SYM-AM-20-063

PROCEEDINGS
OF THE
SEVENTEENTH ANNUAL
ACQUISITION RESEARCH SYMPOSIUM

Acquisition Research:
Creating Synergy for Informed Change

May 13–14, 2020

Published: April 13, 2020

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

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

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Analysis of Contractor Data in Federal Acquisition Databases

Ningning Wu—is Professor of Information Science at the University of Arkansas at Little Rock. She received a BS and an MS degree in electrical engineering from the University of Science and Technology of China and a PhD in information technology from George Mason University. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

M. Eduard Tudoreanu—is Professor of Information Science at University of Arkansas at Little Rock. Professor Tudoreanu has expertise in human-computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis and has extensive experience in software development and user interface design. Professor Tudoreanu was the founding Technical Director of the Emerging Analytics Center. He has been the keynote speaker at ABSEL 2010 and has served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in computer science in 2002 from Washington University in St. Louis. [metudoreanu@ualr.edu]

Sanketh Siruvolu—is an enthusiastic graduate student who always has a zeal for learning. He has a bachelor's degree in computer science. He worked at Amazon, where he started learning to analyze data and learned how deeply it is having an impact in the current world of technology, where almost everything is becoming dependent on data. His work on researching data, finding credible sources for the data, and analyzing data to give estimates on the given input has been an amazing experience and a great opportunity to learn data analytical skills. [ssiruvolu@ualr.edu]

Richard Wang—is Director of the Massachusetts Institute of Technology (MIT) Chief Data Officer and Information Quality Program. He is also the Executive Director of the Institute for Chief Data Officers (iCDO) and Professor at the University of Arkansas at Little Rock. From 2009 to 2011, Wang served as the Deputy Chief Data Officer and Chief Data Quality Officer of the U.S. Army. He received his PhD in information technology from the MIT Sloan School of Management in 1985. [rwang@mit.edu]

John R. Talburt—is the Acxiom Chair of Information Quality and Director of the Information Quality Graduate Program at the University of Arkansas at Little Rock. Previously, he led the Data Research and Development group for Acxiom Corporation, where he implemented their Total Data Quality Management program. He is currently serving as a member of the U.S. Technical Advisory Group (TAG) to the International Organization for Standardization (ISO) in the area of data quality. He is an inventor for several patents related to customer data integration and the author of numerous research papers and articles on information quality and entity resolution. [jrtalburt@ualr.edu]

Abstract

The publicly accessible acquisition databases at www.usaspending.gov and www.fpds.gov contain rich contract information from all U.S. departments for the last two decades. Hidden in these huge data repositories may be patterns that are unknown yet potentially useful for acquisition research and practice. This proposal is a continuation of the last two years of research that were included in the Acquisition Research Program Symposium. It aims to utilize information from the web to identify the distribution of federal contractors in areas at high risk of natural disasters. The work is part of an ongoing research project whose ultimate goal is to develop a risk assessment framework that can help acquisition decision-makers assess potential risks to a supply chain or project and mitigate such risks.

Key words: data analytics, risk assessment, natural disasters

Introduction

The vast quantity of online information provides great opportunities to harvest and enrich our data and knowledge. Indeed, searching for information on the web has become a common practice, thanks to the data explosion and the advancement of search engines and web technologies. If properly utilized, online information may help not only assess and improve the quality of the data, but also crystalize the needed information.

The paper presents techniques for leveraging the information available in the public domain to complement internal acquisition and purchasing data in order to support effective acquisition planning and management. Acquisition planning and management is a complex system engineering task that involves many decision-making and action-taking processes. It covers a complex environment including actual acquisition, contracting, and fiscal, legal, personnel, and regulatory requirements. Many factors need to be taken into account to ensure success in a project's planning and execution. This paper focuses on natural disasters and their potential impacts to acquisition projects. It aims to provide a tool for program managers to assess and control the risks of natural disasters. By utilizing the information from authoritative sources such as National Centers for Environmental Information (NCEI, www.ncei.noaa.gov) and Federal Emergency Management Agency (FEMA, www.fema.gov), this research describes methods of identifying areas with high risk of natural disasters as well as the distribution of U.S. industry in those areas. As a typical acquisition project might involve hundreds, if not thousands, of contractors and subcontractors located in different geographical areas, the information on high-risk areas can help program managers assess the possible impacts of natural disasters to a project and its contractors and can help mitigate disruptions to mission-critical projects by selecting appropriate redundancies from industry located in low-risk areas.

This research is a component of an ongoing project that aims to develop a risk assessment framework to help acquisition decision-makers assess potential risks of a project. It focuses on identifying the areas with high risks of natural disasters and the industry types that are highly concentrated in those areas. The information on high-risk areas would help program managers select right contractors for a project to reduce the risk factors that might contribute to failure.

The work will focus on two factors: natural disasters and types of business that are mainly located in risk areas. The type of business is identified by North American Industry Classification System (NAICS, 2017) code. A NAICS code can be attached to many products and many companies. However, if a NAICS code has only a few companies associated with it, then it can be considered as a high-risk business type, because if one of these companies failed, it would be difficult to find an alternative source.

The proposed research is composed of two tasks. First, it determines salient techniques to classify U.S. counties by their natural disaster risk levels. It will integrate NCEI and FEMA data to evaluate the severity of a disaster and its impact to the affected areas. Such a classification system is helpful to acquisition management for risk assessment. Second, it analyzes the distribution of federal contractors and different types of industries (identified by NAICS code) in various risk areas. As a case study, we apply the analysis of the percentage of high-risk contractors to past project data.

Related Work

Previously, policy makers and researchers have recognized the need to employ data as a multifaceted means of increasing the agility of the acquisition process (Krzysko & Baney, 2012). To this end, research has looked at automatic means of dealing with the heterogeneous acquisition data sources from text processing (Zhao et al., 2015), systems engineering (Cilli,

Parnell, Cloutier, & Zigh, 2015), and business (Gaither, 2014) perspectives. Our paper is different both in content and in the approach—in content in that we rely on big data to identify hidden risk factors, and in the approach in that our expertise in information visualization, data quality, data governance, policy (chief data officers), and data science provides a value-based perspective.

Tudoreanu, Franklin, Rego, Wu, and Wang (2018) investigated employment data in an attempt to correlate changes in employment with negative modifications to contracts. Such correlations can be exploited to infer hidden and undisclosed contractors that are part of the defense acquisition network. Hidden contractors may pose the risk of becoming a weak stress point of a project and would affect the overall outcome of the project.

Wu, Wang, and Tudoreanu (2018) proposed a framework based on a data science approach that aims to utilize the online information to assess and improve acquisition database quality as well as to find the hidden patterns to further acquisition research. The main component of the framework is a web search and text mining module, whose main function is to search the internet and identify the most credible and accurate information online.

Apte, Rendon, and Dixon (2015) explored the use of big data analytics techniques to explore and analyze large data sets that are used to capture information about Department of Defense services acquisitions. The paper described how big data analytics could potentially be used in acquisition research. As the proof of concept, the paper tested the application of big data analytic techniques by applying them to a data set of Contractor Performance Assessment Report System (CPARS) ratings of 715 acquired services. It also created predictive models to explore the causes of failed services contracts. Since the data set used in the research was rather small and far from the scope of big data, the techniques explored by the paper mainly focus on traditional data mining techniques without taking into account big data properties.

Black, Henley, and Clute (2014) studied the quality of narratives in CPARS and their value to the acquisition process. The research used statistical analysis to examine 715 Army service contractor performance reports in CPARS in order to understand three major questions: (1) To what degree are government contracting professionals submitting to CPARS contractor performance narratives in accordance with the guidelines provided in the CPARS user's manual? (2) What is the added value of the contractor performance narratives beyond the value of the objective scores for performance? (3) What is the statistical relationship between the sentiment contained in the narratives and the objective scores for contractor evaluations?

Research Results

Our study focuses on federal purchasing data hosted by www.usaspending.gov and www.fpds.gov. To allow more flexibility in the choice of analytics tools and more efficiency for data processing, we set up a database server and stored the data downloaded from both sites. For natural disaster information, we downloaded the weather data of all U.S. counties between the years 1950 and 2018 from www.ncei.noaa.gov and the natural disaster assistance data from www.fema.gov covering 1953 to 2020. This paper takes a data-centric approach and is composed of two tasks: (1) produce a natural disaster risk map for U.S. counties, and (2) examine the distribution of federal contractors and different business types in high-risk areas.

Natural Disaster Risk Map for U.S. Counties

National Centers for Environmental Information (NCEI) has been collecting natural disaster data since 1950 and has that information available for each U.S. county. The data covers a wide range of natural disasters, including flood, tornado, hurricane, blizzard, high wind, flash flood, hail, dust storm, etc. Our previous analysis on NCEI data revealed that it is

challenging to categorize a disaster by its intensity and damage level because disasters lack such a categorization system. Even though some disaster types, such as tornadoes and hurricanes, do have a categorization system, it is often difficult to assess an incident's impact to the local communities without other supporting information. Furthermore, a comparison of impact of disasters of different types is not easy to perform.

One objective of this research was to identify U.S. areas with a high risk of natural disasters. To assess an area's risk level, we considered both the number of disasters and the impact of those disasters. For example, an area might have experienced several minor natural disasters during the period we studied, but none of them were serious, while another area has fewer incidents, but some of them were serious and had a serious effect on local communities and economy. For the purpose of acquisition risk management, the second area in our example should be considered as a high-risk zone.

To gain a better understanding of a disaster in terms of its intensity and impact, NCEI data is enhanced with FEMA disaster mitigation and recovery data. Table 1 lists the FEMA data fields:

Table 1: FEMA Data Fields

1	femaDeclarationString	**13**	incidentBeginDate
2	disasterNumber	**14**	incidentEndDate
3	state	**15**	disasterCloseoutDate
4	declarationType	**16**	fipsStateCode
5	declarationDate	**17**	fipsCountyCode
6	fyDeclared	**18**	placeCode
7	incidentType	**19**	designatedArea
8	declarationTitle	**20**	declarationRequestNumber
9	ihProgramDeclared	**21**	hash
10	iaProgramDeclared	**22**	lastRefresh
11	paProgramDeclared	**23**	id
12	hmProgramDeclared		

FEMA data shows the beginning and ending dates of an incident, its location information, and the FEMA assistance program declared. The types of assistance programs actually provide a good indicator on the damage level and scope of an incident. Below are their short descriptions:

- ihProgramDeclared: denotes whether the Individuals and Households program was declared for this disaster.
- iaPorgramDeclared: denotes whether the Individual Assistance program was declared for this disaster
- paProgramDeclared: denotes whether the Public Assistance program was declared for this disaster
- hmProgramDeclared: denotes whether the Hazard Mitigation program was declared for this disaster.

Among the four assistance programs, ihProgram is the highest level of assistance and aims to help communities that are significantly affected by a major disaster, and hmProgram is

the lowest. This paper proposed to use the declared assistance programs to assess a natural disaster's intensity and damage level. More specifically, given the number of disasters in an area during a period and the corresponding number of different assistance programs declared, a weighted sum of disaster number, hence termed **weighted disaster score** (WDS), is calculated as follows:

$$s = \sum_{i=1}^{4} w_i \times n_i$$

Where n_i is the number of a specific type of assistance programs, and w_i is the corresponding weight for the type. The weight for each assistance program is defined as follows:

- Disaster mitigation: 0.25
- Public assistance : 0.50
- Housing assistance: 0.75
- Individual assistance: 1.0

Table 2 shows the five-number summary (i.e., minimum, first quartile [Q1], medium, third quartile [Q3], and maximum values) for the WDS scores and the number of disasters of all U.S. counties between 1953 and 2020, respectively.

Table 2: Five-number Summary of WDS and Number of Disasters of FEMA Data

Statistic	WDS	# of Disasters
Min	0.25	1
Q1	6.25	10
Medium	10	15
Q3	13	19
Max	56.25	105

To facilitate comprehension for a wide range of domain experts and program officers, we define three risk levels, namely low, medium, and high. The first quartile and the third quartile of WDS are used as the cutoff points for the risk classes, as shown in Table 3. A natural disaster risk class is assigned to each county based on its WDS value. Counties with a WDS value less than 6.25 are considered to have low risk of natural disasters, counties with a WDS value in between 6.25 and 13 are considered to have medium risk, and counties with a WDS value greater than 13 are considered to have high risk.

Table 3: Risk Level Categorization

Level	WDN Values
Low	$wdn < 1^{st}$ quartile
Medium	1^{st} quartile $< wdn < 3^{rd}$ quartile
High	$wdn > 3^{rd}$ quartile

Figure 1 shows the distribution of U.S. counties by risk levels. It shows that about 28% of counties are located in a high-risk area, 52% in medium risk, and 20% in low risk. Figure 2 shows the disaster risk class on a geographical map of the U.S.

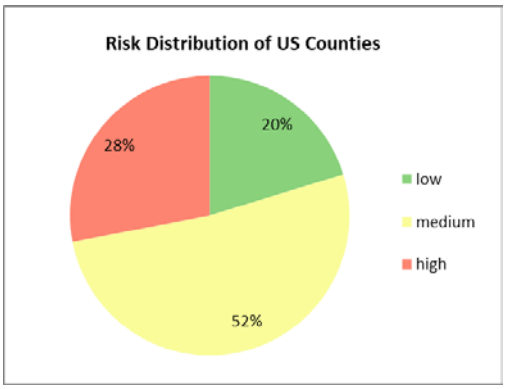


Figure 1: County Distribution by Risk Levels

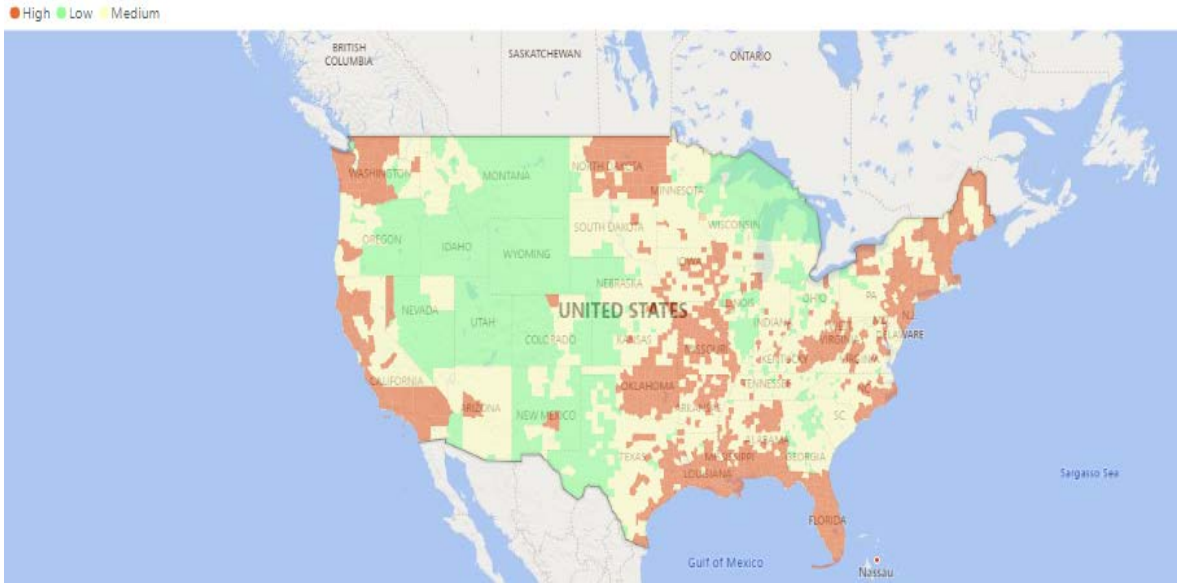
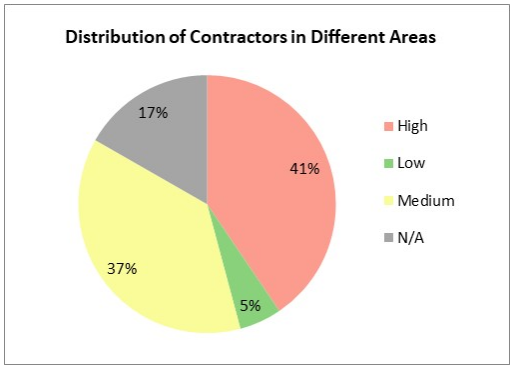


Figure 2: Natural Disaster Risk Class (Low, Medium, High) Displayed for Each County in the United States

The index takes into count both the number of occurrences of disasters and their magnitude. Red encodes high risk, yellow medium, and green low.

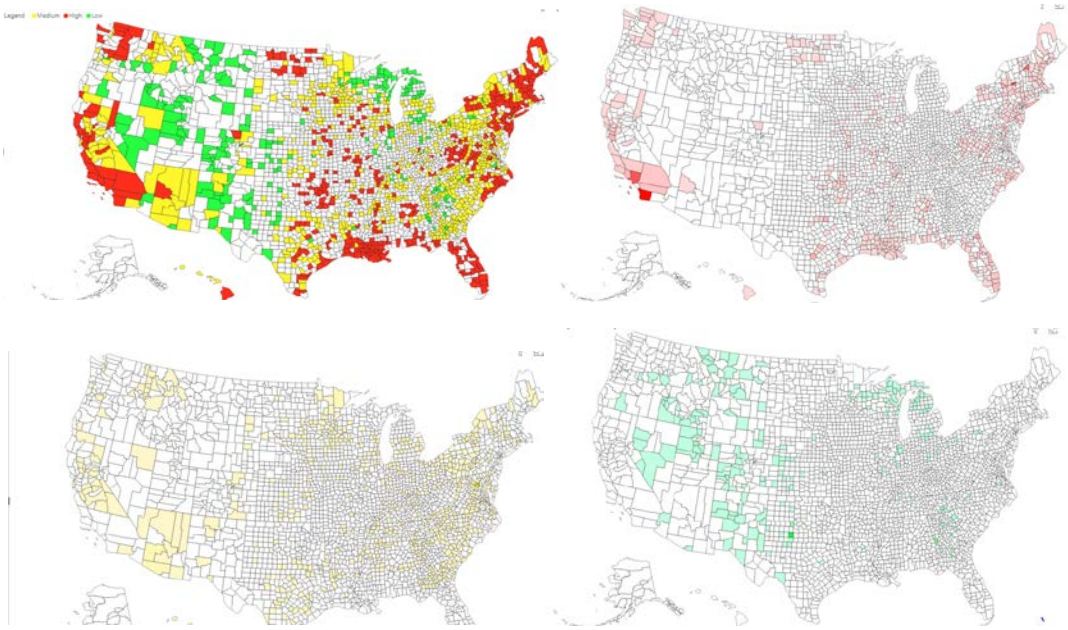
Distribution of Federal Contractors and Business in Different Risk Areas

Not all counties in the United States have federal contractors or business being conducted for a federal project, and the focus of this subsection is to investigate how various types of Navy industries relate to disaster areas. To this end, we consider information from www.fpds.gov and www.usaspend.gov and join it with the data from NCEI and FEMA. Figure 3 shows the distribution of federal contractors, divided by class of natural disaster risk. It shows as high as 41% of contractors are located in high-risk areas. There are about 17% of contractors that are not located in the United States, so their natural disaster risk levels are not assessed by this research. Figure 4 provides somewhat of a reverse analysis, where each county in the United States is mapped based on the risk type and number of awards received from the Navy. A map such as this can be made interactive and computed in real-time to show the distribution of any acquisition project and help a domain expert plan or run the project.



Not all business takes place in the United States, and thus the risk level is not known for 17%.

Figure 3: Contractors' Distribution by Risk Levels



Red is high risk, yellow is medium, and green is low. Top-left show the overall view of counties with at least one award. Top-right shows only counties with high risk, using color intensity to encode the number of awards in that county. Bottom-left depicts counties with medium risk level, using color intensity for number of awards. Similarly, bottom-right shows low-risk counties and number of awards.

Figure 4: Place of Performance of Navy Awards Correlated to Natural Disaster Risk

North American Industry Classification System (NAICS; Office of Management and Budget, 2017) is the standard used by federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. NAICS code describes the business specialization of a company. There are 379 distinct NAICS codes among all contractors in our database copy of the federal purchasing information, and 355 of them have contractors located in natural disaster risk areas.

Generally, a specific type of industry would be robust and not susceptible to natural disasters if most of the companies providing services and products in that NAICS code are located in low-risk areas. Figure 5 shows the clusters of NAICS codes based on the percentage of companies located in high-risk areas. Fifty-two of the 355 NAICS codes have all companies doing business with the Navy in high-risk areas. A closer look at these 52 NAICS codes reveals that the majority of them have only one contractor. Thus, companies of these NAICS codes can increase the system risk of any acquisition project in which they participate.

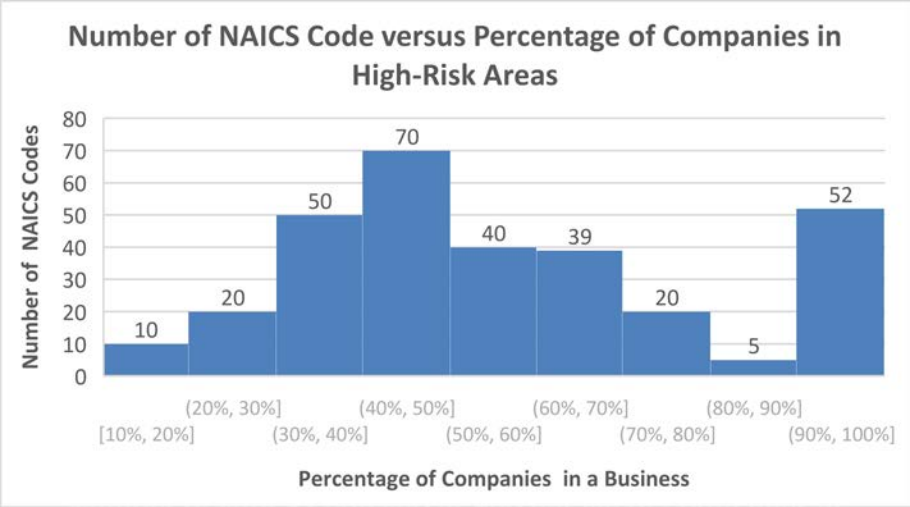


Figure 5: Clustering NAICS Codes by Percentage of High-risk Companies

Distribution of High-risk Contractors by Project

This section analyzes the distribution of high-risk contractors, as defined by their NAICS code and location, in past federal projects. Considering data taken from www.usaspending.gov, there are about 13,435 distinct projects with subcontractor information. The percentage of high-risk contractors for each project is calculated, and it reveals that more than half of the projects have high-risk contractors. A closer look at these projects shows that 90% are single contractor projects, and over 97% have no more three total contractors. The maximum number of contractors in these projects is 35. Figure 6 shows the number of projects in each percentage bin. It shows that, except for the projects in the last bin, the majority of the remaining projects have between 30% to 50% high-risk contractors.

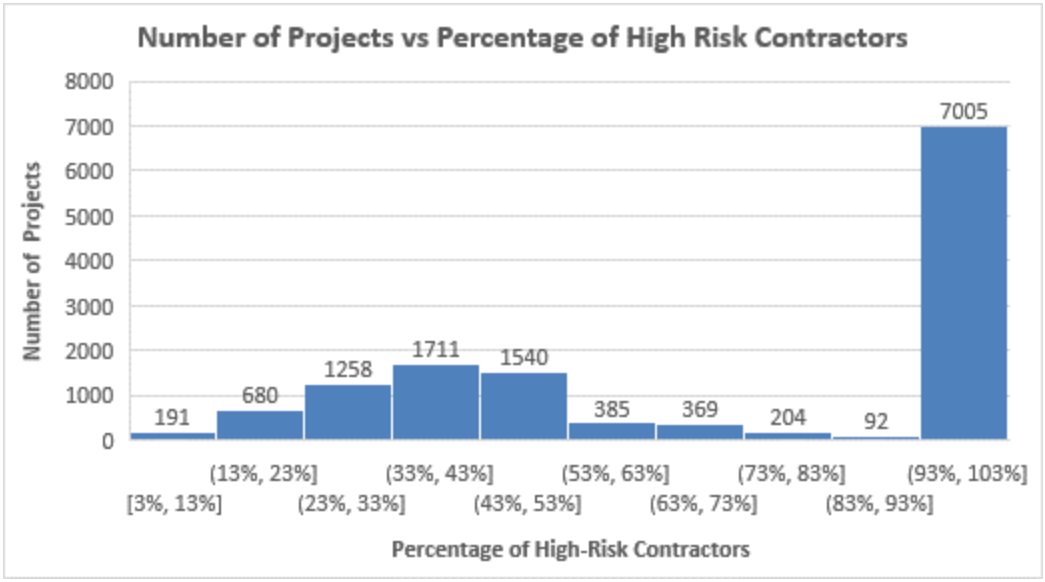


Figure 6: Clustering of Projects by Percentage of High-Risk Contractors

Because the number of contractors in a project varies, we analyze the distribution of number of contractors and present the results in Figure 7, which shows the scatter plot of the

total number of contractors in a project and the percentage of the high-risk ones. It shows that the majority of projects have fewer than 250 contractors. There are several projects with more than 1,000 contractors. Figures 8(a) and 8(b) partition the projects into two groups, the one with more than 100 contractors and the one with fewer than 100 contractors, and it shows the percentage of high-risk contractors for the projects in each group.

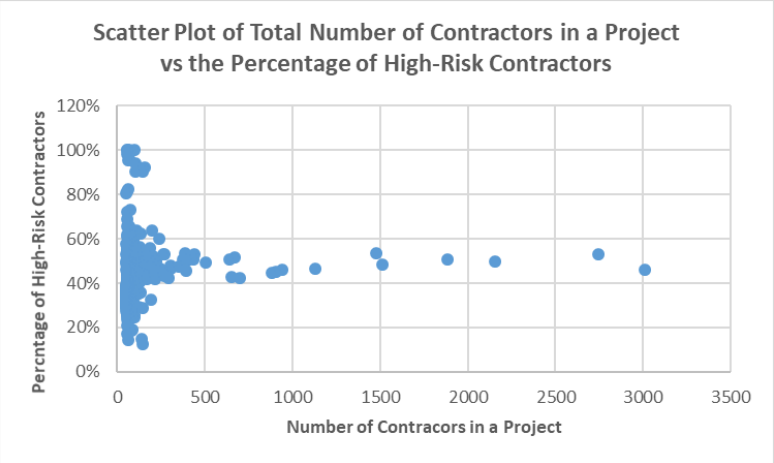


Figure 7: Percentage of High-risk Contractors versus Total Number of Contractors

(a) Fewer than 100 contractors per project, (b) Greater than 100 contractors per project

Figure 8: Percentage of High-Risk Contractors versus Total Number of Contractors

Conclusion and Future Work

The paper presented a framework for classifying U.S. counties by their risk levels of natural disasters. The framework takes into account both the frequency and severity of disasters by leveraging the weather-related information form authoritative websites including www.ncei.noaa.gov and www.fema.gov. Based on this framework, the paper identified federal contractors located in high-risk areas as well as U.S. industries that are highly concentrated in those areas. The research results would be helpful for the acquisition management and planning to control the risks of natural disasters and their impacts to a project.

Our future work will focus on two directions. First, investigate other risk factors in order to develop a comprehensive, data-driven risk assessment framework that can be applied quickly to both large and small acquisition and purchasing projects. Second, we will research on appropriate visualization methods to present information to acquisition professionals and

decision-makers that is clear, intuitive, timely, and comprehensive, thus improving the quality of the available information.

References

Apte, U., Rendon, R., & Dixon, M. (2016). Big data analysis of contractor performance information for service acquisition in DoD: A proof of concept. In *Proceedings of the 13th Annual Acquisition Research Symposium* (Vol. 2, pp. 398–418). Retrieved from https://calhoun.nps.edu/handle/10945/53512

Black, S., Henley, J., & Clute, M. (2014). *Determining the value of Contractor Performance Assessment Reporting System (CPARS) narratives for the acquisition process* (Master's thesis, Naval Postgraduate School). Retrieved from https://calhoun.nps.edu/handle/10945/42583

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised Defense Acquisition System. *Systems Engineering, 18*(6), 584–603.

Gaither, C. C. (2014). Incorporating market-based decision-making processes in defense acquisitions. *International Journal of Defense Acquisition Management, 6*, 38–50.

Krzysko, M., & Baney, B. (2012). The need for "acquisition visibility." *Journal of Software Technology, 15*(1), 4–9.

Office of Management and Budget. (2017). *North American Industry Classification System: United States, 2017.* Retrieved from https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf

Tudoreanu, M. E., Franklin, K., Rego, A., Wu, N., & Wang, R. (2018). Searching hidden links: Inferring undisclosed subcontractors from public contract records and employment data. In *Proceedings of the 15th Annual Acquisition Research Symposium* (Vol. 2, pp. 491–499). Retrieved from https://calhoun.nps.edu/handle/10945/58788

Wu, N., Wang, R., & Tudoreanu, M. E. (2018). Utilizing public data for data enhancement and analysis of federal acquisition data. In *Proceedings of the 15th Annual Acquisition Research Symposium* (Vol. 2, pp. 94–113). Retrieved from https://calhoun.nps.edu/handle/10945/58763

Zhao, D., MacKinnon, D., Douglas, J., Gallup, D., & Shelley, P. (2015). *Lexical link analysis (LLA) application: Improving web service to Defense Acquisition Visibility Environment* (Acquisition Research Program Sponsored Report Series, Naval Postgraduate School). Retrieved from https://www.semanticscholar.org/paper/Lexical-Link-Analysis-(LLA)-Application%3A-Improving-Zhao-MacKinnon/4bbc796cd335b1aac5ab3975d686e893658cc7ad

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net