Wildfire Emergency Response Hazard Extraction and Analysis of Trends (HEAT) through Natural Language Processing and Time Series

Sequoia R. Andrade
HX5 LLC.
Under Prime Contract No. 80ARC020D0010
with the NASA Ames Research Center,
Moffett Field, United States
sequoia.r.andrade@nasa.gov

Hannah S. Walsh
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, United States
hannah.s.walsh@nasa.gov

***Abstract*—A methodology for Hazard Extraction and Analysis of Trends (HEAT) is proposed and conducted on a data set of wildfire incident response forms, known as ICS-209-PLUS. The HEAT processes: (1) extract a set of hazards from a data set, (2) calculate hazard-relevant metrics in a primary analysis, (3) analyze trends over time in metrics using time series, and (4) examine potential explanations for metric trends using a secondary analysis. Hazards are extracted from narrative data in the ICS-209-PLUS based on a framework previously developed by the authors, using natural language processing. Metrics examined for each hazard include operational time to occurrence, rate of occurrence, frequency, and severity. Primary results include a taxonomy of hazards present in the data set with relevant quantitative metrics. The most frequent hazards identified are environmental and include hazardous terrain. Most hazards occur on average between 35-55% containment. Incidents with hazards tend to have a higher average severity score when compared to the average score for all incidents. Time series of the metrics and relevant predictors, including fire characteristics, fire intensity, and operations, are created to facilitate further analysis. Secondary results used to determine which factors best predict hazard frequency include a correlation matrix and regression analysis. These findings are relevant to safety for current, as well as emerging wildfire operations, and are an exploratory first step in developing historical data-driven risk assessment models.**

***Index Terms*—hazards, risk, safety, systems, emergency response, wildfire, UAS, time series analysis, NLP, topic modeling, text mining, correlation, trend analysis**

I. Introduction

In recent years, destruction caused by wildfire has intensified in specific regions, such as warm, dry regions and coniferous forests in California [1]. Significant research has been conducted analyzing the trends in the fires themselves, including the finding that regions with trends toward hotter temperature and less precipitation experience greater increases in fire activity [2]. According to NASA's Earth Observatory, since 1950 trends indicate there are more wildfires and these wildfires have grown in size. Only 11% of the western U.S. has burned, with the same areas burning repeatedly [3]. Existing research highlights the impact wildfire will increasingly have on the Wildland Urban Interface, which includes effects on public infrastructure, damage to residential communities, and harm to agricultural resources [4]. While there is a substantial body of research on longitudinal trends in wildfire characteristics, there is less research on longitudinal trends in wildfire response operations, specifically suppression efforts. In addition to research on operations trends in general, specific analysis of hazards during and from operations can provide insight on safety gains possible from new technology.

Operators and safety assurance specialists need information on existing system risk, which can be gained from examining historical data. While anecdotal or experience-based assessments are valuable, they may be biased and may fail to capture the complete set of patterns present in historical data. As often noted, manual human analysis of large data sets is tedious and limiting. In contrast, a large-scale study of incident reports leverages the vast historical knowledge that is available and presents the results with less human bias. However, large-scale studies of the hazards associated with wildfire response and their correlation with specific operational characteristics have been, to date, limited. Identifying what characteristics are related to hazards, when hazards are most likely to occur, and how often, can provide insight for potential risk mitigation. Natural language processing techniques, as well as the availability of wildfire incident reports, has made possible a large-scale analysis of wildfire hazards and trends. The recently released ICS-209-PLUS database, a set of wildfire incident response form ICS-209, consists of reports on wildfire operations with qualitative narratives alongside quantitative data [5]. Incident commanders use this form to document how an incident unfolds and summarize the results of the incident. The portion of the database used in this research spans from 2006 to 2014, and describes fires that are sufficiently large and/or complex to necessitate an incident commander.

In this research, a methodology for Hazard Extraction and Analysis of Trends (HEAT) is proposed and conducted on the ICS-209-PLUS database [5]. The HEAT processes: (1) extract hazards from a data set, (2) calculate metrics for each hazard, (3) analyze longitudinal trends in metrics, and (4) examine potential explanations for metric trends. Hazards are extracted from the narrative data using natural language processing. Operational time to occurrence, frequency, rate of occurrence, and severity metrics are calculated for each hazard and compared to understand which hazards pose the greatest risk at what times of operation. Next, using time series, a longitudinal analysis of trends in the metrics is performed to examine how occurrences of the hazards have changed from 2006 to 2014. Finally, to examine potential causes of hazard trends, trends in operations, intensity, and fire characteristics are also analyzed using time series. Fire characteristics include the annual number of acres burned, days burning, and frequency of fires, while intensity includes number of injuries, fatalities, damaged structures, and destroyed structures. Operational trends include the annual number of aerial assets, personnel, and cost. The relationship between trends in hazard frequency, operations, intensity, and fire characteristics are examined through a correlation matrix. Linear regression is performed to determine if trends in hazards can be explained by trends in operations, intensity, and/or fire characteristics. HEAT enables the discovery of trends and risks within wildfire response and can be used to better understand how to improve the safety of such systems.

II. Wildfire Response

Emerging operational concepts for wildfire response, such as NASA's Scalable Traffic for Emergency Response Operations (STEReO) [6], include novel technologies and are both safety-critical and complex. Assessing the risk of emerging operations as compared to conventional operations is a necessary prerequisite to introducing these new technologies. To understand the context of the hazards and results produced in this work, present and emerging wildfire response operations are described in the following sections.

A. Present Operations

Wildfire response involves the coordination of multiple assets, such as ground crews and aircraft, with the goal of containing a wildfire and thereby minimizing losses of property and other resources [7]. Large and/or complex response operations are managed according to the Incident Command System (ICS), a system designed for various types and sizes of incidents [7]. The ICS was developed with inter-agency operations in mind and focuses on command, operations, finance, planning, and logistics [7]. Activities are centered around the construction of a fireline, which contains the fire. The operation is complex, with assets often operated by different organizations, such as multiple fire departments for large fires. The number of assets scales depending on the size of the fire, with large wildfires often requiring hundreds of personnel. Smaller fires may be managed using only ground crews, whereas large or inaccessible fires may require tanker aircraft to be flown in from miles away. As a fire grows, the number of assets can scale dynamically, up to what is known as a Type 2 or Type 1 attack in the ICS [8]. In fires large enough to necessitate aerial support, aircraft perform various tasks to support ground crews building a fireline. Retardant drops slow the spread of the fire, while water drops cool the fire to allow ground crews to safely work nearby. Aerial surveillance missions are also necessary to provide updated information, including fire location, to the commander overseeing the operation.

Aircraft may fill one of several different roles. Aerial supervisors coordinate the aerial assets being applied to the wildfire event [9]. They additionally relay weather information to operators. The airspace coordinator is in charge of managing potential conflicts in the airspace, including with the temporary flight restriction (TFR) zone around the incident. Lead planes direct tankers to drop locations, indicating the location by releasing white smoke. These tankers may drop either water or retardant on the fire. There are several sizes and configurations of air tankers. Specialized scooper aircraft refill without landing. Finally, helicopters may perform bucket drops, cargo and personnel transport, and surveillance. Details are summarized in Table I. Table II additionally includes hazards for the operation.

B. Emerging Operations

In emerging operations, there are several opportunities to leverage technology to improve performance as well as safety of wildfire response. Unmanned Aerial Systems (UAS) provide a potentially cost-effective way of (1) reallocating dangerous tasks to autonomous systems and (2) improving performance [7]. NASA's UAS Traffic Management, or UTM [14], has strong potential for better coordination of UAS with conventional aircraft in wildfire response, as well as the ability to improve communications and data relay [7]. NASA's Scalable Traffic Management for Emergency Response Operations (STEReO) investigates the application of NASA technologies, including UAS and UTM, to wildfire response and other emergency response applications. There are several emerging technologies developed for STEReO that will enable UAS to maintain a high standard of safety and performance. Specific technologies include Safe2Ditch [15], Independent Configurable Architecture for Reliable Operations of Autonomous Systems (ICAROUS) [16], Safeguard [17], Radar on Autonomous Aircraft to Verify ICAROUS Navigation (RAAVIN) [18], and Detect and Avoid iN The Cockpit (DANTi) [19]. These technologies are designed to mitigate hazards specific to UAS, regardless of operational scenario. Wildfire response is a safety critical system that may exhibit hazards relevant to UAS that are not addressed by these systems. Hazards discovered by HEAT must be considered in the context of these emerging technologies to understand the potential additional safety and performance considerations, as these new operational concepts have numerous expected benefits.

TABLE I

OVERVIEW OF AIRCRAFT TYPES, MODELS, AND OTHER INFORMATION FOR AIRCRAFT USED IN WILDFIRE RESPONSE [7], [9]–[13]

Category	**Sub-Category**	**Example Aircraft Type or Model**	**Selected Avionics**
Aerial Supervision	Air Tactical Ground Supervisor (ATGS) Aerial Supervision Module (ASM) Lead Plane	Typically, twin-engine fixed wing aircraft	Traffic collision avoidance system (TCAS/TCAD) Headsets for each person Boom microphone on headsets Audio panels for each person Voice activated intercom Volume and squelch controls for each person
	Helicopter Coordinator (HLCO)	Typically, a helicopter	
Airtanker	VLAT	DC-10, B-747	Automatic Dependent Surveillance –Broadcast (ADS-B) Global Positioning System (GPS) using datum NAD27, NAD83, and WGS84 Traffic collision avoidance system (TCAS/TCAD)
	Type 1	B-737	
	Type 2	Convair 580, Q400	
	Type 3	Air Tractor AT-802 F	
	Type 4	Air Tractor AT-802/602	
	Water Scooper	CL215/415	
Helicopter	Type 1	Sikorsky S-64E	Aircraft VHF-AM/FM Radio Global Positioning System (GPS) Paper or electronic map of area
	Type 2	Bell B-214	
	Type 3	Bell B-206 B3	
UAS	Type 1-2	Scan Eagle, Aerosonde, Penguin C	Mode C transponder Communication via radio
	Type 3-4	3DR Solo (RW) and FireFly6 (FW)	

TABLE II

SELECTED HAZARDS IN THE OVERALL AVIATION OPERATION [7], [9].

Hazard Category	**Selected Hazards**
Aircraft	Avionics failure, Field of view, Engine performance, Fueling errors, Low maneuverability of VLAT
Communication	Radio frequency congestion, Inaccurate information
Environmental	Topography, Visibility, Wind, Smoke, Weather, Turbulence
Human Factors	Fatigue, Task saturation, Risk normalization, Low situational awareness, Lack of pilot skill or familiarity with aircraft
Mission	Low altitude flight, Collision potential, Wake turbulence, Proximity of airstrip/airport to fire

III. EXISTING RESEARCH

A. *Risk Management in Wildfire Response*

Some experts believe advances in risk management and analytics have the potential to improve wildfire management operations [20], [21]. Previous research highlights there is limited data and knowledge about the effectiveness of planning, decision making, and resource allocation during a wildfire response. This lack of data in turn makes it difficult to identify pivot points necessary to make the system more resilient. Even at the current level of knowledge, there are deficiencies evident in the system, including increased hazards and inefficient allocation of resources [20]. Previous research outlining the need for analytics for risk-informed decision making identify three types of analytics: descriptive, predictive, and prescriptive [21]. Descriptive analytics present lessons learned and patterns from past data, predictive analytics use descriptive data coupled with predictive modeling to estimate future risk, and prescriptive analytics suggested recommended actions to reduce risk [21]. HEAT is an algorithm that yields descriptive analytics on the present-day system. The hazards extracted and trends analyzed in this paper provide insight on potential weak points in today's system.

B. *Natural Language Processing for Information Extraction*

This paper utilizes natural language processing to identify and extract hazards from a data set of wildfire incident response forms. Natural language processing (NLP) leverages machine learning capabilities on unstructured text to extract information that would typically be near impossible or time consuming to process by hand. Previous research on the intersection of NLP and emergency response focuses on real-time information using text from social media websites, such as Twitter, rather than text from official reports. Specifically, research conducted uses NLP to classify tweets as disaster or non-disaster related in real-time [22]. Additional research using NLP and classification filters false alarm and true fire event tweets to produce real-time geographic information about fire location [23]. Similar work detects disasters and maps the location using NLP on tweets [24]. Existing research demonstrates NLP applied to extract fire information, yet this methodology is often applied to Twitter instead of official reports and focuses on real-time rather than long term analysis of trends. Use of reports, rather than social media, for text mining provides information useful to engineers and related more to system operations than civilian impacts, though both are important. HEAT focuses on using NLP on official records to gain information on system safety, including operational time to occurrence and frequency of hazards.

Natural language processing has been applied to engineering applications to extract qualitative failure-related information, as well as perform quantitative analysis. Accidents, hazards, and failure events have been extracted from a database of reports using text mining and classification [25]. [26] uses text mining to identify accidents and their causes into a fault tree, then creates a Bayesian network to determine the probability of events. Both [27] and [28] extract mean time between failures (MTBF) from data. [28] extracts reliability information and MTBF directly from web data, rather than calculating the values based on text data, whereas [27] use text mining and lexicons to identify failure-related reviews on electronics prior to using quantitative information from the review, such as date of purchase and time of review, to calculate MTBF for each component. The use of NLP to identify hazards and failures in engineering applications motivates the development of the HEAT algorithm. HEAT is unique from previous research as it is uses official records and analyzes longitudinal trends.

Previous research by the authors [29] developed the hazard extraction methodology used in HEAT. In particular, this work describes a method for using topic modeling on text-based information about incidents to extract failures, their causes, and recommendations. In this research, we extend this framework by analyzing extracted hazards with respect to quantitative data from the same incidents.

IV. Method

A. ICS-209-PLUS Data Set

This research uses the ICS-209-PLUS data set, which contains a complete record of all situation reports, incident summary reports, and complex incident summaries for all incidents [5]. Incidents include multiple types of natural disasters, but the majority (98.5%) of the records are on wildfires [5]. The wildfires in the data set compromise about 1-2% of all wildfires, as they are only the fires large enough to require the incident command system [5]. The raw reports that comprise the data set come from the Fire and Aviation Management Website in the form of PDF scans of documents filled out by hand [5]. St. Denis et al. clean and format each document to produce the ICS-109-PLUS data in the form of CSVs [5]. Since the raw data originates from three distinct reporting eras (1999-2001, 2002-2014, 2014 to present) with different styles, there are differences in quality and included sections in the reports. Discrepancies in data were checked and edited accordingly using other databases, such as FPA-FOD and MTBS [5]. The data set and code used to clean and format the data is open source.

There is a total of 120,803 situation reports and 24,458 summary reports from 1999-2014. Since the mandated use of the reporting system did not occur until 2006 [5], the analysis presented in this work only uses data from 2006-2014. From this time period of interest, there are 53,879 situation reports and 16,800 summary reports. The authors of the ICS-209-PLUS examine key variables related to fire characteristics over regions of the United States, and verifies the trends in the data set match expectations [5] . Rather than focus on the fire information in the data set, this paper utilizes the operations and text information present in the ICS-209-PLUS. Situation reports are used for the extraction of hazards and related metrics, while incident summary reports are used for developing time series for predictor variables of interest. Since this research utilizes both the incident summary and situation reports, only fire incidents with both types of reports are used. Further data cleaning results in a final data set containing 8,991 incident summaries and 44,363 incident reports.

B. Data Cleaning

Since the data was transformed from original handwritten format [5], there are some discrepancies in the ICS-209-PLUS and additional data cleaning is necessary prior to HEAT. In particular, for a small number of of reports there are missing or incorrect values for dates, personnel, and aerial support. The reports use the number of the day in the year, so December 31st is day 365 and January 1st is day 1. When a fire begins towards the end of a year and continues into a new year, dates in the previous year are corrected to negative values to avoid errors. For example, December 31st becomes -1 rather than 365. In some reports, the fire start date is later than the report date. In this case, it is assumed that the dates are transposed and they are corrected accordingly. Occasionally the report date is filled in as the count of days since the fire started. This is corrected to the day of the year by adding the start date. If a single fire has multiple situation reports with different start dates, the earliest start date is assumed the correct one. This is cross checked with the start date for the fire in the summary reports. Some summary reports are missing the start date, which is corrected using the start date in situation reports. When a summary report is missing the containment date, situation reports are used to find the approximate date. If no start date or containment date is found for a a given incident, all reports for the incident are removed from the data set. Similarly when a summary report for an incident has missing values for the total number of personnel and/or aerial assets, these numbers are estimated by summing up the totals from the situation reports for the incident. If this calculation still reports no personnel, then all reports for the incident are removed. Any duplicate reports containing information identical to another report are removed.

Because the data set is so large and cannot reasonably be manually checked, not all discrepancies can be accounted for during cleaning and some outliers are still present after calculations. Outliers are defined as any data point that is a distance 1.5 times the interquartile range from the boundaries of the range .These outliers are safely removed when calculating average operational time to occurrence and severity, resulting in more accurate calculations.

C. Hazard Extraction and Analysis of Trends (HEAT)

There are four stages to the proposed HEAT method: (1) hazard extraction from text-based data, (2) primary analysis, which involves metric calculation from quantitative data, (3) time series analysis of trends in metrics, and (4) secondary analysis, which is an examination of potential explanations for metric trends. The method is summarized in Fig. 1.

1) Hazard Extraction

Hazard extraction is performed after preprocessing the narrative text in the “major problems”, “significant events summary”, and “remarks” sections of situation reports. First, the text from the three sections is combined into one body and standard preprocesssing is performed. Preprocessing procedures include converting all text to lower case, lemmatization, which converts a word such as “running” to its stem form “run”, punctuation removal, and stop word removal. Stop words are common words that do not provide relevant information, and so they are removed from the text. In addition to a standard dictionary of English stop words, a domain specific dictionary is created. The domain specific stop word dictionary

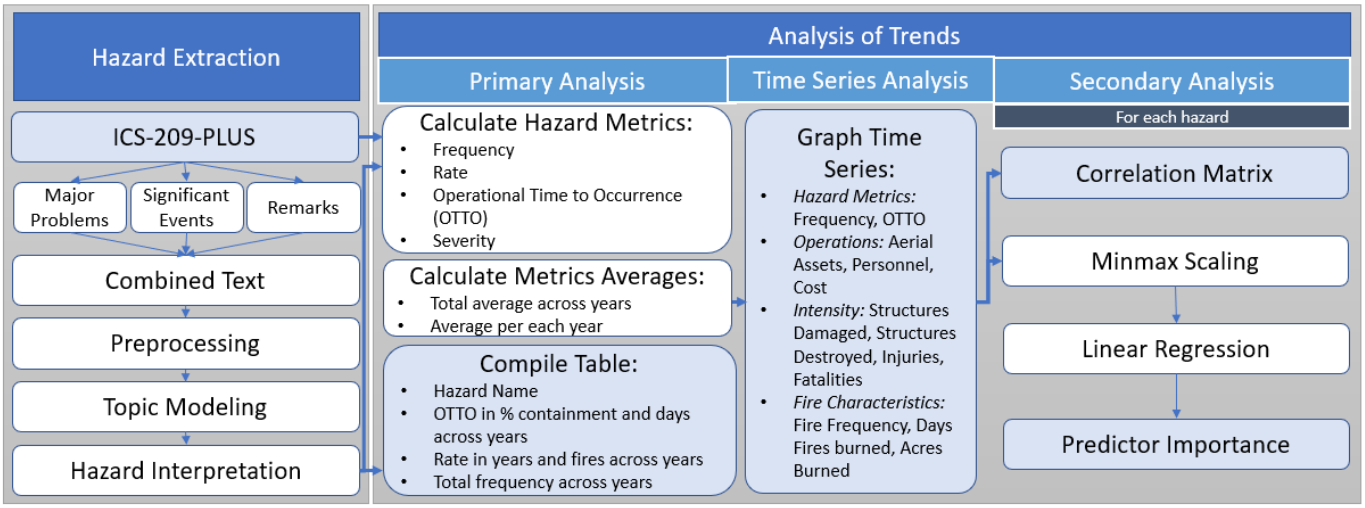

Fig. 1. Overview of HEAT method. Analysis of text-based information is outlined in the Hazard Extraction block, while analysis of quantitative information is outlined in the Analysis of Trends block.

includes the names of all states, counties, national forests, state forests, national parks, national grasslands, and rivers. Additional preprocessing includes removing words that appear in over 50% of documents, as these words likely do not include hazard-specific information. After preprocessing is complete, Latent-Dirichlet Allocation (LDA) [30] topic modeling is performed to identify and group together common themes among narrative reports. LDA assumes each document is composed of a mixture of latent topics, where each topic is composed of a distribution of words. With d documents, the distribution of topics for each document, θ_i with $i \in \{1, 2, .., d\}$, is drawn from a Dirichlet distribution with parameter α. For the K topics, the distribution of words over each topic, ϕ_j with $j \in \{1, 2, .., K\}$ is drawn from a Dirichlet distribution with parameter β. LDA topic modeling outputs the word distribution for the latent topics.

The output of topic modeling requires filtering and interpretation, which is performed to identify specific hazards based on subject-action words Since the database includes documents with no hazards, some topics do not contain hazards and thus filtering is necessary. Interpretation includes examining the topic words with high probability ($p \geq 0.01$) alongside the best documents for each topic, manually identifying hazard relevant words, and aggregating any similar sets of hazard words. To identify which individual reports contain hazards, the specific hazard relevant subject-descriptor word pairs extracted from the raw topic modeling output are used. Documents must contain both a hazard relevant subject and descriptor word in order to denote a hazard occurrence. For example, hazards relating to aerial groundings are identified by reports that contain words related to aircraft, such as *'tanker'* or *'helicopter'*, in addition to words indicating a grounding, such as *'suspend'* or *'unable'*. Once the documents containing hazards are identified, metrics are calculated for each hazard.

2) Primary Analysis

Following hazard extraction, a combination of quantitative and qualitative data is used to calculate metrics for each hazard. Metrics include average operational time to occurrence (OTTO), average rate of occurrence, frequency, and average severity. Operational time to occurrence provides the time frame during which the hazard is most likely to occur, which is vital for targeted risk mitigation. Frequency indicates how many times a hazard occurs, while rate describes how often the hazard is expected to occur. Both frequency and rate are related to likelihood of risk, and severity is a measure of damage. The operational time to occurrence is first calculated for each individual report containing a hazard, then averaged across all reports containing that hazard. For a given report with a hazard, the hazard operational time to occurrence is calculated both in terms of days and percent containment. This is performed to avoid potential skews from large fires lasting many days. The metric measured in days is calculated by subtracting the fire start date from the report date, while the percent containment is used directly from the report, seen in (1). Since multiple reports can be filed per fire, the frequency is calculated as the number of fires during which the hazard occur. This avoids over-counting the frequency of hazards that occur during fires with multiple reports. The rate of occurrence is calculated in two methods in (2), one measuring number of hazards per year and the other measuring the number of fires until hazard occurrence. Severity is calculated for each fire containing a hazard, and is defined as the sum of the total structures destroyed, structures damaged, injuries, and fatalities. For primary results, the average operational time to occurrence, average rate, frequency, and average severity are calculated across the entire range of 2006-2014.

$$OTTO = \begin{cases} report\ DOY\ \ - start\ DOY \\ report\ \%\ containment \end{cases} \tag{1}$$

$$rate = \begin{cases} \frac{hazard\ frequency}{year} \\ \frac{fire\ frequency}{hazard\ frequency} \end{cases} \tag{2}$$

3) Time Series Analysis

Longitudinal trends in the hazard metrics are examined by calculating the average for each year and evaluating time series. Studying the changes in these metrics over time reveals whether there may be other trends affecting these hazards, such as changes in fire characteristics or operations. For instance, we expect operational time to occurrence and number of fires per hazard to be stable over time. If those metrics are not stable over time, this may indicate that another factor is affecting that hazard. In addition to studying hazard metrics over time, trends in predictors, including fire characteristics, intensity, and operations, are also studied using time series. Predictors' time series are calculated using the quantitative data in incident summary reports, rather than the situation reports used for hazard extraction. Fire characteristics time series include total number of fires per year, the total acres burned per year, and the total days fires burn per year. Intensity measures include total number of fatalities per year, total injuries per year, total structures damaged per year, and total structures destroyed per year. Finally, operations measures include total personnel per year, total aerial assets per year, and total cost per year.

4) Secondary Analysis

The relationship between longitudinal trends in hazard frequency, and trends in operations, intensity, and fire characteristics are examined in a secondary analysis consisting of a correlation matrix and regression analysis. Prior to performing these analyses, the time series data is normalized using min-max scaling. The correlation matrix describes how closely each hazard frequency is correlated with each individual measure of operations, intensity, and fire characteristics. Linear regression is performed to determine if trends in hazards can be entirely explained mathematically trends in operations, intensity, or fire characteristics. Typically regression is used as a prediction algorithm; however, here it is used solely to identify important predictors for future hazard and risk prediction models. Given a certain set of continuous inputs, $X = (x_1, x_2, ...x_n)$, linear regression predicts the value of a continuous target variable, Y. The regression uses a linear combination of X to produce the prediction, $\hat{Y}$ seen in (3). The quality and goodness of fit of a regression is measured using the coefficient of determination, R^2, calculated using (4). For this application, the target variable Y is hazard frequency, and the predictor variable X is a vector composed of measures for fire characteristics, intensity, and operations. Due to the experimental design, there are a limited number of Y data points, specifically only 9 years of data from 2006-2014. There are more predictors than data points and the predictors exhibit signs of collinearity. While multiple regression is preferred since it controls for multiple variables at a time and examines the impact of a combination of variables, it is inappropriate to use in this instance as the model will over fit the data. Instead, simple linear regression, where $X = X_1$, must be performed for each predictor individually. The importance of a predictor is measured using R^2, where more important predictors have higher values than less importance predictors. This is performed for each hazard in order to identify the most important predictors and possible explanations for trends in its frequency. Together the regression analysis and correlation matrix determine which factors are most important in predicting frequency for each hazard .

$$\hat{Y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{3}$$

$$R^2 = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}, \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{4}$$

V. Results

A. Hazard Extraction

The LDA topic model used for hazard extraction consisted of $k = 160$ topics, with hyper parameters $\alpha = 1.0$ and $\beta = 0.0001$. Since α affects the amount of topics per document, the value chosen for this model allows for documents to have either only one topic or a mixture of many topics. Quality of a topic model is often measured using C_v coherence, where $0 < C_v < 1$ and $C_v = 1$ indicates perfect coherence. The model used had an average $C_v = 0.615$ with standard deviation of 0.055. For some topics, the raw topic words clearly indicate a certain hazard. For example, topic 34 describes injuries with the words: *'injury', 'hospital', 'minor', 'report', 'transport', 'firefighter', 'medical', 'block', 'member', 'heat', 'ankle', 'injured', 'related', 'release', 'accident', 'occur', 'treatment', 'laceration', 'shoulder', 'yesterday'*. From this output, hazard subject words include *'injury'* and *'hospital'*, while descriptor words include *'minor'* and *'report'*. Accuracy of hazard extraction is measured as the proportion of documents that contain the hazard said to be identified. For each hazard, 15 reports identified as containing the hazard are randomly selected, then manually examined to confirm if the reports contain the hazard. The average accuracy across the hazards is found to be 0.953 with a standard deviation of 0.061. Seen in Table III, aerial groundings have the lowest accuracy at 0.800.

B. Primary results

Each hazard's category and description is displayed in Table III, as well as the average operational time to occurrence, average rate, frequency, and average severity from 2006-2014. Cultural resource hazards include threatened sites such as Native American land, archaeological sites, and heritage sites. Ecological hazards primarily relate to threats to protected habitats and species. Military Base hazards occur when the operation of a base is impacted due to fire or suppression activities. Infrastructure hazards are specifically related to electric power, water, and gas. Resource issues include general shortages of assets, sharing assets with nearby fires, and needs for a specific asset. The most common hazard categories are mission and environmental, which is primarily due to the information documented in the data set.

The operational time to occurrence (OTTO) identifies the most likely period each hazard may occur, which provides insight on risk during a given period. When measured in days, the OTTO provides insight on the size of fires during which the hazard occurs. Specifically hazards with larger average OTTO measured in days likely occur in fires that span more days than those with a smaller measure. When measured in percent containment, the longevity of the fire is controlled for. The average operational time to occurrence in percent containment varies, with livestock hazards and resource issues having the lowest averages at 35.592% and 35.531% respectively, and law violations having the largest at 98.367%. The standard deviations for the OTTO in percent containment are fairly similar across hazards, and generally indicate a wide range for OTTO. Law violations have the most consistent range, with the lowest standard deviation of 11.634%, while command transitions have the largest range with a standard deviation of 36.433%. Command transitions occur throughout a fire, including at the beginning and end, and thus it makes sense that this hazard has the largest spread of time of occurrence.

For all hazards, the total frequency is greater than the frequency of fires the hazard occurs in, and the difference varies from hazard to hazard. For law violations, the two frequencies are almost the same, indicating law violations are often reported only once per fire. The total hazardous terrain frequency is about six times the fire frequency, which implies hazardous terrain is repeatedly reported. Hazardous or dangerous terrain is the most common hazard, followed by extreme weather, command transitions, traffic impacts, and resource issues. The least common hazard is flooding, followed by military bases, ecological hazards, and law violations. Hazardous terrain is encountered on average about 238 times per year and once every four fires. Floods only occurs approximately five times per year and once every 209 fires.

After removing outliers, the average severity across all incidents is 0.322 with standard deviation 0.616. Floods have the greatest severity metric, followed by injuries and infrastructure hazards. The severity of military base hazards is most similar to the severity across all incidents, and thus these hazards have less of an impact. All other hazards have an average severity larger than the score across all incidents, indicating incidents with hazards also experience more severe damage than incidents without hazards. This finding could be due to the hazards themselves, or the conditions which lead to hazard occurrence.

C. Time Series results

Trends from 2006-2014 in predictors of fire characteristics, intensity, and operations are displayed in Fig. 2. The predictors' time series are generated using the quantitative data in incident summary reports, described in Section IV. While there is some variation, generally 2009-2010 is a minimal point for all predictors while both 2006-2008 and 2011-2013 show peaks for most trends. As seen in Fig. 3, hazard frequencies appear to follow the same pattern exhibited in Fig 2. That is, the global minimum frequencies mostly occurs in 2009-2010, while the local maximums occur between 2006-2008 and again 2012-2013. Overall the hazard trends are consistent with the predictor trends, which indicates hazard occurrences may be explained by the trends in the predictors. Fig 4 shows that for the majority of hazards, operational time to occurrences is stable across years. This is evident by the averages occurring within the error ranges. Most hazards occur in range 40-60% containment, and thus there may be more system wide risk in general in this period. Early containment and late containment generally have fewer hazards; however, both law violations and floods often occur in later containment, while livestock hazards occur earlier.

Fig. 2. Fire characteristics, operations, and intensity trends.

Fig. 3. Hazard Frequency trends, where frequency is the number of fires a hazard occurs during.

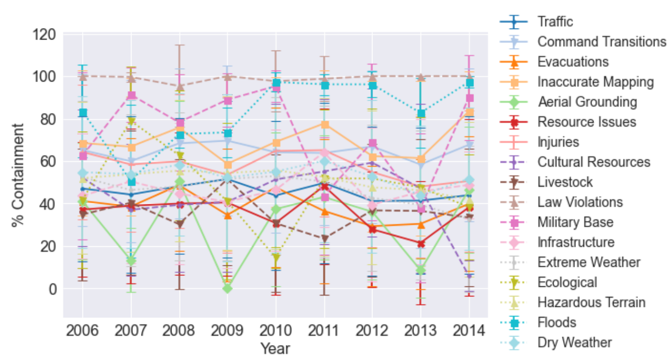

Fig. 4. Trends in average hazard operational time to occurrence with standard deviation error bars.

D. Secondary results

To determine the most important predictors for hazard frequency, the secondary analysis correlation matrix and regression results are considered. The correlation matrix in Fig. 5 shows the pairwise correlations between each predictor and hazard frequency. Generally the correlations are positive and moderate to strong, except damaged structures are weakly negatively correlated with the hazard frequencies. Both cost and structures damaged have weaker correlations than the

TABLE III
HAZARDS EXTRACTED FROM THE ICS-209-PLUS AND THEIR METRICS. OPERATIONAL TIME TO OCCURRENCE AND SEVERITY MEANS AND STANDARD DEVIATIONS ARE SHOWN AS $\mu \pm \sigma$. RATE OF OCCURRENCE METRICS ARE CALCULATED USING THE FREQUENCY OF FIRES A HAZARD OCCURS IN.

Hazard			Metrics						
Category	*Description*	*Accuracy*	*Operational Time to Occurrence*		*Average Rate of Occurrence*		*Frequency*		*Severity*
			days	*% containment*	*per year*	*fires until*	*total*	*# fires*	
Mission	Traffic	0.867	9.831±10.889	45.399±35.27	128.111	7.798	3896	1153	1.597±2.19
	Command Transitions	1.000	15.95±21.251	64.612±36.433	135.111	7.394	2882	1216	2.111±2.733
	Evacuations	1.000	11.109±12.66	38.31±32.025	115.556	8.645	3395	1040	2.86±3.864
	Inaccurate Mapping	1.000	6.201±6.422	68.63±32.711	71.889	13.896	827	647	1.493±2.26
	Aerial Grounding	0.800	16.086±17.537	17.537±30.373	13.889	71.928	219	125	3.239±4.162
	Resource Issues	0.867	10.546±12.107	35.531±33.546	92.667	10.781	2218	834	1.979±2.84
	Injuries	1.000	9.901±10.425	58.098±33.323	61.222	16.318	902	551	4.215±4.439
Wildland Urban Interface	Cultural Resources	0.933	8.534±8.306	49.153±35.295	22.333	44.731	861	201	3.406±4.367
	Livestock	1.000	8.421±9.204	35.592±32.514	30.889	32.342	728	278	2.547±3.357
	Law Violations	1.000	2.128±6.797	98.367±11.634	21.111	47.321	196	190	1.047±0.791
	Military Base	1.000	11.112±16.800	67.894±36.015	6.889	145.016	181	62	0.352±0.643
	Infrastructure	0.933	10.943±12.062	46.984±34.227	23.333	42.814	816	210	3.607±5.154
Environmental	Extreme Weather	0.933	5.701±6.95	47.022±35.963	154.000	6.487	3343	1386	1.482±2.217
	Ecological	0.933	9.942±9.539	49.778±32.954	13.000	76.846	535	117	1.99±2.909
	Hazardous Terrain	1.000	10.892±12.212	51.541±35.964	238.111	4.196	12802	2143	1.074±1.695
	Floods	1.000	26.545±28.85	85.518±20.829	4.778	209.093	156	43	8.333±13.013
	Dry Weather	0.933	9.143±10.1	54.621±35.983	99.111	10.080	2857	892	1.841±2.779

other predictors. The predictors that have the greatest number of statistically significant correlations are acres burned, aerial assets, and personnel. Fire frequency, injuries, and fatalities also have a number of significant correlations. Every hazard has a statistically significant correlation with at least one predictor, but law violations and military base hazards both have much weaker correlations with the predictors. The injuries hazard has a perfect positive correlation to the injuries predictor variable, indicating the frequency calculated using hazard extraction has high validity for injuries. The correlations between the predictors are also examined and show mostly positive, moderate to strong correlations, indicating some predictors may be dependent on other predictors, such as fire frequency. These findings indicate a subset of the predictors, including acres burned, aerial assets, and personnel, may be most important in predicting frequency for most, but not all, hazards.

As expected, the multiple regression model including all ten predictors from fire characteristics, operations, and intensity, exactly fits the hazard frequency trends from 2006-2014. As stated in section IV, in this work simple linear regression is used to determine the best predictors for hazard frequency. In Table IV, the best predictor for each hazard is noted by *. Confirming the results from the correlation matrix, personnel and injuries are the most important predictors for a majority of hazards, followed by aerial assets and fire frequency. Other predictors still show high levels of importance for most hazards, including fatalities, days burning, and acres burned. Fatalities are the most important predictor for floods, which is in agreement with the severity metric for floods. Since fire frequency is not the best predictor for most hazards, hazard occurrence cannot be explained only by fire occurrence. Hence there are likely other predictors, perhaps related to fire complexity, that have a stronger relationship to hazard occurrence.

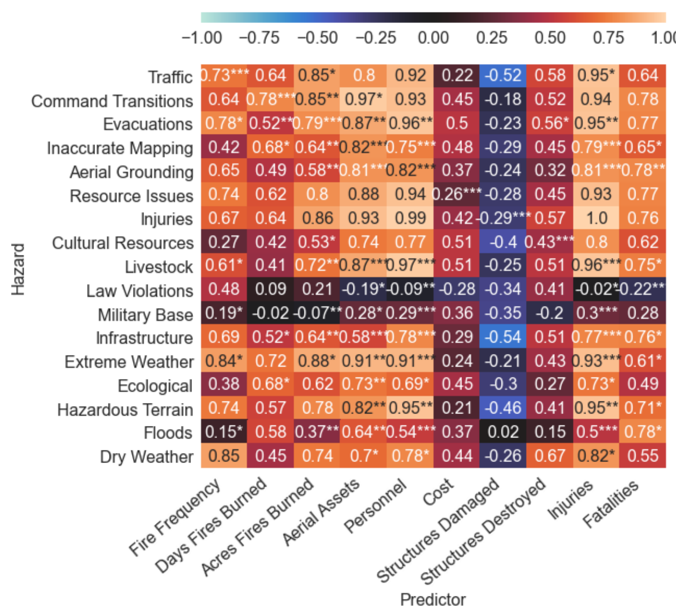

Fig. 5. Correlation matrix for trends in hazard frequency, fire characteristics, intensity, and operations. Statistically significant correlations are denoted *.

VI. DISCUSSION

The hazards extracted are consistent with current knowledge, but hazards at the wildland urban interface may be of concern for emerging wildfire response operations as these hazards tend to have more of an impact on civilian life than mission or environmental hazards. As expected, operational time to occurrence is mostly consistent from 2006-2014 for the extracted hazards. This finding suggests that targeting certain times of operational occurrence may be beneficial and effective risk mitigation, nonetheless further detailed analysis is necessary. Generally hazard frequency time series follow the same trends seen in time series for fire characteristics, operations, and intensity. It can be deduced from the time series that hazard frequency can mostly be explained by the predictors. Thus as the predictors, such as fire frequency, increase, hazard frequencies are expected to increase. Since incidents with hazards have higher damage severity, increases in hazards likely result in increases of damage. The secondary analysis demonstrates that personnel, injuries, and aerial assets

TABLE IV
LINEAR REGRESSION COEFFICIENT OF DETERMINATION R^2 SCORES. PREDICTORS ARE FROM FIG.2.

Hazard	**Full Model**	**Fire Frequency**	**Days Burning**	**Acres Burned**	**Aerial Assets**	**Personnel**	**Cost**	**Structures Damaged**	**Structures Destroyed**	**Injuries**	**Fatalities**
Traffic	1.0	0.531	0.412	0.730	0.633	0.843	0.047	0.266	0.337	0.898*	0.4135
Command Transitions	1.0	0.415	0.610	0.716	0.935*	0.860	0.198	0.034	0.266	0.887	0.6151
Evacuations	1.0	0.613	0.272	0.627	0.751	0.913*	0.252	0.052	0.319	0.911	0.591
Inaccurate Mapping	1.0	0.178	0.462	0.416	0.672*	0.556	0.226	0.086	0.206	0.620	0.423
Aerial Grounding	1.0	0.422	0.236	0.334	0.658	0.671*	0.136	0.058	0.100	0.656	0.61
Resource Issues	1.0	0.543	0.378	0.639	0.772	0.880*	0.065	0.078	0.204	0.870	0.5865
Injuries	1.0	0.447	0.410	0.732	0.864	0.981	0.179	0.087	0.320	1.000*	0.582
Cultural Resources	1.0	0.075	0.175	0.281	0.540	0.591	0.261	0.162	0.188	0.633*	0.388
Livestock	1.0	0.375	0.172	0.522	0.750	0.937*	0.264	0.063	0.263	0.924	0.564
Law Violations	1.0	0.233*	0.009	0.043	0.034	0.009	0.077	0.117	0.171	0.000	0.047
Military Base	1.0	0.036	0.000	0.005	0.076	0.083	0.133*	0.120	0.039	0.088	0.078
Infrastructure	1.0	0.473	0.272	0.415	0.334	0.602*	0.082	0.296	0.255	0.588	0.571
Extreme Weather	1.0	0.710	0.513	0.780	0.820	0.824	0.057	0.043	0.188	0.863*	0.374
Ecological	1.0	0.141	0.465	0.383	0.535*	0.476	0.203	0.089	0.073	0.530	0.242
Hazardous Terrain	1.0	0.550	0.326	0.603	0.679	0.895	0.046	0.214	0.170	0.900*	0.505
Floods	1.0	0.022	0.333	0.140	0.408	0.295	0.138	0.000	0.023	0.250	0.608*
Dry Weather	1.0	0.730*	0.205	0.553	0.497	0.608	0.193	0.070	0.447	0.672	0.305

are the best predictors for the frequency of most hazards. This relationship could possibly be explained by a third variable or combination of variables, such as fire frequency, terrain, and burned acres. Large, complex, fast spreading, and difficult to access fires often require more personnel and aerial assets. Thus it is possible that the underlying causes for increases in injury, increases in personnel, and increases in aerial assets are responsible for the relation between the predictors and hazard frequency. This analysis does not yield causes for hazards, rather undirected relationships between hazards and predictors; therefore, some trends in predictors may cause the trends in hazards, while other predictor trends are caused by hazard trends.

Hazards and relevant metrics identified using HEAT provide insight on existing system risk, as well as opportunities for mitigation and improvements. Specific use cases for UAS, including surveillance and logistic delivery, could mitigate inaccurate mapping and resource issues. Inaccurate mapping is an issue for current operations as both paper and electronic maps are often only updated daily at best. Fire size may be over or under estimated, and fast spreading regions may go unnoticed for periods of time. However, maps used in current operations are effective and may prioritize protecting certain valuable regions, including areas with critical infrastructure, ecological importance, military bases, and cultural resources. The advanced geofencing capabilities of emerging UAS technology described in STEReO, such as ICAROUS and Safeguard, could consider these valuable threatened areas with precision and less effort than human operators. Traffic hazards and evacuations may create areas with higher density civilian populations than normal, resulting in elevated risk should a catastrophic aerial incident occur overhead. TCAS reduces the likelihood of incident occurrence, while advanced geofencing capabilities of UAS could also reduce incident likelihood by avoiding flying above these high-density areas. Current operations have little ability to mitigate environmental hazards resulting in unsafe operational conditions, and thus must balance the safety of the operators alongside the urgency of containing the fire. On the other hand, UAS may operate and perform tasks in these circumstances. While few of the hazards extracted directly result in a catastrophic event, the hazards do tend have higher damage severity scores. Thus there are opportunities for both hazard and damage mitigation, potentially through advanced capabilities.

The results presented are a preliminary exploration and serve as an initial analysis of the ICS-209-PLUS, as well as a demonstration of HEAT. While a filtering step is required to obtain hazards because the data set contains a non-trivial quantity of non-hazard related narratives, topic modeling significantly decreases the bias, work load, and time required for hazard extraction when compared to human tagging. The list of hazards complied is exhaustive from the ICS-209-PLUS only, thus there could be additional hazards that exist in wildfire emergency response that were not documented in this data set. Since ICS-109-PLUS is comprised of form ICS-209, a different form, such as SAFENET or SAFECOM, could be a better source for hazard-specific data. The ICS-209-PLUS has not yet been updated with data from 2015-present, and this data would be useful for verifying and validating the results presented. This research only used nine years worth of data, which limits the analysis capabilities and robustness. Despite this constraint, the findings presented provide insightful quantitative information on hazards in wildfire response.

VII. CONCLUSION

In this paper a methodology for Hazard Extraction and Analysis of Trends (HEAT) is presented and applied to the ICS-209-PLUS data set consisting of wildfire response incident reports. The exploratory HEAT methodology and results yield a large-scale historical data driven study on hazards and related factors in wildfire response. Due to increasing wildfire severity and system complexity, both current and emerging operations require safety assurance through risk analysis. Presence of hazards may result in more severe damage, thus to mitigate damage, potential causes of hazard occurrence must be explored. The findings presented suggest predictors, including fire characteristics, operations, and intensity, can explain trends in hazard frequency, and thus future real-time risk models may consider some of these predictors when investigating hazard occurrence. Additionally, the identification of when a hazard is likely to occur in a single operation, how often it is expected to occur over a given time period, when it may next occur, and its severity, provides insight relevant to targeted risk-mitigation models.

VIII. Future Directions

Future work includes applying HEAT to additional databases that contain hazard-relevant and aerial operations specific information. The findings presented in this paper will be further analyzed for validity, specifically by comparing results for different regions of the U.S. and identifying any results that are artifacts of the data set. Regression analysis predicting damage severity for a given fire based on hazard occurrence, fire characteristics, operations, and intensity will be performed. Additional analysis on the ICS-209-PLUS includes multiple regression to understand how different combinations of predictors are related to hazard occurrence, as well as analysis on an individual fire scale, rather than the annual scale used in this work. The current work is an exploratory first step in developing historical data-driven risk assessment models for wildfire response and can help targeted risk mitigation by identifying predictors and critical, high-risk times of operation. Future research using HEAT-like algorithms will allow for a similar analysis on causes of hazardous events. Together the causes and hazards can create various predictive models. For example, a model can be created to estimate real-time risk during a single operation based on current reported conditions.

Acknowledgment

We would like to thank Dr. Misty Davies at NASA Ames Research Center, Project Manager of the System Wide Safety project, for her encouragement and feedback on this work. We would also like to thank Dr. Carlos Paradis for his feedback.

References

[1] J. D. Miller and H. Safford, "Trends in wildfire severity: 1984 to 2010 in the sierra nevada, modoc plateau, and southern cascades, california, usa," *Fire Ecology*, vol. 8, pp. 41–57, 2012.

[2] P. Dennison, S. Brewer, J. Arnold, and M. Moritz, "Large wildfire trends in the western united states, 1984-2011," *Geophysical Research Letters*, vol. 41, 04 2014.

[3] J. Davis and K. T. Weber, "Spatio-temporal relationships of historic wildfires: Using the nasa recover historic fires geodatabase to perform long-term analysis of wildfire occurrences in the western united states," Idaho State University GIS Center, Tech. Rep., 2018.

[4] D. E. Calkin, J. D. Cohen, M. A. Finney, and M. P. Thompson, "How risk management can prevent future wildfire disasters in the wildland-urban interface," *Proceedings of the National Academy of Sciences*, vol. 111, no. 2, pp. 746–751, 2014. [Online]. Available: https://www.pnas.org/content/111/2/746

[5] St. Denis, Lise A., Mietkiewicz, Nathan P., Short, Karen C., Buckland, Mollie, and Balch, Jennifer K., "All-hazards dataset mined from the us national incident management system 1999–2014," *Scientific Data*, vol. 7, p. 64, 2020.

[6] J. Mercer, R. Mcswain, and C. Ippolito, "Scalable traffic management for emergency response operations (stereo)," in *AIAA Aviation Conference*. National Aeronautics and Space Administration, June 2020. [Online]. Available: https://hsi.arc.nasa.gov/groups/AOL/downloads/AIAA-STEReO-2020-v4-web.pdf

[7] H. S. Walsh, E. Spirakis, S. R. Andrade, D. E. Hulse, and M. D. Davies, "SMARt-STEReO: Preliminary concept of operations," NASA Ames Research Center, Tech. Rep. NASA/TM-20205007665, 2020.

[8] "Wildland fire incident management field guide," National Wildfire Coordinating Group, Tech. Rep. PMS 210 NFES 002943, 2014.

[9] "NWCG standards for aerial supervision," National Wildfire Coordinating Group, Tech. Rep. PMS 505 NFES 002544, 2020.

[10] "NWCG standards for airspace coordination," National Wildfire Coordinating Group, Tech. Rep. PMS 520, 2018.

[11] "NWCG standards for fire unmanned aircraft systems operations," National Wildfire Coordinating Group, Tech. Rep. PMS 515, 2019.

[12] "NWCG standards for airtanker base operations," National Wildfire Coordinating Group, Tech. Rep. PMS 508, 2021.

[13] "NWCG standards for helicopter operations," National Wildfire Coordinating Group, Tech. Rep. PMS 510, 2019.

[14] L. J. Glaab, C. V. Dolph, S. D. Young, N. C. Coffey, R. G. McSwain, M. J. Loga, and D. E. Harper, "Small unmanned aerial system (uas) flight testing of enabling vehicle technologies for the uas traffic management project," Tech. Rep. NASA/TM–2018-219816, April 2018.

[15] L. J. Glaab, P. C. Glaab, P. C. Lusk, B. J. Petty, R. W. Beard, C. V. Dolph, and R. G. McSwain, "Safe2ditch autonomous crash managementsystem for small unmanned aerial systems: Concept definition and flight test results," Tech. Rep. NASA/TM–2018-220110, 2018.

[16] M. Consiglio, C. Muñoz, G. Hagen, A. Narkawicz, and S. Balachandran, "Icarous - integrated configurable algorithms for reliable operations of unmanned systems," in *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference (DASC)*, 2016.

[17] S. D. Young, E. T. Dill, K. J. Hayhurst, and R. V. Gilabert, "Safeguard: Progress and test results for a reliable independent on-board safety net for uas," in *Proceedings of IEEE/AIAA Digital Avionics Systems Conference (DASC "17)*, 2017.

[18] G. N. Szatkowski, L. A. Ticatch, C. M. Morris, and A. A. Cavone, "Indoor ground testing of a small uas sense and avoid airborne doppler radar," Tech. Rep. NASA/TM–2019–220280, 2019.

[19] J. Chamberlain, M. Consiglio, and C. Muñoz, "Danti: Detect and avoid in the cockpit," in *2017 AIAA Aviation and Aeronautics Forum*, 2017.

[20] C. Dunn, D. Calkin, and M. Thompson, "Towards enhanced risk management: Planning, decision making and monitoring of us wildfire response," *International Journal of Wildland Fire*, vol. 26, p. 551, 01 2017.

[21] M. Thompson, Y. Wei, D. Calkin, C. O'Connor, C. Dunn, N. Anderson, and J. Hogland, "Risk management and analytics in wildfire response," *Current Forestry Reports*, vol. 5, 12 2019.

[22] S. Goswami and D. Raychaudhuri, "Identification of disaster-related tweets using natural language processing," in *International Conference on Recent Trends in Artificial Intelligence, IOT, Smart Cities & Applications (ICAISC-2020)*, 05 2020.

[23] K.-G. Thanos, A. Polydouri, A. Danelakis, D. Kyriazanos, and S. Thomopoulos, "Combined deep learning and traditional nlp approaches for fire burst detection based on twitter posts," 04 2019.

[24] J. McCarty, E. Levin, K. Endsley, S. Aden, and J. Bialas, "Relating big data to local natural hazards: Lessons learned from data mining the twitter api for volunteered geographic information on earthquakes, wildfires, and prescribed fires in the contiguous united states," 04 2015.

[25] F. Zhang, H. Fleyeh, X. Wang, and M. Lu, "Construction site accident analysis using text mining and natural language processing techniques," *Automation in Construction*, vol. 99, pp. 238 – 248, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0926580518306137

[26] S. Sarkar, S. Vinay, and J. Maiti, "Text mining based safety risk assessment and prediction of occupational accidents in a steel plant," 03 2016.

[27] X. Pan, H. Wang, W. You, M. Zhang, and Y. Yang, "Assessing the reliability of electronic products using customer knowledge discovery," *Reliability Engineering & System Safety*, vol. 199, p. 106925, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832019307732

[28] H. Dussault, P. S. Zarubin, S. Morris, and D. Nicholls, "Harvesting reliability data from the internet," in *2008 Annual Reliability and Maintainability Symposium*, 2008, pp. 322–327.

[29] S. R. Andrade and H. S. Walsh, "Knowledge discovery for early failure assessment of complex engineered systems using natural language processing," in *ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2021)*, 2021.

[30] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. null, p. 993–1022, Mar. 2003.