.1	0.5
16	0.3	0.6	0.1
17	0.3	0.5	0.2
18	0.3	0.4	0.3
19	0.3	0.3	0.4
20	0.3	0.2	0.5
21	0.3	0.1	0.6
22	0.2	0.7	0.1
23	0.2	0.6	0.2
24	0.2	0.5	0.3
25	0.2	0.4	0.4
26	0.2	0.3	0.5
27	0.2	0.2	0.6
28	0.2	0.1	0.7
29	0.1	0.1	0.8
30	0.1	0.8	0.1

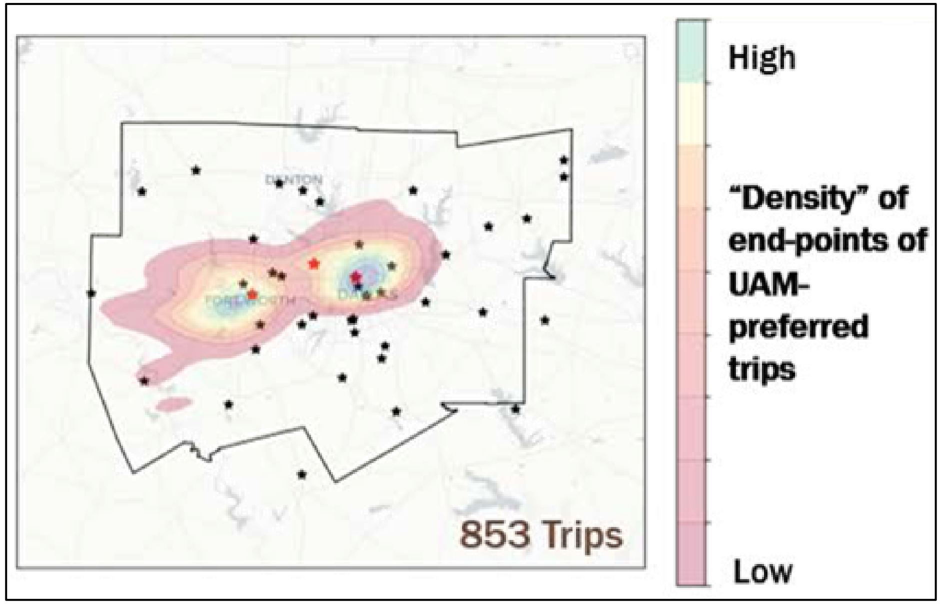


Example 2 - External uncertainty: acquisition decisions in competitive market

Problem setup

- Urban Air Mobility (UAM). Passengers can move between destinations with ground transportation or regional UAM
- Passenger's choice is based on cost of the two alternatives, income, and perceived value of time
- Service providers must decide how many 1/2/4 passenger vehicles to acquire and on which of the available routes, to maximize income
- Dallas / Fort Worth area. Info on past traffic (2010-2018) from North Central Texas Council of Governments (NCTCOG). Information on passenger income per area also available for past years
- Uncertainty on future trends, **unknown available share of market, suboptimal passenger decision**





Example 2 - External uncertainty: acquisition decisions in competitive market

Procedure and results

- Due to scarcity of data from NCTCOG, simple linear regression
- Three vertiport locations (A, B, C), six routes
- Constrained optimization on number of vehicles per type per route and number of passengers per route per day, run on predicted future with assumptions on market share, ticket price, and operational cost

Predictive analysis and full market share information

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB - BA	209	0	83
AC - CA	31	160	0
BC - CB	0	0	157
Passengers per day	2522	3840	11448
Expected income	$ 1,954,910.73		

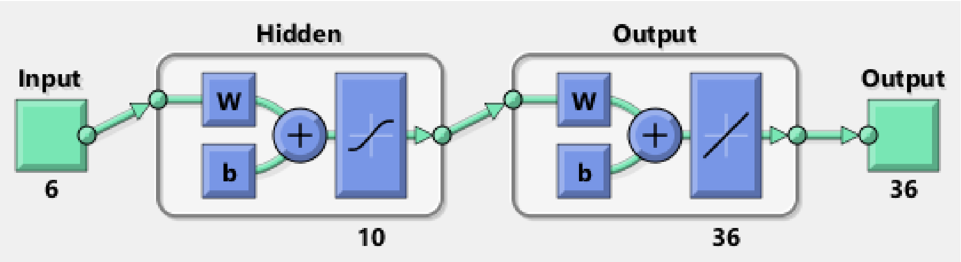
- On two routes with more passenger availability, larger vehicles are preferable even if they produce less income per passenger
- Actual income (1,000 runs): $1.893M

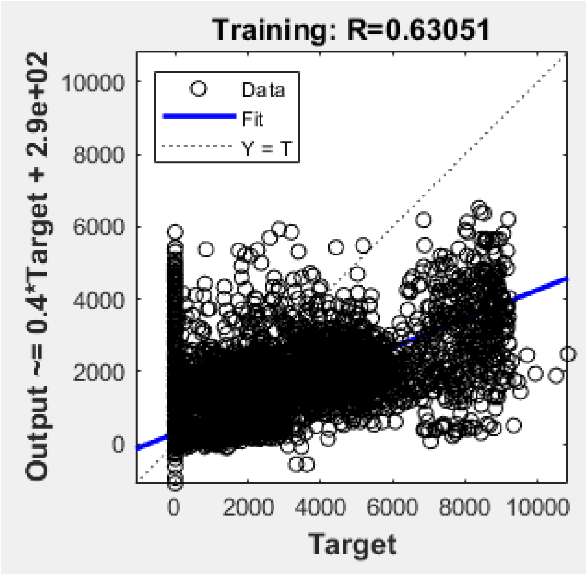
Predictive analysis and partial market share information

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB - BA	218	0	67
AC - CA	22	160	0
BC - CB	0	0	173
Passengers per day	2488	3840	11832
Expected income	$ 1,956,211.60		

- Small differences, but already suboptimal decisions (5% loss)
- Actual income (1,000 runs): $1.858M

Machine Learning to support decision-making





- Initial attempt at using Artificial Neural Network to identify relations between feature of the SoS architecture (in this case, scenarios and choices of vehicles) and resulting passenger flow and income
- All variables together result in not great fit
- The network can already drive some high-level decision

Output of NN where 1-passenger and 2-passengers vehicle have same max number of daily flights, market fraction, and gain margin on AB and BA

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB - BA	**245**	**151**	129
AC - CA	22	94	144
BC - CB	2	41	105

Output of NN where 1-passenger vehicles have more allowed max number of daily flights, market fraction, and gain margin than 2-pass on AB and BA

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB - BA	**274**	**125**	179
AC - CA	24	159	221
BC - CB	2	45	118

Conclusions and future steps

- Predictive and prescriptive analytics very useful for SoS acquisition, with caveats
  - Predictive analytics need to be chosen carefully based on problem at hand, and consequent limitations
  - Prescriptive analytics often cannot be fully executed because of amount of uncertainty
    - Modify to combined use of stochastic analysis / Probabilistic Risk Assessment
  - Machine Learning to combine predictive and prescriptive for decision support
- Need to expand to more SoS problems to establish full methodology
- Addition of Uncertainty Quantification will expand ML / NN into analysis of “why” we observe certain behavior, and which variables count more

Thank you!