Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)

Topics and Trends in Incident Reports

Using Structural Topic Modeling to Explore Aviation Safety Reporting System Data

Kenneth D. Kuhn
Engineering and Applied Sciences
RAND Corporation
Santa Monica, USA
kkuhn@rand.org

***Abstract*—The Aviation Safety Reporting System includes over a million confidential reports describing safety incidents. Natural language processing techniques allow for relatively rapid and largely automated analysis of large collections of text data. Meaningful interpretation of the results and further investigations by subject matter experts can follow. This article describes the application of structural topic modeling to Aviation Safety Reporting System data. Results reveal that the application is able to identify known issues. The method also has the potential to identify previously unknown connections that may warrant further, more manual, study. Results reported here highlight the importance of fuel pump, tank, and landing gear issues and the relative insignificance of smoke and fire issues for private aircraft. The results also uncovered evidence of the prominence of the Quiet Bridge Visual and Tip Toe Visual approach paths at San Francisco International Airport in safety incident reports.**

Keywords-aviation safety; machine learning; structural topic modeling; natural language processing;

I. INTRODUCTION

Public and private agencies in Europe, the United States, and elsewhere are implementing many substantive changes to air transportation operations, driven by concern about the impacts of increasing air traffic. For example, the Federal Aviation Administration (FAA) is in the midst of implementing "Wake Recategorization" procedures at busy airports, "reducing separation criteria for multiple runway operations" [1]. Wake Recategorization is a part of the NextGen initiative to modernize air traffic control.

Researchers have defined generic methodologies for assessing the safety implications of proposed operational changes. Relevant studies include [2] and [3]. Researchers have also used simulation data to study the safety implications of specific changes to policies, procedures, and infrastructure. For example, [4] studied alternative configurations for the North Airfield at Los Angeles International Airport (LAX).

This article describes an exploratory analysis of recently reported empirical, primarily text, data on aviation safety. Structural topic modeling (STM), a technique from the field of machine learning, was applied to a large corpus of incident reports. STM identifies topics contained within a (potentially very large) set of documents, finding themes and providing structure for quantitative analysis. STM also estimates the impacts of covariate data, including time, on topic prevalence. The methodology can reveal trends in the frequency with which topics with intuitive meanings appear.

One goal of this study was to evaluate the usefulness of this novel method for identifying safety issues. A more ambitious goal was to begin using the method to find previously unreported connections or themes in incident reports.

II. AVIATION SAFETY REPORTING SYSTEM DATA

A. Data Description and Initial Exploration

The Aviation Safety Reporting System (ASRS) lets pilots, air traffic controllers, airline dispatchers, and others submit confidential reports of safety incidents. The FAA and the National Aeronautics and Space Administration (NASA) developed and manage the ASRS, in part to "provide data for planning and improvements to the future National Airspace System" [5]. Similar systems exist elsewhere, including CHIRP in the United Kingdom and REPCON in Australia. In 2015, the ASRS database included over 1.3 million records and was adding roughly 7,500 additional reports each month [5].

Analysts anonymize submitted reports and code the results into a database. ASRS database records include narrative portions, lengthy blocks of free text. Figure 1 shows the beginning of a narrative portion of an ASRS record.

Roughly 68% of the incidents reported between January 2010 and April 2015 refer to passenger flights, 14% to personal flights, 6% to cargo flights, and the remaining 13% to some other category of flights. The reporting organization was classified as an "Air Carrier" for 58% of reports, as "Government" for 16%, as "Personal" for 12%, and as some other category (including "Military" and "Corporate") of organization for the remaining 13% of reports.

> JFK Tower cleared us for takeoff on 31R; Kennedy 1 Departure (Breezy Point Climb). After our takeoff roll the Tower cleared a heavy aircraft into position on 22R Intersection YA and hold at idle thrust at an intersecting runway. At around 100 knots we received a pretty good jolt from his thrust buffet. Quick left rudder and left aileron was used to counteract the thrust buffet…

Figure 1: Example of a Narrative from an ASRS Record.

This research was partially funded by the NASA NextGen – Concepts and Technology Development Project under program announcement NNH14ZEA001N-CTD1.

In addition to the narrative portion, each ASRS record contains information on conditions during the incident including: the month, "locale" (e.g., "LGA.Airport" or "TUL.TRACON"), meteorological conditions, phase of flight (e.g., "Climb" or "Descent"), flight mission (e.g., "Passenger" or "Skydiving"), etc.

There have been prior research efforts exploring ASRS data, searching for issues that could be resolved to prevent or reduce the frequency of specific types of incidents. For example, [6] report on the role of cockpit alarm systems in incidents. [7] describe a project that lasted for seven years, involving manual classification of incident reports and examination of relationships in derived data. An example of the authors' conclusions is that "The most common controller errors involve failure to coordinate traffic with other elements of the air traffic control system" [7].

There has been recent progress in the fields of machine learning, computational linguistics, and natural language processing. There are established theories and tools for performing largely automated and relatively rapid analysis of a corpus. As an example, there are fast and easy-to-implement algorithms that identify the most frequently observed phrases. Tab. I shows the results when applying such an algorithm to ASRS records reported between January 2010 and April 2015. (Some of the results may reflect how analysts code submitted reports. For example, FO appears to be a code inserted into narratives to highlight a reference to the first officer of a flight.)

TABLE I. FREQUENTLY OBSERVED PHRASES IN ASRS NARRATIVES

Phrase	*Observation Count*
5 word phrases	
first officer FO pilot flying	38
cleared visual approach runway R	33
climb via SID except maintain	30
declared emergency returned departure airport	29
we cleared visual approach runway	28
4 word phrases	
first officer pilot flying	286
in future I will	213
I asked first officer	206
cleared visual approach runway	160
aircraft maintenance manual AMM	149
3 word phrases	
air carrier X	1,096
first officer I	892
at point I	872
at time I	697
landed without incident	614

B. Natural Language Processing of Aviation Safety Data

Tab. I reveals the importance of the first officer and pilot. The application of other natural language processing techniques holds the promise of more interesting results. This explains the recent proliferation of papers describing applications of natural language processing to ASRS data.

[8] describes a way to visualize narratives from ASRS records on a 2-D graph based on "latent relationships" among the narratives evident in word and phrase use.

[9] introduces a technique the authors call Semi-supervised Impurity based Subspace Clustering – Multi Label (SISC-ML) and applies this technique to ASRS (and other) data. SISC-ML classifies text records, linking each document to multiple labels. An important input is a training data set that includes previously identified labels for each of several documents. The focus of the article is on the description and testing of SISC-ML, rather than on aviation safety per se.

[10] applies two different natural language processing methods to identify the "cause types" of aviation safety incidents as reported in ASRS records. The techniques introduced provide an automated way to identify "shaping factors," first described and assigned manually in [11]. This is another example of a classification model requiring previously established training data.

Topic modeling, a form of machine learning that aims to identify "the main themes that pervade a large and otherwise unstructured collection of documents" [12], could be useful. Topic modeling has gained prominence recently as analysts search for ways to organize the large volume of text data available on the internet. [13] describes a suite of methods called sparse machine learning for topic modeling and other tasks, testing the methods on ASRS data. The authors "reveal causal and contributing factors in runway incursions" and "automatically discover four main tasks that pilots perform during flight" [13]. The four tasks that pilots perform are: aviate, navigate, communicate, and manage systems [13]. The automated analysis of runway incursions reveals specific runway/taxiway intersections that are frequently mentioned in incident reports.

[14] provides an overview of the promise and difficulty of applying natural language processing techniques to aviation safety data. The authors describe topic modeling, classification model fitting, and other methods. The authors note that they have developed tools that "are currently in test or in use both at the national and international levels, by airline companies as well as by regulation authorities" [14]. One interesting conclusion is that "It appears that topic modelling is very suitable for [incident report] data" and that identified topics highlight "relevant aspects of [these] documents, as can be seen through an expert's interpretation" [14]. The authors highlight the importance of interactive analysis, where human experts explore results identified by algorithms. Specific findings directly related to aviation safety are not provided.

III. STRUCTURAL TOPIC MODELING

This article describes applications of structural topic modeling to ASRS data. STM is a form of topic modeling, a probabilistic way to describe documents in terms of topics.

The most common form of topic modeling is latent Dirichlet allocation (LDA). A brief introduction to LDA is provided here.

LDA assumes documents and the words within them are derived from a "generative probabilistic model" [15]. Each document is, in theory, generated via the following process:

- The number of words N is a random variable drawn from a Poisson(ξ) distribution.
- The parameter θ is a random variable drawn from a Dirichlet(α) distribution.
- The topic of each word z_n is a random variable drawn from a multinomial(θ) distribution.
- Each word w_n is a random variable based on another draw from a multinomial distribution defining $p(w_n|z_n,\beta)$ terms.

Fig. 2, originally appearing in [15], shows the 'plate notation' representation of the LDA model. There are single corpus-wide parameters α and β. There is one 'plate,' and an associated parameter θ for each of the M documents in a corpus. Then, there is another 'inner plate' that is replicated for each of the N words in each document. z and w, topics and words, appear in this inner plate. There is a topic linked to each specific place where a word appears in each specific document. According to the theory, the topic determines the distribution used to generate the word. Different topics can generate the same English language word.

Estimates of the parameters of the model described above provide researchers with data on topic representation within each document and within the corpus. These data also reveal the words most associated with each topic, allowing analysts to ascribe intuitive meanings to topics. In its most general form, LDA can also be applied to non-text data and has proven useful in image recognition.

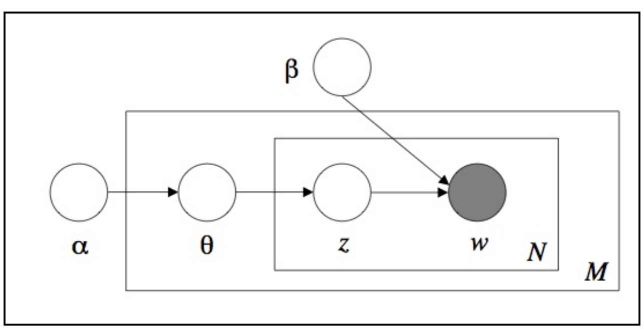

Figure 2: Latent Dirichlet Allocation Model, in Plate Notation [15].

Applying a Bayesian approach, the "key inferential problem" is to compute the posterior distribution of the latent variables given the text from a document [15]. This can be expressed as evaluating the following application of Bayes' Law (also from [15]).

$$p(\boldsymbol{\theta},\boldsymbol{z}|\boldsymbol{w},\alpha,\beta) = p(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{w}|\alpha,\beta) \,/\, p(\boldsymbol{w}|\alpha,\beta) \qquad (1)$$

Although it is not typically feasible to directly evaluate equation (1), there are many ways to find approximate solutions using expectation-maximization (EM) algorithms.

Note that within an LDA model, the probability of observing a particular word at a particular location within a document is a function only of the relevant topic and the model parameters. The topic is a function only of the model parameters. In particular, LDA does not allow us to model changes in the representation of topics and words within documents over time or as a function of (other) covariate data.

Structural topic modeling is an alternative to LDA that allows researchers to link topics to covariate data and to model changes in topic prevalence over time. STM has recently been applied to scientific texts on climate change, revealing links between corporate funding and the framing of scientific studies [16]. It has also been applied to social media data in a variety of ways. One study shows that "STM can be used to detect significant events such as the downing of Malaysia Air Flight 17" when applied to twitter data [17]. Another study shows how STM can be used to explore relatively large data sets including course evaluations and discussion forum posts from a Massive Open Online Course [18].

In an application of STM, the model parameters describing topic proportions (θ terms) are assumed to be random variables drawn from Log-normal distributions that are parameterized based on covariate data. The relevant topics in each document are assumed to be drawn from a distribution specific to the document based on the covariate data. The word distributions are similarly specific to the document and topic. [19] provides further technical details on structural topic modeling.

Intuitively, STM allows the model to capture, for example, a situation where "Democrats [use] the word 'estate' more frequently than Republicans while discussing taxation" [19]. In the context of this study of aviation safety, one could build a model that determines whether a pilot would be more likely to mention the word 'rain' while describing an incident during the 'landing' phase of flight than while describing an incident during the 'takeoff' phase of flight. This would not be possible when applying the base form of LDA.

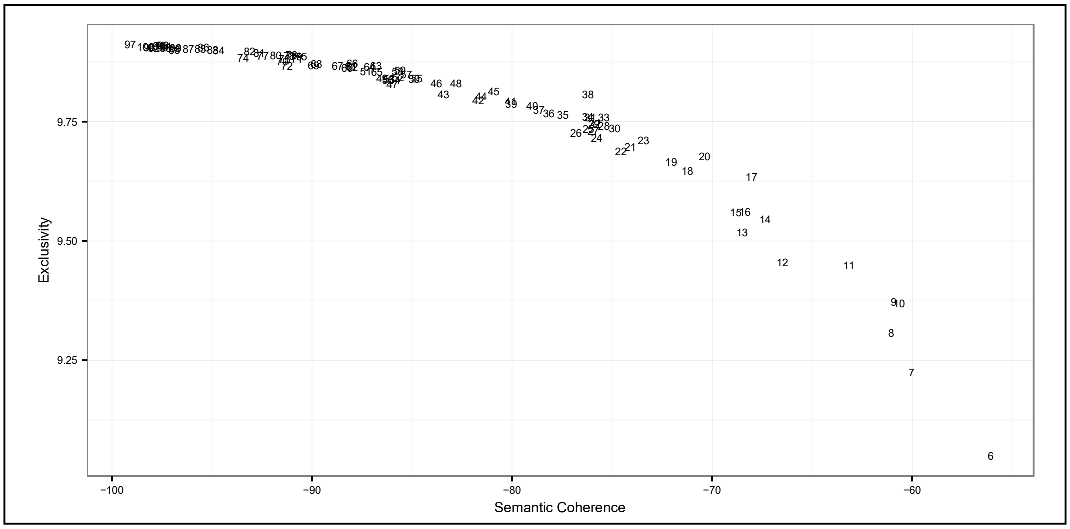

Figure 3: Statistics Used to Select Number of Topics.

IV. Results: All Recent Reports

Structural topic modeling was applied to the ASRS data covering incidents that occurred between January 2010 and April 2015. The stm package for structural topic modeling using the R statistical software was used [20]. Punctuation, whitespace, and stop words were removed from the corpus as a first step in the analysis. The flight mission, the phase of flight, and the time at which the incident was reported were selected as covariate data to be studied.

A. *Selecting the Number of Topics*

A natural first problem when applying STM involves identifying the number of topics. There is no single correct way to address this issue, but one possibility involves studying the trade-off between *semantic coherence* and *exclusivity.*

Semantic coherence is based on measures of how frequently individual words occur and pairs of distinct words co-occur. Such measures can help analysts avoid defining topics that are problematic for one of several specific reasons. For example, words may be linked in a chain. The word 'wall' is linked to the word 'Trump' which is also linked to the word 'Hillary' but the words 'wall' and 'Hillary' should not be assigned to the same topic in a topic model. This issue and its remedy via the semantic coherence measure are introduced in [21]. As the number of topics in a model increases, the semantic coherence will decrease, generally speaking.

A topic is considered to be exclusive if the words that have a high probability of appearing conditional on that topic have low probabilities conditional on other topics. For example, there may be a topic present within our ASRS data that refers to problems with landing gear. If the word 'gear' frequently occurs when this topic comes up but rarely occurs otherwise, then this finding would be evidence of the topic's exclusivity. As the number of topics in a model increases, the exclusivity of the model as a whole will typically increase.

Fig. 3 graphs the values of semantic coherence and exclusivity when applying STM to select between 6 and 100 topics in ASRS data. Each data point is based on a distinct analysis and a distinct set of topics. The label shows how many topics were found. So the location of the data point labeled 97 reflects the semantic coherence and exclusivity of a model that identified 97 separate topics in the ASRS data. The fact that the observed values of exclusivity (semantic coherence) are between 9 and 10 (between -100 and -50) reflects details about the frequency of word occurrence in the ASRS data that are unimportant here. Attention should instead focus on comparisons among the data points when selecting the number of topics.

Figure 3 shows the expected trends for, and trade-off between, semantic coherence and exclusivity. There is no clear correct number of topics in the data. A few observations do stand out, including the cases where 9, 10, 11, 17, 20, and 38 topics were found. Arguably the biggest outlier is the case where 17 topics were found. The following sub-sections focus on this case.

B. *Identified Topics and Intuitive Meanings*

After applying STM, specific words in narratives were linked to topics. These topics do not have pre-existing labels or definitions. In order to assign intuitive meanings to these topics, one must study the words linked to each topic.

The most obvious way to do this would be to look at the words that have the highest probability of occurring conditional on each topic. The problem here is that certain words such as 'aircraft' and 'airport' will show up as high probability words for many topics. An alternative would be to focus on *lift*, the probability of word occurrence conditional on the topic divided by the probability of word occurrence across the corpus. [22] suggests using the *FREX* statistic, the ratio of word frequency conditional on a topic to word-topic exclusivity (described informally in the preceding section of this article).

TABLE II. TOPICS IDENTIFIED AND LABELED

Topic	***Criteria***	***Word 1***	***Word 2***	***Word 3***	***Word 4***	***Word 5***	***Labels***	***Expected Topic Proportion***
1	Prob	arriv	clearance	atc	departure	cross	**ATC**	0.078
	Lift	domno	fms	mistook	sefr	trup		
	FREX	fms	restrict	rnav	sid	waypoint		
2		runway	tower	aircraft	taxi	clear	**surface, routing**	0.074
		backtaxi	ogg	quebec	foxtrot	papa		
		taxiway	runway	hold	short	taxi		
3		land	emerg	airport	fire	declar	**smoke, fire**	0.064
		smoke	midcabin	fire	fum	tailpipe		
		fire	smoke	declare	emerg	divert		
4		fuel	gear	land	engine	tank	**fuel pump, tank, landing gear**	0.041
		sputter	desert	enrich	pump	imbalance		
		tank	gear	pump	fuel	hydraulic		
5		flight	plan	dispatch	crew	hour	**fatigue**	0.057
		circadian	polar	nighter	fdp	awake		
		fatigue	schedule	sleep	hour	duty		
6		engine	pressure	cabin	start	oil	**engine, oil, pressure**	0.034
		buy	outflow	psi	bleed	pressure		
		oil	bleed	pressure	mask	temperature		
7		aircraft	control	traffic	sector	airspace	**airspace**	0.082
		apreq	datablock	dside	jurisdic	loa		
		sector	carrier	train	dside	separate		
8		approach	visual	final	land	runway	**approach**	0.071
		phanom	mateo	glidepath	stable	loc		
		approach	visual	tcas	sight	terrain		
9		get	time	just	said	need	**human factors**	0.107
		deep	stupid	leadership	pride	imagine		
		thing	something	think	know	realize		
10		airport	pilot	radio	flight	traffic	**low-altitude traffic**	0.061
		civilian	foreflight	laser	tfr	tfrs		
		helicopter	ctaf	class	tfr	pattern		
11		aircraft	speed	wind	weather	turbulence	**weather**	0.041
		pub	recat	vortex	chop	turbulence		
		turbulence	wake	wind	encounter	moderate		
12		captain	flap	takeoff	first	officer	**thrust, flaps**	0.056
		dual	rto	asymmetric	thrust	autothrust		
		flap	trim	thrust	autothrottle	lever		
13		aircraft	mechanic	inspect	install	remove	**mechanic**	0.040
		jobcard	rii	washer	bolt	bundle		
		install	card	cable	repair	bolt		
14		maintain	aircraft	system	control	mel	**maintenance, fault**	0.047
		veil	dmi	nef	mel	elac		
		mel	inop	fault	maintain	breaker		
15		flight	passenger	door	attend	captain	**passengers, cargo**	0.047
		clinic	csr	lightheaded	mail	monoxide		
		agent	door	galley	cargo	bag		
16		aircraft	brake	left	right	ramp	**tug, brake**	0.049
		tug	wand	rope	traction	towbar		
		tug	wheel	brake	deice	snow		
17		altitude	climb	feet	level	atc	**climb**	0.052
		barometric	altimeter	gyro	rime	compass		
		climb	altitude	cloud	altimeter	feet		

Tab. II shows, for each topic found in the ASRS database, the five most highly ranked words when ordering by probability of occurrence conditional on topic (Prob), by lift (Lift), and by the FREX statistic (FREX). Recall that different topics can generate the same word.

A few labels that have intuitive meanings are suggested for each of the topics. These labels are based on the words linked to each topic and expert judgment. Some topics appear to cover distinct issues or systems, and are therefore assigned more than one label. For example, topic 15 is linked to words describing passengers and other words describing aircraft cargo. Since the goal is to succinctly describe the topics, however, no topic is assigned more than three labels.

Tab. II also includes data on how frequently each topic is observed in the ASRS data, in the form of expected topic proportion. The human factors topic is the topic found to be the best represented in ASRS records. Air traffic control issues including those related to the 'air traffic control' and airspace topics are also surprisingly common. The individual topics that describe mechanical issues, including, among others, those ascribed the engine, oil, pressure, fuel pump, tank, and landing gear labels are each relatively rare. This analysis has not revealed any single mechanical issue that frequently appears in aviation safety incident reports. The topic assigned the label surface is more common than the approach topic, which is more common than the climb topic.

The results analyzed here do not tell a complete story but do provide valuable reference data and a starting point for future aviation safety studies.

Fig. 4 presents a visualization of the correlations among the topics listed in Tab. II. Note that topic 3, assigned the intuitive labels smoke and fire is linked to topic 6, assigned the labels engine, oil, and pressure. Topic 3 has a substantial negative correlation with topic 9, the human factors topic. Topic 6 is also linked to topic 14, which is assigned the labels maintenance and fault.

C. *Impact of Covariates on Topic Prominence*

Tab. II lists expected topic proportions across all documents. We can also look at the effects of covariate data on topic proportions. In the aviation safety context, we can look at the prominence of topics in incident reports that arise in certain, specific situations. We can also study confidence intervals around estimated topic proportions and estimates of marginal effects.

Fig. 5 shows the estimated marginal effects of the phase of flight of the primary aircraft listed in the incident report for the air traffic control (ATC) and human factors topics. These and other charts were generated using the R package described earlier [20]. The dots on the chart depict the expected values of marginal effects while the horizontal bars illustrate confidence intervals.

When the aircraft is reported to be in the takeoff, cruise, or landing phases of flight, the ATC topic is more prominent than the human factors topic. In all other phases of flight, the human factors topic is more prominent than the ATC topic. This is particularly true when aircraft are reported to be on the surface of an airport.

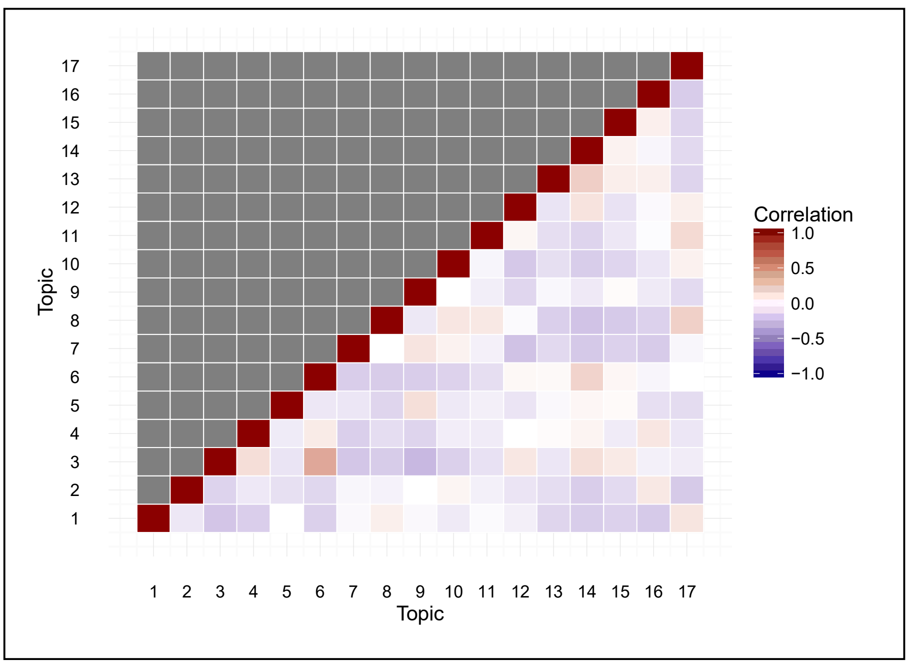

Figure 4: Correlations Among Topics.

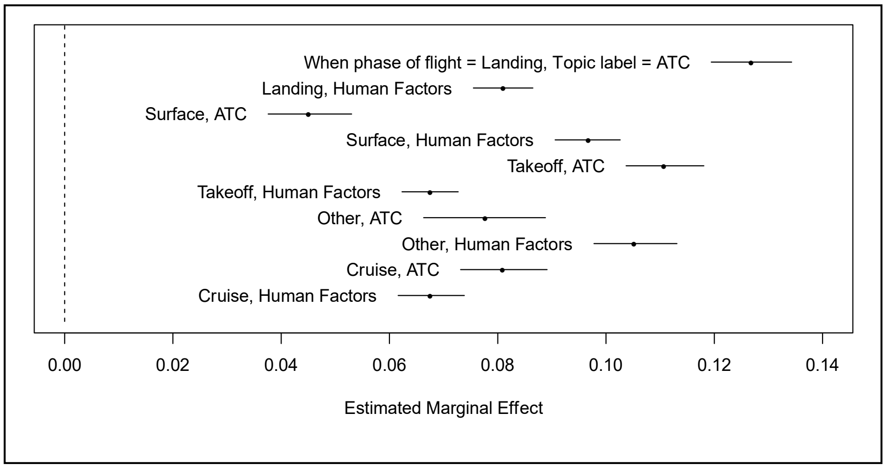

Figure 5: Phase of Flight and Estimated Topic Proportion.

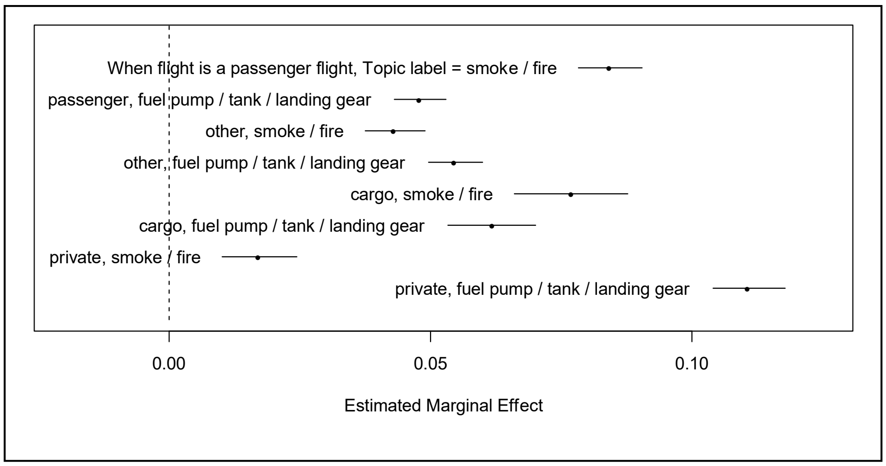

Figure 6: Flight Mission and Estimated Topic Proportion.

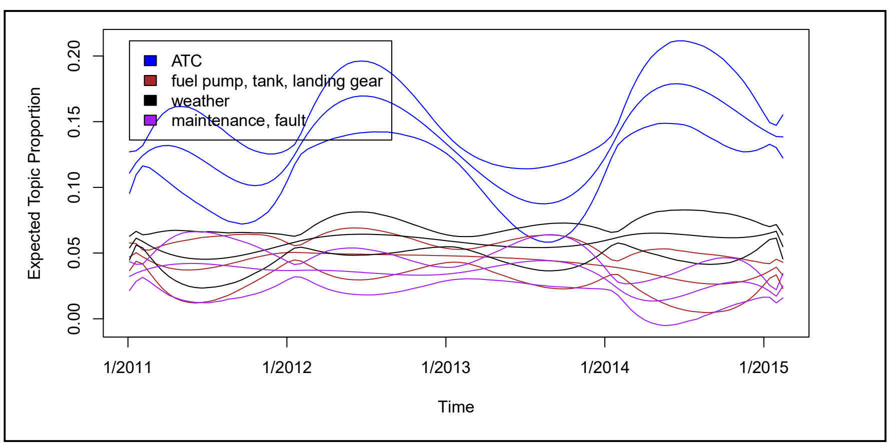

Figure 7: Estimated Topic Proportion over Time.

The results presented in Fig. 5 also show that the most prominent (phase of flight, topic) pairs involve the ATC topic and the takeoff and landing phases of flight. The least prominent pair involves ATC issues in reports issued involving an aircraft on the airport surface. This makes some intuitive sense as a primary challenge in ATC involves make efficient use of busy runways while ensuring wake vortex temporal separation standards among arriving and departing flights are met.

Fig. 6 shows the estimated effects of flight mission on the smoke, fire topic and the fuel pump, tank, landing gear topic proportions. Issues involving smoke and fire are more prominent for cargo and, particularly, passenger flights. Issues involving fuel pumps, tanks, and landing gear are more prominent for other flights, and particularly for private aircraft.

The data points that stand out the most in Fig. 6 reveal the significance of fuel pump, tank, and landing gear issues, and the insignificance of smoke / fire issues, in incident reports where the primary aircraft listed in the report is a private aircraft. A cursory inspection of a sample of the relevant reports reveals incidents where there was smoke or an unpleasant smell or an alarm from a smoke detector in the cabin. All of these issues would seem to be more likely on a larger commercial transport aircraft as opposed to a smaller private aircraft. For example, there would seem to be more opportunities for passengers or cargo to create issues on the larger commercial aircraft. Further analysis by subject matter experts would be helpful here, to come to more definitive conclusions.

Fig. 7 shows the estimated marginal effects of time on the prominence of topics 1, 4, 11, and 14 as listed in Tab. II. This chart shows a seasonal pattern in the prominence of the ATC topic. The topic appears to be more important during the spring and summer months and less important in fall, although 2013 is a notable exception. The other topics are included here to show that this pattern is unusual. The author cannot offer an intuitive explanation of this result.

Further research is needed to explore and extend the results presented here. For example, a more thorough inspection of incident reports would explain if and perhaps why there are more aviation safety incidents related to air traffic control but not weather in the Spring and Summer months. Is this simply a result of misunderstanding the meaning of two topics or indicative of something interesting regarding the prominence of air traffic control-related incidents in certain months of the year?

V. RESULTS: SAN FRANISCO INTERNATIONAL

The preceding sections of this article are based on an analysis of all of the ASRS records reported between January 2010 and April 2015. There might be more specific, actionable insights gained from analysis of particular subsets of this data. This section applies similar methods to data where the locale is listed as “SFO.Airport.” San Francisco International Airport (SFO) was chosen because of the large number of records linked to it and its familiarity to the author.

Tab. III shows the most common phrases, three words or longer, found in ASRS narratives linked to SFO. Many of the most commonly occurring phrases reference the San Mateo Bridge, include the word “visual,” or reference either “Quiet Bridge Visual” or “Tip Toe Visual.” The San Mateo Bridge is a well-known landmark on the routes of many aircraft arriving at runways 28L and 28R at SFO. The Quiet Bridge Visual and Tip Toe Visual are two of the approach paths into these runways that pass over the bridge. The phrases mentioned directly or indirectly relate to the use of Simultaneous Offset Instrument Approach (SOIA) procedures, which come up frequently in discussions of safety at SFO.

Tab. IV describes the topics identified in the narratives from SFO. (Results not shown here recommended the identification of seven topics.) The most prominent topic is assigned the labels visual and approach. Topics related to the taxiway, pushing, and holding / runways are individually less prominent but together point to (separate) issues on the surface at SFO.

Fig. 8 shows the correlations among the topics listed in Tab. IV. There isn’t evidence of any strong positive correlations. Topic 5, assigned the labels visual and approach, is negatively correlated with topics 3 (taxiway), 6 (push), and 7 (hold, runway). This highlights the difference between incidents on approach versus on the surface at SFO.

TABLE III. PHRASES IN ASRS NARRATIVES FROM SFO

Phrase	***Observation Count***
FMS bridge visual	38
bridge visual 28R	22
FMS bridge visual 28R	19
hold short line	19
air carrier X	18
maintain visual separation	16
Quiet Bridge Visual	15
visual approach runway	15
Tip Toe Visual	13
first officer FO	13
cross runway 28L	12
San Mateo Bridge	11
hold short runway	11
I pilot flying	11

TABLE IV. TOPICS IN ASRS NARRATIVES FROM SFO

Criteria	*Word 1*	*Word 2*	*Word 3*	*Label*	*Exp. Topic Prop.*
Prob Lift FREX	feet feet difficult	altitude fms rnav	arrive altitude quiet	**altitude**	0.17
	aircraft carrier air	carrier air carrier	departure separate nct	**separate**	0.12
	taxiway taxiway taxiway	taxi taxi taxi	ground ground ground	**taxiway**	0.10
	captain flap increase	flap captain gear	speed wind flap	**flap, gear**	0.12
	approach visual tcas	visual approach sight	aircraft sight visual	**visual, approach**	0.22
	flight push push	time flight maintain	call crew dispatch	**push**	0.13
	runway hold hold	aircraft runway line	tower takeoff across	**hold, runway**	0.14

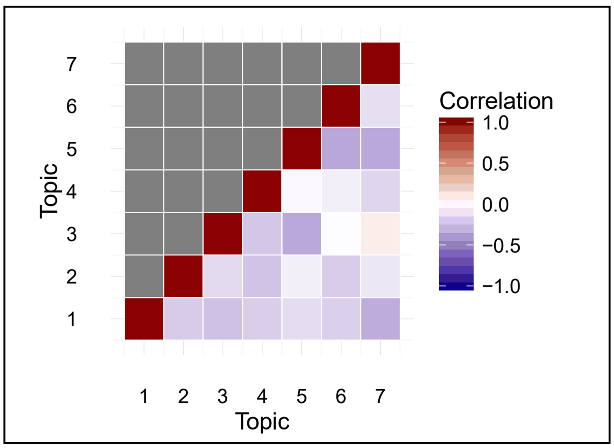

Figure 8: Topic Correlations at SFO.

Fig. 9 shows the estimated topic proportions for the taxi and visual, approach topics for different types of flights. It is interesting to note that passenger flights are more likely to report problems related to the approach topic. Private flights are (slightly) more likely to report problems related to the taxi topic. There is a relatively large amount of uncertainty when estimating topic proportions for private or cargo flights at SFO. This is because there are relatively few recent incident reports in the ASRS database at SFO where the primary aircraft is a cargo or private aircraft.

The application of natural language processing tools and techniques revealed the prominence of the Quiet Bridge Visual and Tip Toe Visual approach paths in reports of aviation safety incidents at SFO. Many traffic managers and analysts familiar with the airport will already know of issues related to the use of these approach paths. One benefit of applying machine learning is that it allows analysts to learn of the importance of specific approach paths and the like. Thus, machine learning can help focus discussion or set priorities for further analysis. It is also important to note that analysts can learn of features and topics for any airport or region of airspace without surveying the relevant subject matter experts. Analysts can also quantify, in an objective way, the prominence of the features and topics which they uncover in the data.

VI. CONCLUSION

The ASRS database is a useful resource for aviation systems researchers interested in safety. There are over a million incident reports in the database. Techniques developed in the fields of natural language processing and machine learning can be used for analysis of this database. This article describes applications of structural topic modeling to ASRS records from January 2010 through April 2015. STM was applied both to all the relevant records and to the subset consisting of reports from SFO.

Methods highlighted the Quiet Bridge Visual and Tip Toe Visual approach paths as particularly prominent in incident reports at SFO. Looking nationwide, results demonstrated the importance of human factors and air traffic control, with the former being more prominent on the airport surface and the latter more prominent during flight. The frequency of fuel pump, tank, and landing gear issues and the sparsity of smoke and fire issues for private aircraft were also recorded.

The results demonstrate that methods tested here are able to identify known issues. These methods are also able to uncover some issues that have not been previously reported, but do not necessarily provide detail that could be used to produce actionable insights. Subject matter expertise is important and needed to assign intuitive meanings to topics and to otherwise interpret the results of topic modeling efforts.

The degree to which the ASRS data are representative of all aviation safety incidents in the United States is unclear. It may be that certain types of incidents are over- or under-represented in the data. This issue warrants further study.

Analysts may want to apply structural topic modeling and similar techniques to specific subsets of the ASRS database or other data sets to produce more detailed and meaningful results. These methods provide a relatively fast and relatively automated way for analysts to identify areas where further, more detailed discussion and investigation are warranted. The results could also be used to set priorities when planning future aviation safety research.

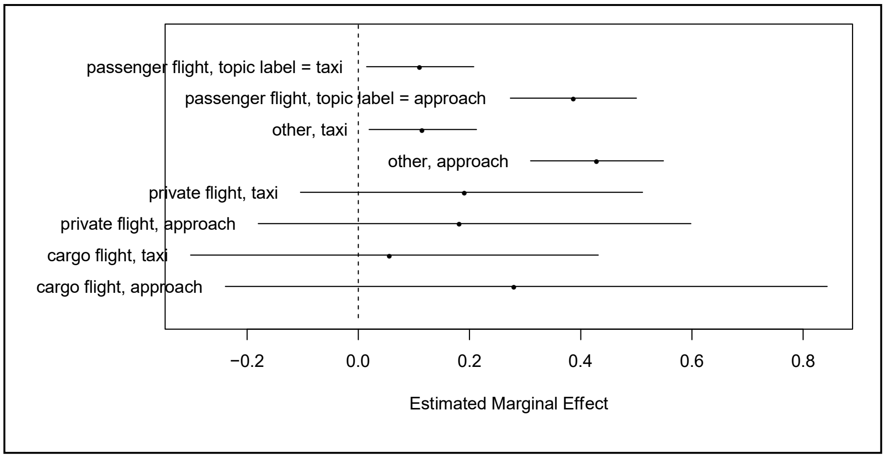

Figure 9: Flight Mission and Estimated Topic Proportion at SFO.

ACKNOWLEDGMENT

K. Kuhn thanks Heather Arneson, Jeremy Eckhause, Amy Kim, and the anonymous reviewers of an earlier draft of this paper.

REFERENCES

[1] Federal Aviation Administartion, NextGen Priorities Joint Implementation Plan, 2017-2019. Available online at: https://www.faa.gov/nextgen/media/ng_priorities.pdf.

[2] C. Fleming, N. Leveson, "Including safety during early development phases of future air traffic management concepts." ATM R&D Seminar, 2015.

[3] C. Fleming, N. Leveson, M.S. Placke, "Assuring safety of NextGen procedures." ATM R&D Seminar, 2013.

[4] A. Barnett, M. Ball, G. Donohue, M. Hansen, A. Odoni, A. Trani. "Collision course? The North Airfield Safety Study at Los Angeles International Airport (LAX)." Trans. Res. Part A, vol. 77, pp. 14-34, July 2015.

[5] National Aeronautics and Space Administration, ASRS Program Briefing. Available online at: https://asrs.arc.nasa.gov/docs/ASRS_ProgramBriefing2015.pdf.

[6] J. Bliss, M. Freeland, J. Millard. "Alarm related incidents in aviation: A survey of the Aviation Safety Reporting System database." Proc. of the Human Factors and Ergonomics Society Ann. Meeting, vol. 43(1), pp. 6-10, 1999.

[7] C.E. Billings, W.D. Reynard. "Human factors in aircraft incidents: results of a 7-year study." Aviation, Space, and Envir. Med., vol. 55(10), pp. 960-965, 1984.

[8] S.D. Robinson, "Visual representation of safety narratives." Safety Science, vol. 88, pp. 123-128, 2016.

[9] M.S. Ahmed, L. Khan, N. Oza, M. Rajeswari, "Multi-label ASRS dataset classification using semi supervised subspace clustering." Conf. on Intel. Data Underst., pp. 285-299, 2010.

[10] M.A. Abedin, V. Ng, L. Khan, "Cause identification from aviation safety incident reports via weakly supervised semantic lexicon construction." Journ. of Artif. Intel. Res., vol. 38, pp. 569-631, 2010.

[11] C. Posse, B. Matzke, C. Anderson, A. Brothers, M. Matzke, T. Ferryman, "Extracting information from narratives: An application to aviation safety reports." IEEE Aero. Conf., pp. 3678-3690, March 2005.

[12] D.M. Blei, "Probabilistic topic models." Communications of the ACM, vol. 55(4), pp.77-84, 2012.

[13] L. El Ghaoui, V. Pham, G.C. Li, V.A. Duong, A. Srivastava, K. Bhaduri, "Understanding large text corpora via sparse machine learning." Stat. Ana. and Data Mining, vol. 6(3), pp. 221-242, 2013.

[14] L. Tanguy, N. Tulechki, A. Urieli, E. Hermann, C. Raynal, "Natural language processing for aviation safety reports: from classification to interactive analysis." Computers in Industry, vol. 78, pp. 80-95, 2016.

[15] D. Blei, A. Ng, M. Jordan, "Latent dirichlet allocation." Journ. of Machine Learning Res., vol. 3, pp. 993-1022, 2003.

[16] J. Farrell, "Corporate funding and ideological polarization about climate change." Proc. of the Nat. Acad. of Sciences, vol. 113(1), pp. 92-97, 2016.

[17] A. Mishler, E.S. Crabb, S. Paletz, B. Hefright, E. Golonka, "Using structural topic modeling to detect events and cluster Twitter users in the Ukrainian crisis." Int. Conf. on Hum.-Comp. Inter., pp. 639-644, August 2015.

[18] J. Reich, D.H. Tingley, J. Leder-Luis, M.E. Roberts, B. Stewart, "Computer-assisted reading and discovery for student generated text in massive open online courses." Journ. of Learning Analytics, vol 2(10, pp. 156-184, 2015.

[19] M. Roberts et al., "Structural topic models for open-ended survey responses." Amer. Journ. of Pol. Sci., vol. 58(4), pp. 1064-1082, 2014.

[20] M.E. Roberts, B.M. Stewart, D. Tingley, D., "stm: R package for structural topic models." R package. 2014.

[21] D. Mimno, H.M. Wallach, E. Talley, M. Leenders, A. McCallum, "Optimizing semantic coherence in topic models." Proc. of the Conf. on Emp. Methods in Nat. Lang. Proc., pp. 262-272, July 2011.

[22] J. Bischof, E.M. Airoldi, "Summarizing topical content with word frequency and exclusivity." Proc. of the 29th Int. Conf. on Mach. Learning, pp. 201-208, 2012.

AUTHOR BIOGRAPHY

Kenneth Kuhn received a B.A. in Math in 2001 from Johns Hopkins University and both a M.S. in Industrial Engineering and Operations Research in 2002 and a Ph.D. in Civil and Environmental Engineering in 2006 from the University of California, Berkeley. He has worked at the National Aeronautics and Space Administration as an Aerospace Engineer and at the University of Canterbury in Christchurch, New Zealand as a Lecturer. Kenneth currently works as an Operations Researcher at the RAND Corporation.