SYM-AM-22-073

Excerpt from the Proceedings of the Nineteenth Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 11–12, 2022

Published: May 2, 2022

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

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

Cesare Guariniello—is a Research Scientist at Purdue's School of Aeronautics and Astronautics. He received his PhD in Aeronautics and Astronautics from Purdue University in 2016 and his master's degrees in astronautical engineering and computer and automation engineering from the University of Rome "La Sapienza." Guariniello works as part of the Center for Integrated Systems in Aerospace and is currently engaged in research projects funded by the DoD, the National Science Foundation, and JPEO. His main research interests include modeling and analysis of complex systems and SoS architectures—with particular focus on space mission architectures, aerospace technologies, and robotics.

Prajwal Balasubramani—is a Graduate Research Assistant at Purdue's Center for Integrated Systems in Aerospace and is currently pursuing a PhD in aerospace engineering. He graduated with a BTech in aerospace engineering from Jain University, India, in 2018, and received a master's degree in Aeronautics and Astronautics at Purdue University in 2020. His primary research interests are in the areas of systems engineering, system dynamics, multi-agent systems, mechanism design, and digital engineering. He has also worked on various projects involving operations research, agent-based modeling, and system-of-systems.

Daniel A. DeLaurentis—is a Professor at Purdue University's School of Aeronautics and Astronautics. He leads Purdue's Center for Integrated Systems in Aerospace, and he is also the director of Purdue University's Institute for Global Security and Defense Innovation (i-GSDI). His primary research interests are in the areas of problem formulation, modeling, and robust system design and control methods for aerospace systems and systems-of-systems. This includes agent-based modeling, network theory, optimization, and aerospace vehicle modeling. His research is conducted under grants from NASA, the FAA, the Navy, the DoD Systems Engineering Research Center (SERC) UARC, and the Missile Defense Agency.

Abstract

System-of-Systems (SoS) capability emerges from the collaboration of multiple systems, which are acquired from independent organizations. Even though the systems contribute to and benefit from the larger SoS, the data analytics and decision-making about the independent system is rarely shared across the SoS stakeholders. The objective of the research presented in this paper is to identify how the sharing of datasets and the corresponding analytics among SoS stakeholders can lead to an improved SoS capability. Our objective is to characterize how appropriate use of data sets may lead to deployment of different predictive (predicting an outcome from data) and prescriptive (determining a preferred strategy) analytics and lead to better decision outcomes at the SoS level. We build and demonstrate a framework for this objective, based on extensive literature review, which utilizes appropriate predictive and prescriptive methodologies for SoS analysis. Additionally, we propose to utilize machine learning techniques to predict the achievable SoS capability and identify sources of uncertainty derived by sharing partial datasets. A case study demonstrates the use of the framework and prospects for future improvements.

Introduction

Acquisition in the context of System-of-Systems (SoS) presents additional challenges due to the independence of stakeholders, which is a characteristic trait of this category of complex systems. Data availability can be affected by uncertainty due to the independence of stakeholder decisions. Therefore, an approach is necessary which is suitable to understand how different SoS scenarios in acquisition can be addressed with appropriate strategies to minimize the risk due to uncertainty. The use of predictive analytics to model expected behavior of variables of interest, combined with prescriptive analytics which will support adequate decision-making in the presence of uncertainty, constitutes a first step to address the difficulties in SoS acquisition. However, it is often important not only to identify best practices to address specific scenarios, but to be able to assess patterns that characterize different types of problems. We therefore propose to utilize machine learning techniques to assess achievable SoS capability that can be achieved by sharing pertinent datasets and to prescribe the information links between systems to enable this sharing. This combination of predictive, prescriptive, and machine learning methods is the foundation to our acquisition support framework.

We use a case study to demonstrate the use of the framework and to identify future steps. Previous research used a Decision Support Framework (DSF), developed by researchers at Purdue University, to simulate and analyze a fictitious multi-domain battle scenario, where the different stakeholders do not agree on the relative weight of the different achievable SoS capabilities. This example did not make use of predictive and prescriptive methodology and addressed only uncertainty due to different stakeholder objectives. The case study in this work models an acquisition problem for an Urban Air Mobility service, where a stakeholder entering the market faces uncertainty in population commute data because of partial information on potential customers and competitor market strategies. Here we use predictive (regression modelling) and prescriptive analytics to provide support towards the decision-making and locally optimal acquisition, after properly modeling the interactions due to the dynamic nature of this SoS problem. This framework is then leveraged to conduct multiple experiments with varying scenarios for stakeholders to play out, in order to build a data set on which machine learning algorithms can be applied to extract key dependencies and factors in the market space. These insights then favor acquisition decisions to build an SoS.

Background of Research and Literature Review

Acquisition in a System-of-Systems Context

System-of-Systems (SoS) capability emerges from the collaboration of multiple systems, which are acquired from independent organizations. The systems within an SoS serve two purposes: one is to meet their own independent objectives, and the second is to contribute some capability to the SoS from which all constituents can benefit. In recent decades, the fields of machine learning and data analytics have found widespread application in system design and acquisitions. It is unanimously understood that any organization acquiring a complex system employs some form of data analytics to assess a system's independent objectives. Even though the systems contribute to and benefit from the larger SoS, the data analytics and decision-making about the independent system is rarely shared across the SoS stakeholders.

Characteristics of SoS (Maier, 1998) make them quite different from simple systems and the resulting behavior of a SoS is often unpredictable just by knowing its constituent parts, due to the interactions between those parts. Given the interdependencies in SoS, when considering acquisition, it is important to recognize the stakeholders, resources, operations, policies, and economics not only of one system, but of the entire SoS. Uncertainty and possible hidden information are common in SoS acquisition, and since the SoS capability is a multi-faceted enterprise, it is hard to formulate a single set of mathematical equations that would cover all cases. Therefore, in this work we develop research towards an information-centric framework that helps inform early-stage decisions on enterprise level.

Important context for our work comes from the ambitious goals put forth in both defense and commercial sectors for Digital Engineering (DE) and its related components in various engineering functions, such as Model-Based Systems Engineering (MBSE) for the SE domain. DE and MBSE pursue the use of digital models at every phase of acquisition. Within this context, the overarching goal of our framework is to examine the impact that data features (e.g., survey categories, types of variables, ownership/privacy of data, etc.) have on the type and effectiveness of predictive and prescriptive analytics that can be employed and how the outcome can be shaped differently by examining the connectivity of data sets. This is particularly important for SoS acquisition where these data sets exist at the local system level but may not be shared at the SoS/enterprise level or vice versa. Our objective is to characterize how the sharing and the connectivity of data sets may lead to deployment of different predictive and prescriptive analytics (due to data access) and lead to better outcomes at the SoS level.

Overview of Data Analytics

Predictive data analytics provides methodologies to anticipate and predict outcomes by collecting and utilizing prior information (Joseph & Johnson, 2013; Rehman et al., 2016; Waller & Fawcett, 2013). Although using data to guide decision-making has been around since the Babylonian times, where data was recorded on tablets to predict harvest (Lo & Hasanhodzic, 2010), a major shift in the ability to reason over large amount of data emerged in 1940s with the advent of computer development, storage, and machine learning techniques. For application in complex systems, early usage of analytics can be traced back to the 1940s and '50s when data analytics models were used to predict outcomes for the behavior of nuclear chain reactions in the Manhattan Project and the weather forecasting using the ENIAC computer (Lynch, 2008).

Prescriptive data analytics, on the other hand, aims to provide an ability to generate/prescribe the best courses of action based on given information which may be obtained from a predictive data analytic outcome. Starting around World War II, the need to optimize courses of actions stimulated the development of operations research field, which in the proceeding decades led to *Analytics 1.0* for introducing data-based decision making in organizations. As the capabilities of computing and machine learning evolved to handle structured and unstructured large data sets (also known as Big Data), *Analytics 2.0* became the new paradigm across most large enterprises such as Google and Amazon (Davenport, 2013). Today, the Big Data landscape is shaped by the volume, variety, velocity, and veracity of data (known as the big four Vs of data science) and organization's ability to include this *Analytics 3.0* in the decision-making process has become fundamental to its success and profitability. It will not be a generalization to state that most successful organizations employ some form of *Analytics 3.0* for business and product development.

For SoS acquisition and capability development, deployment of *Analytics 3.0* provides a unique challenge where the individual organizations contributing the constituent systems individually employ a suite of predictive and prescriptive analytics tools. However, these analytics and the underlying data sets are rarely shared across the SoS stakeholders. Given that the SoS capability emerges from the collaboration of otherwise independent systems and considering the ever-increasing need of interoperability between systems for transitioning towards DE and MBSE, there is an imperative to connect the data sets across SoS for holistic *Analytics 3.0* capability deployment. Previous work (summarized in the Machine Learning Techniques and Application in the DoD and First Steps from Previous Work: Optimal Acquisition with Uncertainty on Objectives sections) established the significance of utilizing Machine Learning techniques and predictive and prescriptive analytics to address uncertainty in SoS acquisition.

Machine Learning Techniques and Application in the DoD

This research builds upon previous work (Raz et al., 2020) which analyzed the use of Machine Learning techniques in DoD applications. Table 1 summarizes the findings from Raz et al., 2020, describing various Machine Learning methodologies, their assumptions, and applications in DoD research.

Table 1. Summary of ML Methodologies

Method	Key Features	Assumptions	DoD Reference
Supervised Learning			
Linear Regression	Fits quantitative/categorical predictors and continuous response to regression line using OLS	Linear parameters, constant error variance, independent error terms, errors are normally distributed, random sample of observations, no multi-collinearity	Moore and White III (2005)
Ridge Regression	Modification of linear regression that uses L2 norm when multi-collinearity assumption in linear regression is broken	Standardization of predictors, linear parameters, constant error variance, independent errors	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	Model has sparsity, irrepresentable conditions (Zhao & Yu, 2006)	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	Independent observations and errors, binomial distribution of response variable, linearity between logit of response and predictors	Apte et al. (2016)
Support Vector Machine	Uses a linearly separable hyperplane to classify data into two classes	Independent and identically distributed observations, margin is as large as possible, support vectors are most useful data points	Wei et al. (2006)
Artificial Neural Networks	Model consisting of interconnected nodes that receive inputs and return outputs based on an activation function	Independence of inputs	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)	Similar inputs have similar outputs (Weinberger, 2018)	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	Predictors are conditionally independent of each other given the response	Freeman (2013)
Decision Tree	Algorithm that recursively and iteratively partitions the data into homogeneous subsets to identify a target outcome	Entire training set is at root node, quantitative predictors must be discretized	Apte et al. (2016)
Unsupervised Learning			
K-means	Use to identify homogeneous clusters in a data set	Clusters sizes are similar and spherical in form	Zainol et al. (2017)

Different stakeholders might use different Machine Learning techniques for prediction and decision making. In the presence of stakeholder independence, it is important to recognize what information is available and what information will be at least partially hidden (thus causing uncertainty), and then choose appropriate techniques to deal with the uncertainty. When the uncertainty is only on the desired objectives of stakeholders, we can design experiments to treat multiple cases and run predictions based on possible different choices of the stakeholders. However, when the uncertainty is due to multiple, external factors (for example, competitors' decisions, stakeholder preferences, and fluctuation of the market, as shown in Figure 1), simpler predictive analytics are a better choice. Predictive analytics will provide baseline scenarios for subsequence application of prescriptive analytics, which can support educated decision-making that will cause robust decision based on the expected scenarios made available by the predictive analytics.

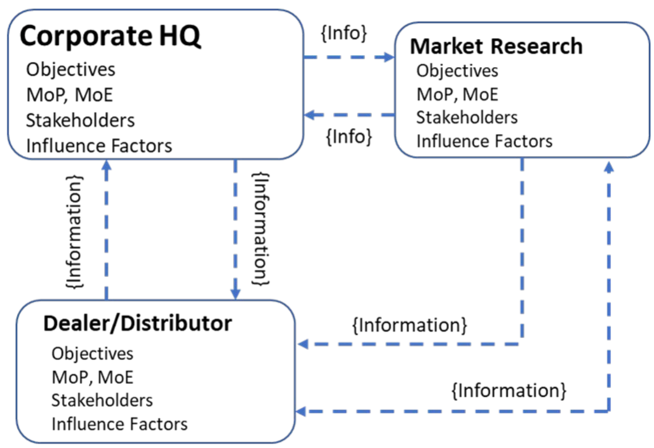

Figure 1. Conceptual Problem to Identify Impact of Data-Set Connectivity

First Steps from Previous Work: Optimal Acquisition with Uncertainty on Objectives

A precursor to this work, an application on a Naval Warfare Scenario (Raz et al., 2020) demonstrated the use of a Decision Support Framework (DSF) to assess optimal acquisition where multiple stakeholders might not agree on System-of-Systems-level objectives. The DSF identifies optimal portfolios of systems that, accounting for operations constraints, budget limitation, and uncertainty on capabilities, provide the best SoS performance.

Since in this case the uncertainty is due to different interpretation and preferences about mission requirements and objectives, the DSF has been simply used to run multiple scenarios, each one having a different combination of preferred SoS-level objectives. The resulting optimal portfolios of systems have then been compared to identify common occurrences of certain systems in optimal portfolios for any given budget limitation and risk acceptance. At the same time, these results identify parts where additional information might be required to support optimal decisions (for example, when the difference in preferred objectives results in extremely different optimal portfolios).

Figure 2 shows the different combinations of weights, representative of the importance given by different stakeholders to various SoS-level objectives in the Naval Warfare scenario. Figure 3 shows pareto fronts of optimal portfolios providing weighted SoS-level performance (vertical axis) based on budget (horizontal axis). The different lines represent cases from Figure 2, i.e., different stakeholder preferences and different weights given to SoS-level objectives.

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

Figure 2. Test Runs with Variation in Weight Distribution

Some stakeholder decisions result in higher SoS-level performance, and different budget levels cause different set of weights (and the resulting optimal portfolios) to provide better performance. We can therefore notice that any uncertainty in SoS capability preferences affects the resulting performance of the SoS portfolios. Since its point on the pareto frontier corresponds to a different portfolio of systems, we can compare the optimal portfolios to identify common systems and to assess areas where further information might be needed, if available.

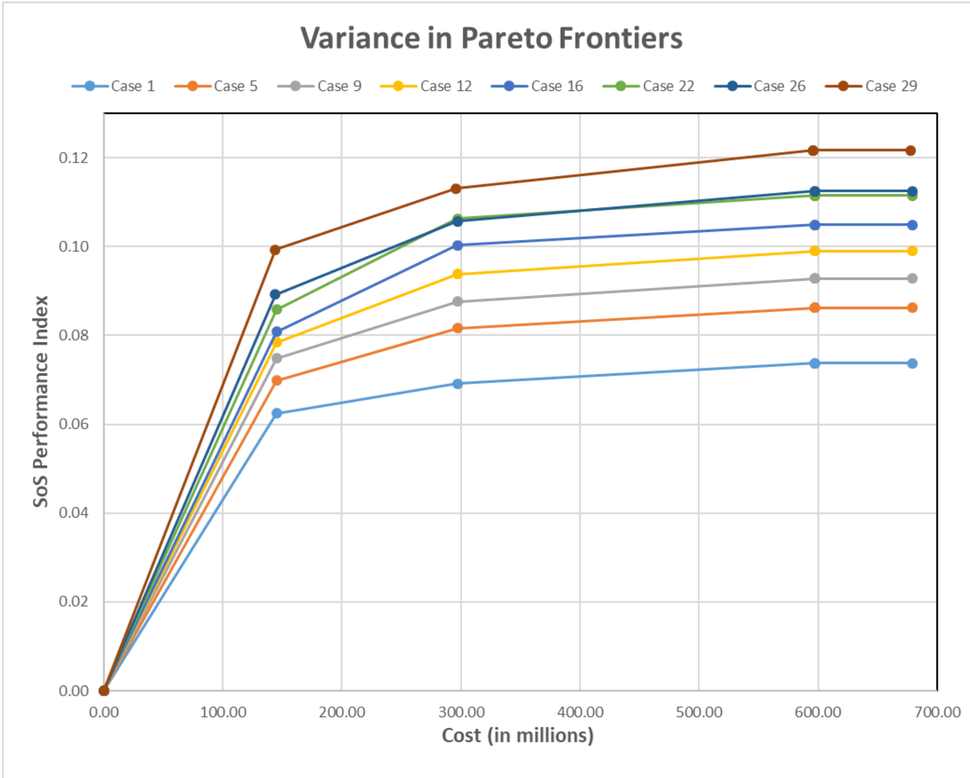

Figure 3. Variation in Pareto Frontiers Across the Cases

Optimal Acquisition with Uncertainty on External Factors

In this application, we consider much less "controllable" sources of uncertainty, and lack of information due to external factors. We model a scenario of Urban Air Mobility (UAM) in the Dallas, TX region, where a new stakeholder plans to participate as a provider of UAM services. The new provider has access (possibly limited) to data from the past that can suggest how many passengers would be willing to use UAM services. However, these data present some degree of uncertainty about the future, due to the many factors that can affect the number of passengers. Furthermore, the new stakeholder might not be fully aware of the decisions of competitors, which would affect the quota of available market to which the new stakeholder would have access. Predictive and prescriptive analytics, together with the use of Machine Learning, can support decision making in this context.

We model the expected user traffic in this transportation SoS by looking at previous years' data from the North Central Texas Council of Governments (NCTCOG). This is data on total population travel frequency between home and work districts, therefore it represents the total number of travelers between different locations. We then need to predict the proportion of travelers willing to use UAM services as an alternative to driving or using public ground transportation. Travelers' decision is affected by their income, the cost of commuting by UAM vehicles, and the perceived value of time. Figure 4, from Maheshwari et al. (2021) shows the process of comparison of UAM routes against ground transportation, when there are no potential competitors. This comparison gets specialized to each specific region of interest.

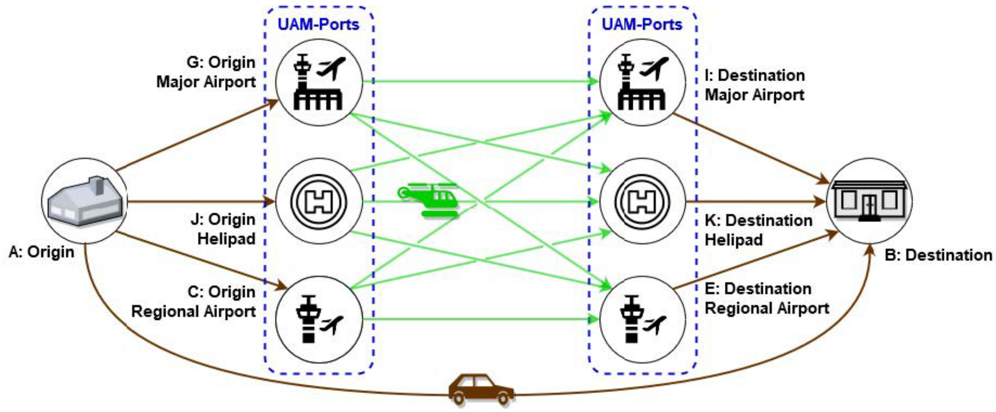

Figure 4. Alternative Routes of Travel from A to B: UAM Vehicles with Different Origins and Destinations and Ground Transportation

Urban Air Mobility Scenario and Problem Setup

Three regions in the Dallas-Fort Worth area (called A, B, and C in the results) are modeled in our study as potential UAM points of operation based on the NCTCOG data on population travel frequency. Figure 5 shows the location of vertiports for UAM in this study and the density of origins and destinations pertaining to travelers who would prefer UAM transportation. We model 6 routes operating between these three regions. Additionally, the stakeholder is provided with acquisition decisions for the type of UAM vehicles with varying seating capacity (1, 2, and 4 maximum passengers). These vehicles have different operational costs and different ticket price. The stakeholder decisions to build their UAM portfolio are motivated to maximize the expected total income per day. Our predictive model uses historical data from NCTCOG to predict the population travel frequency for 2022 for which the UAM network is being modelled as the acquisition problem. Since the NCTCOG data is quite sparse, we used simple linear regression models. We run an optimization problem with constraints on the maximum number of allowed flights per day and the maximum licensed number of vehicles for UAM. The decision variables are the number of acquired vehicles per type (seating capacity) per route. We extend the optimization to include the uncertainty in the predicted data, after running the predictive analysis with a 95% confidence interval.

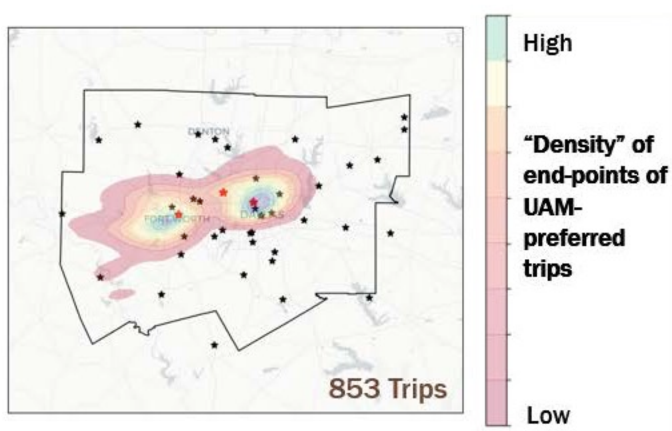

Figure 5. Location of Vertiports (Red Stars) in the Dallas–Fort Worth Area and Areas of Preference for UAM Transportation

Within this context, we also consider that UAM operators might be carrying on their activities in a competitive market, where other stakeholders are present. Based on the amount of available data concerning the travelers' decision (in its turn, influenced by income, cost of commuting by UAM vehicles and perceived value of time), the presence of competition, and some degree of uncertainty, we again compute the expected optimal choices for UAM operators to compete in the market.

Results—PredictiveAanalytics and Optimization

In the first example, we assumed perfect information about the number of passengers willing to fly with UAM vehicles, based on their income, the alternative ground-based travel time, and the personal perception of the value of time. The only source of uncertainty in this case is due to the prediction of travel frequency for 2022, based on past travel data starting in the year 2010. We also assumed perfect knowledge about the share of the market which is already taken by existing competitors. The income is based on ticket price (varying by route and type of vehicle) and on operational costs. We ran an integer linear optimization, where the decision variables are the number of vehicles per type per route, and the number of passengers actually flying with the stakeholder's vehicles. Constraints exist on the maximum number of vehicles that can be acquired and on the maximum number of flights per vehicle per day on each route. Results of this optimization based on the expected values for flying passengers in 2022 are shown in Table 2.

Table 2. Optimal Choices with Full Data Available and Prediction for the Year 2022 in the Dallas Area

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
Income	$ 1,954,910.73		

Results show that on two routes with more passenger availability, larger vehicles are preferable even if they produce less income per passenger. The intermediate-sized vehicle is present only on two routes (from A to C and from C to A), together with the small vehicle. The expected income is about $1.955 million. However, due to uncertainty, the actual income will be slightly smaller. We ran 1000 scenarios according to the expected distribution and using the optimal choice of vehicles, which resulted in an income of $1.893 million.

As a first step towards the study of support for decision-making in SoS, we then ran a scenario where the actual market share is unknown. We assumed that 33% of the passengers are available to fly with the new stakeholder, which is only slightly different from the actual market share that we used in the first case (ranging between 30% and 45%). We also increased the uncertainty in the predictive phase. Despite the small differences, there are already changes in the choice of optimal fleet, as shown in Table 3. The routes AB and BA saw an increase in the number of small vehicles, while more large vehicles were acquired for the routes BC and CB.

Table 3. Optimal Choices with Partial Data Available and Prediction for the Year 2022 in the Dallas Area

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		

The expected larger market share along the routes BC and CB suggests an income slightly larger than the previous case. However, despite the very small differences, the presence of incomplete data causes suboptimal acquisition. When running 1000 scenarios according to the expected distribution of flying passengers, the resulting income is $1.858 million, lower than the income achieved with the optimal choice in the previous case and about 5% lower than the expected income. These results show how even just a small amount of uncertainty can have a large impact on the decision-making process and its outcome.

Machine Learning to Enhance Prescriptive Analytics

To further extend the stochastic optimization, it can be useful not only to know the results of optimization in different scenarios, but also to understand how the different inputs (which are the source of uncertainty) affect the output variables. We therefore trained a Neural Network, implementing 1090 scenarios with variable parameters, which modify the optimization problem. For each route and type of vehicles, the parameters include the maximum number of flights per day, the market fraction available to the new stakeholder, and the feasible gain margin (that is, how much the stakeholder is desiring to earn out of selling tickets. This needs to overcome the operational cost, but high prices of tickets will cause fewer travelers to choose UAM vehicles over ground transportation).

Figure 6 shows a neural network trained in Matlab, where the inputs are different level of maximum number of flights per day, market fraction available, and feasible gain margin on routes AB and BA for vehicles with 1 and 2 passengers. The output variables are the result of the optimization with full data available and prediction for the flying passengers in 2022.

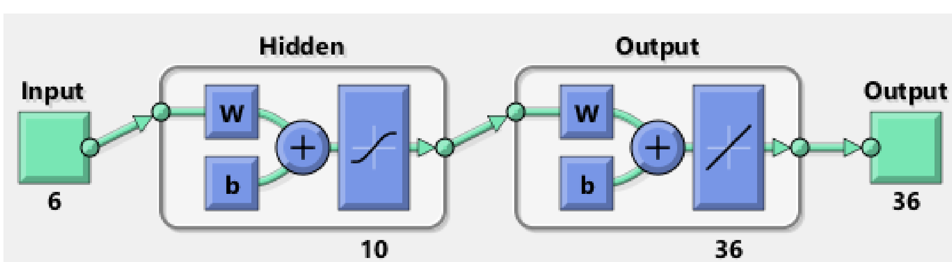

Figure 6. Neural Network for the UAM Scenario in the Dallas–Fort Worth Area

Figure 7 shows the result of the training of this partial Neural Network. Outliers are caused by the fact that this network is based on a partial number of inputs. However, the fit is good enough to utilize the network for prescriptive analytics and to quickly run analysis of a large number of scenarios with changing inputs. For example, Tables 4 and 5 show the outcome of the Neural Network with different inputs for the parameters. When variables pertaining to market fraction and maximum number of allowed flights per day increase for 1-passenger vehicles and decrease for 2-passenger vehicles, we can notice not only how this impacts the route directly affected by the parameters (AB and BA in this case), but also how the variables on available market share and desired gain margin, united with the number of passengers which will decide to use UAM vehicles, affects the choice of acquiring more 4-passenger vehicles. Limitations of this approach and solutions to overcome the limitations are presented in the following section.

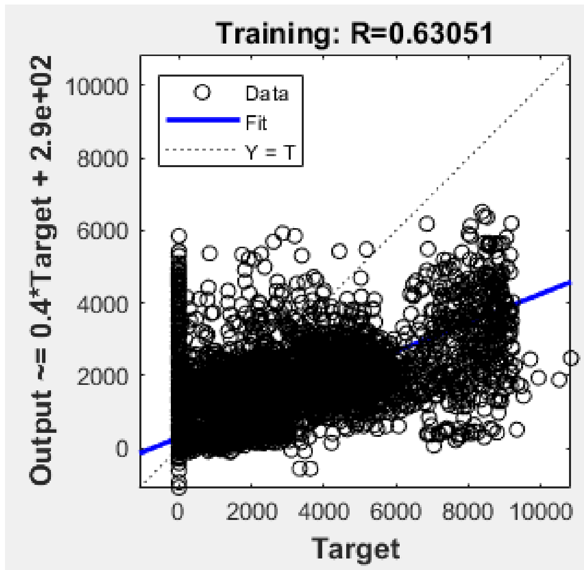

Figure 7. Fit of Training Runs of the Reduced Neural Network for the UAM Scenario

Table 4. Output of the Neural Network with Similar Values for Max Number of Flights Per Day, Market Fraction, and Gain Margin for 1-Passenger Vehicles and 2-Passenger Vehicles on Routes 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

Table 5. Output of the Neural Network with Higher Max Number of Flights Per Day, Market Fraction, and Gain Margin for 1-Passenger Vehicles with Respect to 2-Passenger Vehicles on Routes 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 Work

Building on top of previous research on the use of predictive and prescriptive analytics for acquisition in a System-of-Systems context, we expanded our framework that deals with the uncertainty derived from potential lack of data and information, to treat cases where the uncertainty is due to external factors and can heavily affect the outcome of decisions. The presence of incomplete data, together with uncertainty due to fluctuation of variables in the future and with the presence of independent stakeholders, can produce suboptimal choices in acquisition. Preliminary results showed promising directions for the use of predictive and prescriptive analytics to address this type of problems. An application to an Urban Air Mobility scenario in the Dallas–Fort Worth region was used for various demonstration purposes. First, we showed how decision-making in the presence of full information about some variables (fraction of market availability) while others are affected by limited uncertainty (prediction models for number of passengers willing to use UAM services based on income and perceived value of time) produce results very close to a global optimum in the choice of UAM vehicles to acquire. On the other side, even small changes in the availability of data about market distribution can cause suboptimal decisions. To alleviate the impact of uncertainty and to be able to analyze many scenarios, so as to support prescriptive analytics for decision-making, we propose the use of Neural Networks, that can be trained to provide insight into the dependency of variables of interest (in this case, acquisition decisions) on multiple inputs (in this case, for example, desired gain margin, available market fraction, and maximum allowed number of flights per day). This use of various Machine Learning techniques provides a first step into understanding the reasons for observed outcome, and therefore a step towards robust decision-making in the presence of uncertainty in a SoS. However, we also propose various refinements for future work. First, in this example we trained a Neural Network with a subset of the inputs. The Neural Network is trained as a whole and, other than implementing some basic regularization, does not quantify the impact that each input variable has on the outputs and the variability of these input-output relationships. Therefore, the use of Uncertainty Quantification can greatly improve the approach to these problems, by providing quantitative measurements of the importance that each input variable has on the outputs. This, in turn, will provide information on critical areas where more information (or more caution) is needed. Further research also includes extension of the case studies, to include non-recurrent and recurrent costs, and to use stochastic optimization techniques as additional prescriptive methodologies in support of acquisition decision-making.

References

Apte, U., Rendon, R., & Dixon, M. (2016). *Big data analysis of contractor performance information for services acquisition in DoD: A proof of concept*. Naval Postgraduate School Monterey United States.

Brotherton, T., & Johnson, T. (2001, March). Anomaly detection for advanced military aircraft using neural networks. In *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)*, 6, 3113–3123.

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, *91*(12), 64–72.

Freeman, C. E. (2013). *Multivariate and Naïve Bayes text classification approach to cost growth risk in Department of Defense acquisition programs*.

Huang, C., & Mintz, A. (1990). Ridge regression analysis of the defence-growth tradeoff in the United States. *Defence and Peace Economics*, *2*(1), 29–37.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *It Professional*, *15*(6), 43–48.

Lo, A. W., & Hasanhodzic, J. (2010). *The evolution of technical analysis: financial prediction from Babylonian tablets to Bloomberg terminals* (Vol. 96). John Wiley & Sons.

Lynch, P. (2008). The ENIAC forecasts: A re-creation. *Bulletin of the American Meteorological Society*, *89*(1), 45–56.

Maheshwari, A., Sells, B. E., Harrington, S., DeLaurentis, D., & Crossley, W. (2021). Evaluating impact of operational limits by estimating potential uam trips in an urban area. In *AIAA Aviation 2021 Forum,* 3174.

Maier, M. W. (1998). Architecting principles for systems-of-systems. *Systems Engineering: The Journal of the International Council on Systems Engineering*, *1*(4), 267–284.

Moore, G. W., & White III, E. D. (2005). A regression approach for estimating procurement cost. *Journal of Public Procurement*.

Raz, A., Balasubramani, P., Harrington, S., Guariniello, C., & DeLaurentis, D. A. (2020). *System-of-systems acquisition analytics using machine learning techniques*.

Rehman, M. H. ur, Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*(6), 917–928.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77–84.

Wang, R., & Yang, W. (2016). Estimation of demand and supply in U.S. airline industry. *Unpublished Working Paper*.

Wei, K., Wu, J., Ma, W., & Li, H. (2019). State of charge prediction for UAVs based on support vector machine. *The Journal of Engineering*,*2019*(23), 9133–9136.

Xiao, H., Cai, C., & Chen, Y. (2006). Military vehicle classification via acoustic and seismic signals using statistical learning methods. *International Journal of Modern Physics C*,*17*(02), 197–212.

Zainol, Z., Azahari, A. M., Wani, S., Marzukhi, S., Nohuddin, P. N., & Zakaria, O. (2018). Visualizing military explicit knowledge using document clustering techniques. *International Journal of Academic Research in Business and Social Sciences*, *8*(6), 1127–1143.

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Acquisition Research Program
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net