Federal Aviation Administration

Runway Safety Metric

Weighting Scheme

Firdu Bati, PhD
Fong Lee
Yunsun Li, PhD
Nicholas Scialli

AJI-3
ATO Safety & Technical Training

August 2017

This paper outlines a weighting technique used to develop a Runway Safety Metric that can serve as a monitoring tool for the safety performance of the National Airspace System in the runway environment. It provides an overview of the data preparation tasks, mathematical modeling approach, assumptions based on domain knowledge, and derivation of quantitative measures from relevant accident and incident data.

Contents

1. Introduction 2
2. Data Pre-Processing 2
3. Three-Step Outcome–Based Weighting 3
   3.1. Accident Weights 3
   3.2. Reordering Accident Weights 5
   3.3. Incident Weights 6
4. Application of Weights 11
   4.1. Penalty Term 11
   4.2. Credit Term 11
   4.3. Weights for Incidents 12
   4.4. Combining Penalty and Credit Terms 12
5. Result 13
   5.1. NAS-Level Runway Safety Performance 13
   5.2. Commericial vs Non-Commerical Runway Safety Performance 14
6. Conclusion 16
References 17

Figures

Figure 1: Augmented Naïve Bayes Implementation for Accidents 5
Figure 2: Commercial Accident Relative Weights 8
Figure 3: Commercial Incident Relative Weights 8
Figure 4: Non-Commercial Accident Relative Weights 9
Figure 5: Non-Commercial Incident Relative Weights 9
Figure 6: Simplified Schematic of Weight Calculation by Event Type 10
Figure 7: Moving Average of Commercial and Non-Commercial Weights 14
Figure 8: Moving Average of Commercial Weights 15
Figure 9: Moving Average of Non-Commercial Weights 16

Tables

Table 1: Frequency of Events by Accident and Incident Type 3
Table 2: Commercial and Non-Commercial Fatality Information Gain 5
Table 3: Adjusted Information Gain as Relative Accident Weights 6
Table 4: Incident to Accident Mapping 7
Table 5: Final Commercial and Non-Commercial Relative Weights 7

1. INTRODUCTION

The primary goal of developing a Runway Safety Metric (RSM) is to measure the safety performance of the National Airspace System (NAS) in the runway environment. Unlike existing metrics that focus primarily on Runway Incursions (RIs), the RSM proposed in this paper incorporates all types of relevant runway safety events (accidents and incidents) in the runway environment.

The RSM assumes that the worst possible outcome involves a fatal injury. Therefore, it assigns the highest severity weight to a fatal accident, and all other event types are measured according to their relative "proximity" to a fatal accident. As such, the RSM incorporates all relevant types of events that can occur in the runway environment, including Runway Collisions (RCs), RIs, Runway Excursion (RE) accidents, RE incidents, Surface Collisions (SCs), and Surface Incidents (SIs). By incorporating every type of runway safety event, the RSM reflects the overall safety of the NAS in the runway environment.

The potential downside of the RSM is its sensitivity to fatal accidents; most incidents have a very low probability of becoming a fatal accident. As a result, during periods when fatal accidents are rare, the RSM shows that the system performs at a very high safety performance level. However, with only one fatal accident, the RSM shows a severe decline in the safety performance level[1].

The following activities were performed to gather and prepare all relevant datasets, assign relative weights to event types, and quantify safety performance:

- Runway safety–related event data was extracted, merged, and cleaned.
- A severity weighting scheme was developed for all possible runway safety–related event outcomes in relation to fatal accidents.
- The weighting scheme was applied to historical accident and incident data to quantify safety performance.

2. DATA PRE-PROCESSING

The dataset used for this analysis was derived from the National Transportation Safety Board (NTSB) aviation accident database, the Runway Incursion–Surface Incident (RI-SI) database, and the RE database[2]. For NTSB data, occurrence and phase codes were used to remove events that were not relevant to runway safety analysis. The narratives associated with the remaining NTSB accidents were categorized using a supervised classification model[3], and a manual review was performed to validate the model output. RE and RI-SI data were then imported and combined with the NTSB accident data, and the consolidated data became the main dataset for developing the RSM. Table 1: Frequency of Events by Accident and Incident Type shows the number of events by accident and incident type. A total of 34,457 relevant events were collected from the three data sources. Data through the end of FY 2016 were gathered from each source.

[1]This may not be a limitation, but rather a desired feature for any safety metric. A good metric should capture all undesired outcomes, and it is counterintuitive that the existing metric does not say anything about the most severe cases that tend to be discussed in the public domain.
[2]Runway events that do not involve RIs and REs were excluded from this analysis. Such events include, collisions with the ground, ground loops, etc.
[3]A text mining solution, based on Natural Language Processing, was employed to extract features for supervising the machine learning classification model.

NTSB Database				
Runway Collisions		**Taxiway Collisions**		**Runway Excursions**
324		285		9,520
RI-SI Database				
Category A RIs	**Category B RIs**	**Category C RIs**	**Category D RIs**	**Surface Incidents (SIs)**
237	319	7,170	6,288	9,073
RE Database				
Runway Excursions				
1,241				

Table 1: Frequency of Events by Accident and Incident Type

Operational data and domain knowledge show that runway safety performance varies greatly between commercial and non-commercial operations. Therefore, the dataset was divided into two operational categories: commercial and non-commercial. All Federal Aviation Regulation (FAR) Part 121, 129, and 135 operations were assumed to be commercial, and all other operations were assumed to be non-commercial. Separate weights and metric values were calculated for each category.

3. THREE-STEP OUTCOME–BASED WEIGHTING

The RSM assigns different levels of severity weight to each event type proportional to its proximity to a fatal accident. The weight is the result of a combination of quantitative measures derived from data and assumptions appropriate for the runway safety domain. This technique employs the following three-step approach to calculate the relative weight for each type of event:

I. Weights are assigned to accidents according to their proximity to a fatal accident
II. Assumptions appropriate for the runway safety domain are applied to reorder the weights.
III. Weights are assigned to incident types according to the weights of their corresponding accident types.

3.1. ACCIDENT WEIGHTS

Weights are assigned to accidents (i.e., events with outcomes such as injury and/or aircraft damage) according to their proximity to a fatal accident by using *information gain*. Information gain is the amount of information obtained about one variable when given information about another. This approach considers how the uncertainty of a fatal outcome is affected by observing a non-fatal injury and the extent of aircraft damage. A formal quantification of uncertainty in information theory is *entropy*. Entropy can be used to quantify the uncertainty exhibited in a probability distribution of a fatal accident. For example, knowing the extent of aircraft damage would reduce the uncertainty of a fatality in an accident by some measureable amount, and entropy could be used to quantify it. The entropy can be calculated from the marginal (unconditional) probability of a fatal accident.

The entropy for a discrete distribution can be calculated using its formal definition as,

$$H(X) = -\sum_{x \in X} P(x) log_2 P(x),$$

where X represents a variable of interest.

On its own, the calculated entropy may not be a meaningful metric. To measure the amount of uncertainty it represents, entropy is usually compared to “no uncertainty” and “complete uncertainty” conditions of the variable of interest. For the purpose of fatality distribution, the “no uncertainty” condition involves a situation in which the outcome of fatality is known in a deterministic way: either

100 percent certainty of fatality or 100 percent certainty of no fatality. In the complete uncertainty condition, the probability of either fatality or no fatality is uniformly distributed (i.e., a 0.5 probability value for each outcome). The no uncertainty condition has zero entropy, whereas complete uncertainty results in the highest entropy.

A closely related concept to entropy is *Mutual Information*. Mutual Information is the amount of information gained (i.e., the amount of uncertainty reduced) about a variable given related information about another, and vice versa. It can be used to rank the importance of one variable relative to another (i.e., "closeness"). For example, when fatality is a target, the conditional entropy between Aircraft Damage (*AD*) and Fatal Accident (*FA*) can be measured using entropy as,

$$H(FA|AD) = \sum_{i} P(FA_i)H(FA_i|AD_i)$$

The Mutual Information (*I*) can be calculated as the difference between the marginal entropy of the fatal accident and the conditional entropy of *FA* given *AD*.

$$I(FA, AD) = H(FA) - H(FA|AD)$$

This is equivalent to,

$$I(FA, AD) = \sum_{fa \in FA} \sum_{ad \in AD} p(fa, ad) log_2 \frac{p(fa, ad)}{p(fa)p(ad)}$$

The accident type with the highest mutual information relative to *FA* provides the maximum information gain, which results in a higher weight. The information gain (*I*) in the preceding formulation corresponds to the following outcomes and events: serious injury, minor injury, destroyed damage, substantial damage, minor damage, RC, taxiway collision, and RE.

In practice, some outcomes may not overlap, which renders the unrealistic value of zero probability. For instance, in the commercial dataset, no fatality has occurred with only minor aircraft damage; therefore, direct calculation of the entropy is impractical. A conditional dependence paradigm such as a Bayesian network can resolve this issue by calculating marginal probabilities. Therefore, in the preceding example, the probability of a fatality given minor damage is summed over all the probabilities of other outcomes (i.e., other damage types and non-fatal injuries) correlated with minor damage. In addition to mitigating the zero-probability problem, the modeling approach automates the calculation of the information gain for each pair to avoid tedious manual manipulation. A Tree Augmented Naïve Bayes structure learning algorithm was employed to develop the Bayesian network and calculate the information gain directly from the model. Figure 1: Tree Augmented Naïve Bayes Implementation for Accidents depicts the Naïve Bayes model used to automate this formulation.

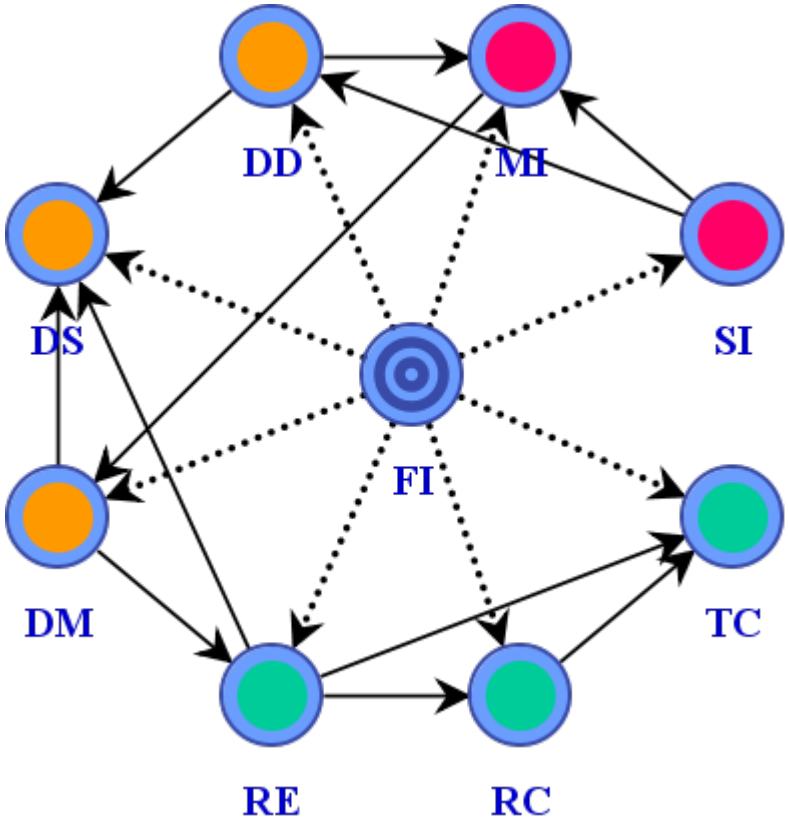

Figure 1: Tree Augmented Naïve Bayes Implementation for Accidents

Table 2: Commercial and Non-Commercial Fatality Information Gain provides the information gain values for both the commercial and non-commercial datasets based on the implementation of the Bayes model.

Outcome	Information Gain (Commercial)	Information Gain (Non-Commercial)
Fatal Injury	1	1
Serious Injury	0.24418	0.069573
Minor Injury	0.108792	0.000276
Destroyed Damage	0.459994	0.264379
Substantial Damage	0.106845	0.19216
Minor Damage	0.061505	0.000283
Runway Collision	0.13936	0.016456
RE Accident	0.033363	0.009677
Taxiway Collision	0.034145	0.000449

Table 2: Commercial and Non-Commercial Fatality Information Gain

3.2. REORDERING ACCIDENT WEIGHTS

The mathematical output shows that the extent of aircraft damage is more indicative of a fatal outcome than non-fatal injuries are; the information gain, therefore, is a pure risk indicator. However, part of the goal of the RSM is to communicate with the public about the overall safety performance of the NAS in the runway environment. This objective assumes that life is more valuable than property, and the RSM should reflect this. Therefore, to align the RSM with this objective, the weighting scheme assigns a higher weight to any type of injury than to any type of aircraft damage, and it assigns a lower weight to any undesired incident than to any type of accident.

The Bayes model calculates the information gain of each outcome with a fatal accident. During the reordering logic, the relative order of the information gain is maintained while enforcing a higher weight to injury than to aircraft damage. After the logical ordering is set, a cumulative risk is calculated to enforce the ordering. The following steps outline the ordering heuristic implemented using cumulative gain and relative significance:

1. Rank each outcome based on the information gain.
2. Calculate the relative information gain (IR) with respect to the highest information gain:

$$IR_i = {}^{I_i}/_{I_h},$$

where I_i is the information gain of outcome *I* and I_h is the highest information gain.

3. Normalize the relative information gain so that it sums to 1, which aligns with the logical use of severity weights.
4. Calculate cumulative information gain *(IC)* and order the outcomes according to the result:

$$IC_i = \sum_{j}^{i-1} IR_j$$

5. Calculate the relative significance to the new highest information gain, which is equivalent to performing step 2 using the result from step 4.
6. Shift the significance by the maximum normalized information gain to calculate the relative significance to fatality:

$$W_i = IC_i - IC_i * IC_{max}$$

Table 3: Adjusted Information Gain as Relative Accident Weights shows the adjusted information gain used as weights for both the commercial and non-commercial datasets using this reordering methodology.

Outcome	Relative Weights (Commercial)	Relative Weights (Non-Commercial)
Fatal Injury	1	1
Serious Injury	0.612859624	0.52213725
Minor Injury	0.30395762	0.420910201
Destroyed Damage	0.157139533	0.103588002
Substantial Damage	0.141287916	0.066919142
Minor Damage	0.133533745	0.065660114
Runway Collision	0.390175896	0.422260574
Taxiway Collision	0.129585506	0.065693717
RE Accident	0.125946876	0.065747103

Table 3: Adjusted Information Gain as Relative Accident Weights

3.3. INCIDENT WEIGHTS

For application of the RSM, when an event does not result in injury or aircraft damage, it is categorized as an incident. In runway safety, relevant incident types include RIs, REs, and SIs. The incident dataset is independent of the NTSB accident data, so no overlap exists between the two

(except in very few cases, which are categorized as duplicates). Therefore, it is impractical to leverage information gain or statistical correlation between various incident types and fatal accidents in a direct way. However, accidents and incidents may fall under similar categories. Based on domain knowledge, most event types have some probability of becoming accidents. For instance, RIs have some probability of becoming RCs. Therefore, by mapping the incident types to the appropriate accident types, weights can be indirectly assigned to incidents. Table 4: Incident to Accident Mapping shows the mapping used to implement this approach.

Incident Type	Accident Type
RI (Category A, B, C, and D)	Runway Collision
SI	Taxiway Collision
RE Incident	RE Accident

Table 4: Incident to Accident Mapping

Because incidents are less severe than accidents, an adjustment term must be introduced to complete the mapping. For each incident type, a reasonable adjustment term is the inverse frequency of that incident type. Therefore, incident weights are calculated as,

$$w_{Inc} = w_{Acc} * \left(\frac{1}{n_{Inc}}\right)$$

This calculation was performed and implemented for the RSM. Table 5: Final Commercial and Non-Commercial Relative Weights shows the final set of relative weights for all relevant event types used in this analysis. Figure 2: Commercial Accident Relative Weights through Figure 5: Non-Commercial Incident Relative Weights show the relative weights for the commercial and non-commercial datasets. The RC, RE Accident, and Taxiway Collision weights were used only to derive incident weights, so they are omitted from both the table and figures.

Outcome	Relative Weights (Commercial)	Relative Weights (Non-Commercial)
Fatal Injury	1	1
Serious Injury	0.612859624	0.52213725
Minor Injury	0.30395762	0.420910201
Destroyed Damage	0.157139533	0.103588002
Substantial Damage	0.141287916	0.066919142
Minor Damage	0.133533745	0.065660114
Category A RI	0.003306575	0.002639129
Category B RI	0.002393717	0.001982444
RE Incident	0.001701985	7.79918E-05
SI	8.26438E-05	8.45261E-06
Category C RI[4]	0.000424104	8.95379E-05
Category D RI	0.000108172	7.86626E-05

Table 5: Final Commercial and Non-Commercial Relative Weights

[4] For the commercial dataset, more Category C RIs were reported than Category D RIs. Category C RIs are known to be riskier than Category D RIs, so the weights were reversed for these two categories.

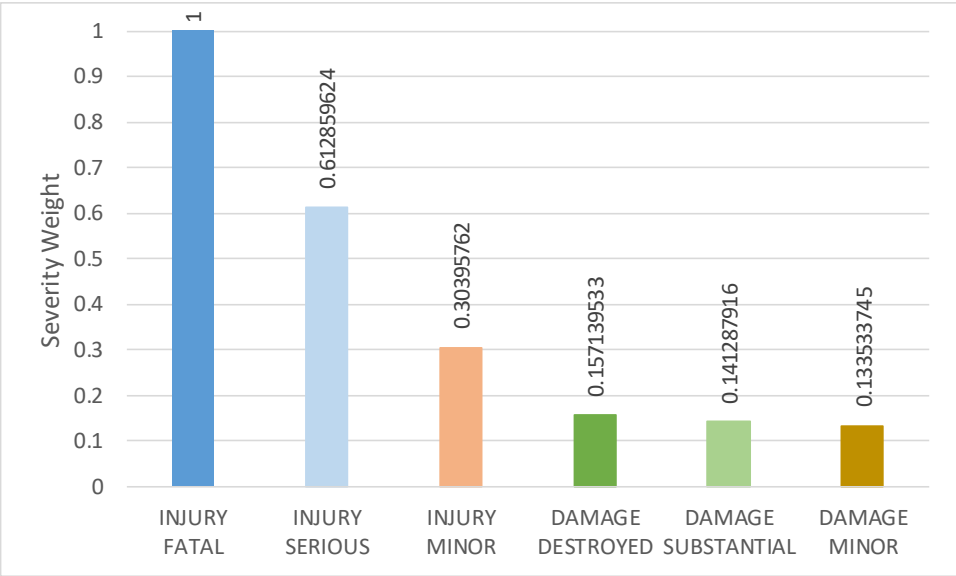

Figure 2: Commercial Accident Relative Weights

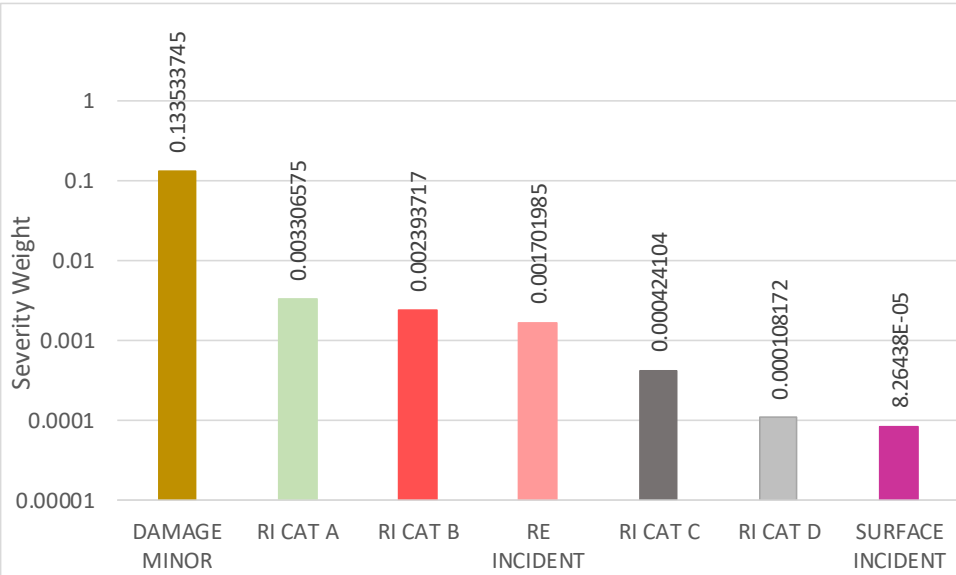

Figure 3: Commercial Incident Relative Weights

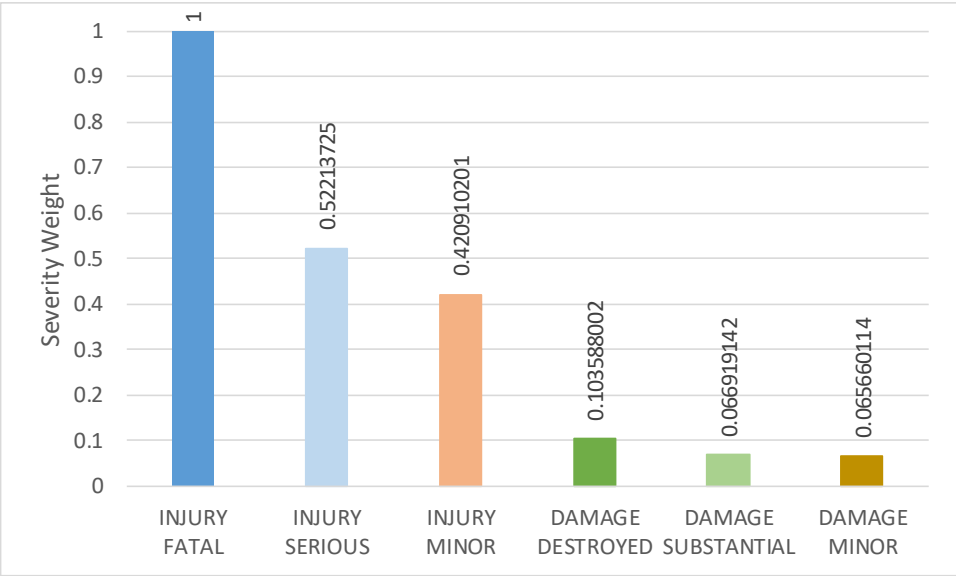

Figure 4: Non-Commercial Accident Relative Weights

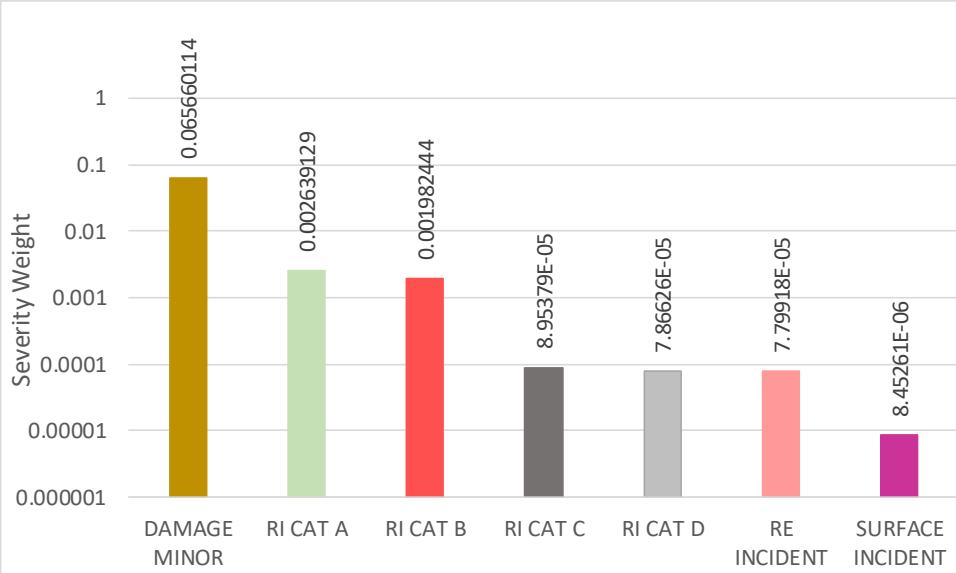

Figure 5: Non-Commercial Incident Relative Weights

As specified in the preceding sections, the RSM was developed using multistep data preprocessing and severity weight calculations. The process involved both quantitative analysis and qualitative constraints. Figure 6: Simplified Schematic of Weight Calculation by Event Type is a simplified schematic that demonstrates the overall process used for deriving the RSM weights.

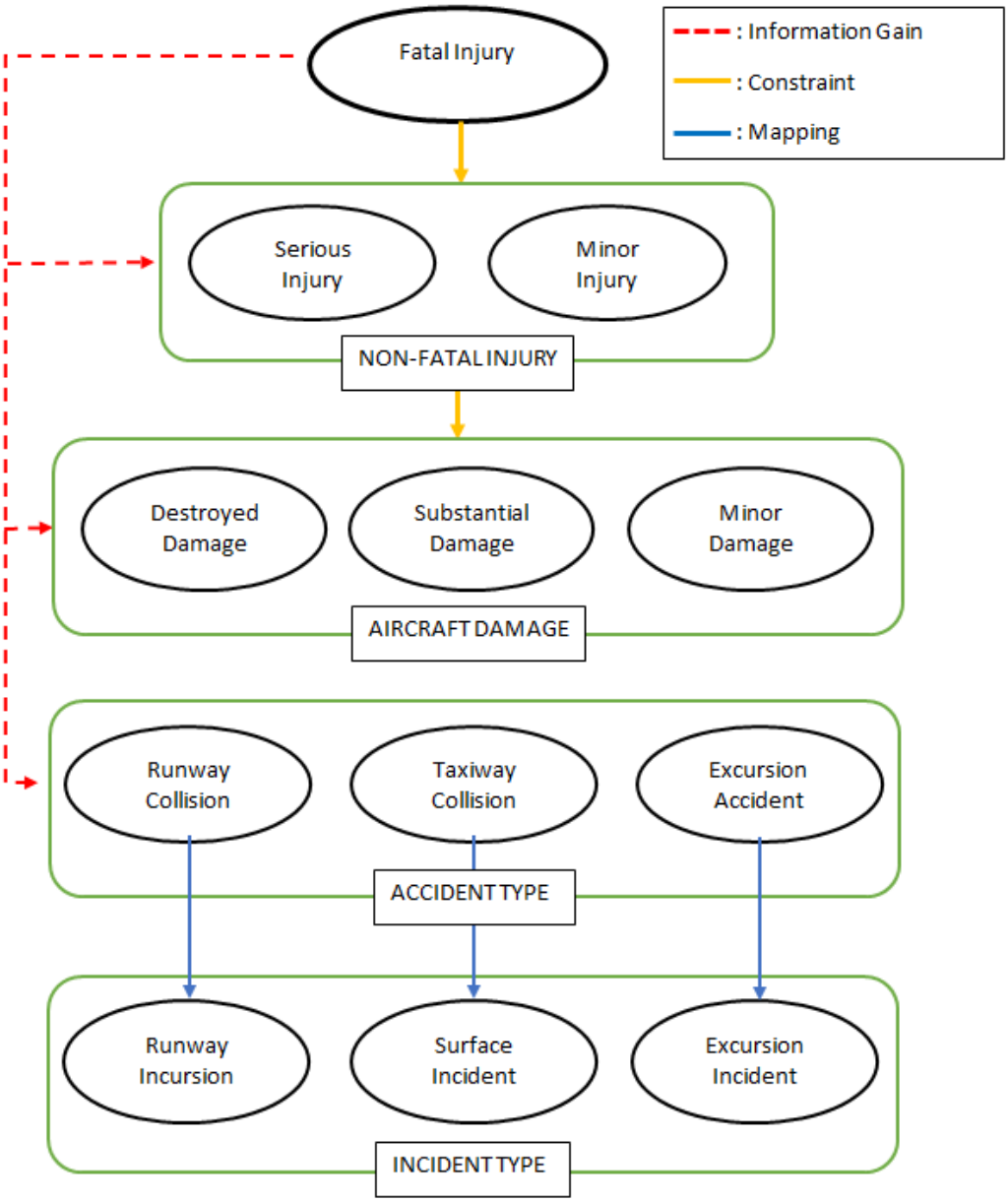

Figure 6: Simplified Schematic of Weight Calculation by Event Type

4. APPLICATION OF WEIGHTS

Weights are calculated for the individual consequences of an event; that is, a corresponding weight exists for each type of injury and extent of aircraft damage. However, each incident type is assigned a single weight because there is no consequence. Therefore, the weight application differs slightly depending on the type of event; that is, events with injuries would consider the number of injuries, and events with aircraft damage would consider the number of damaged airframes. Weights can be applied to accidents in two ways: aggregation, which accounts for all consequences of the event; and worst outcome–based weighting, which assigns only the weight corresponding to the worst consequence.

When applying weights using aggregation, all possible outcomes of an event are considered, and the combined weight is calculated by summing over all the individual weights. For instance, if an accident involves both injury and aircraft damage, weight is assigned for each outcome and the outputs are summed, which results in an aggregate severity index. The advantage of this approach is that all undesirable outcomes are considered. In the worst outcome approach, each event is assigned the weight corresponding to the most severe outcome. For instance, if an accident involves both injury and aircraft damage, it is only assigned the weight corresponding to injury because injury is more severe than aircraft damage. This approach is simple, but it neglects less severe outcomes for the same event. For the purpose of comprehensiveness, the aggregation approach was employed for the RSM.[5]

4.1. PENALTY TERM

In this application, penalty is a term that indicates the overall severity of an event. It is a direct measure of all the consequences of the event. In the aggregation of weights, the weights are summed over all the undesired outcomes. For instance, for an event involving both injuries and aircraft damage, the product of weight for each injury type and number of injuries is added to the product of weight for each damage type and number of damaged aircraft. Thus, the penalty term (P) is calculated as,

$$P = \sum_{i} w_i n_i + \sum_{d} w_d n_d,$$

where w_i is weight *f* for the type of injury, n_i is the number of injuries, w_d is weight for the type of damage, and n_d is the number of damaged aircraft.

The penalty term does not account for the number of people or number of aircraft involved in an event; it reflects only the number of people and aircraft affected by the event.

4.2. CREDIT TERM

A safety performance metric should account for the deficiency of the system as well as the safety improvements and technological advances that contribute to reduction in runway safety event consequences. Over the years, numerous procedural and technological solutions have been deployed in the NAS to reduce the frequency and impact of safety event consequences. Such solutions for the runway environment include Airport Surface Detection Equipment (ASDE) and the Engineered

[5] The choice between worst outcome and all outcome aggregates makes a small difference in the relative risk profile primarily due to the limiting effect of the worst outcome (i.e., adding small numbers to a large number does not significantly change the final result).

Materials Arresting System (EMAS). ASDE is a surveillance system that uses radar, multilateration, and satellite technologies to enable air traffic controllers to track surface movement of aircraft and vehicles. Through early detection of conflicts on the runway, ASDE minimizes the severity of runway events and reduces the likelihood of more severe events and potential collisions. Similarly, EMAS is placed at the end of runways and breaks down under the weight of an aircraft to minimize human injury and aircraft damage.

This RSM is designed to account for such improvements indirectly by incorporating a credit term proportional to non-injured people and non-damaged aircraft involved in an event. The RSM was developed so that events with fewer people injured in proportion to the total number of people onboard and minimally damaged aircraft are assigned relatively smaller weights using a credit term. The credit term is calculated as,

$$C = \sum_i w_i \frac{n_O - n_i - n_{ci}}{n_O} + \sum_d w_d \frac{n_A - n_d - n_{cd}}{n_A},$$

where:

- n_O is the number of people on board;
- n_i is the number of injuries;
- n_{ci} is the number of people already credited at a higher injury level;
- n_A is the number of aircraft involved in the event;
- n_d is the number of damaged aircraft; and
- n_{cd} is the number of aircraft already credited at a higher damage level

The credit term therefore reduces the penalty term assigned to an event proportional to the extent of non-injured people and undamaged aircraft.[6]

4.3. WEIGHTS FOR INCIDENTS

Incidents do not involve injuries or damage. Therefore, the penalty term for an incident calculates the weights as adjusted information gain for each type of incident. The credit term does not apply to incidents.

4.4. COMBINING PENALTY AND CREDIT TERMS

The final severity measure of an event is the difference between the total penalty term and the total credit term. The total penalty term is the sum of all penalty terms of injuries and aircraft damage. Similarly, the total credit term is the sum of the proportion of unaffected people on board and undamaged aircraft. The severity index is calculated as,

$$S = (P_i + P_d) - (C_i + C_d),$$

where P_i is the total penalty term for injuries, P_d is the total penalty term for damaged aircraft, C_i is the total credit term for non-injuries, and C_d is the total credit term for undamaged aircraft.

[6] In this application, a credit term is only defined when there is a penalty term; that is, there is no credit term for events without consequences.

5. RESULT

This RSM demonstrates a critical element of domain knowledge: although events still happen (and in some cases at relatively the same rate as before), their severity has decreased significantly. Even though the number of reported accidents is relatively unchanged and the number of reported incidents has increased, the NAS has become safer overall. There are two potential explanations for this. First, the solutions that have been recently deployed in the NAS (e.g., training, procedures, technological advances) are collectively resulting in significant improvements. Second, the adoption of a Just Culture that promotes increased reporting of unsafe actions and incidents has resulted in a highly efficient, continuously improving system. Therefore, even though the number of reported incidents is steadily increasing, the overall severity of events is trending downward. A fatal commercial accident in the runway environment has not occurred in nearly a decade. Thus, an increasing number of reported incidents may be required to increase general knowledge and improve the overall safety performance of the NAS. As the RSM demonstrates, the increasing number of reported incidents is not necessarily a negative outcome and, therefore, optimal safety decisions cannot be based only on the number of incidents.

Note: *The RSM is sensitive to more severe events. If the agency decides to use this as a formal metric to evaluate the safety performance of the overall system, it should be clearly communicated that the measure is sensitive to catastrophic outcomes. That is, one fatal accident puts the index much higher than a large number of less severe events.*

5.1. NAS-LEVEL RUNWAY SAFETY PERFORMANCE

Figure 7: Moving Averages of Commercial and Non-Commercial Weights shows a moving average of the aggregated weights applied to the entire dataset (both commercial and non-commercial operations) since FY 1997. The solid line shows calculated weight over time, the dashed line depicts a trend of that weight, and the bars show accident and incident counts over time. Figure 7 clearly shows that the increase in the number of incidents is accompanied by a decrease in the overall severity of accidents.

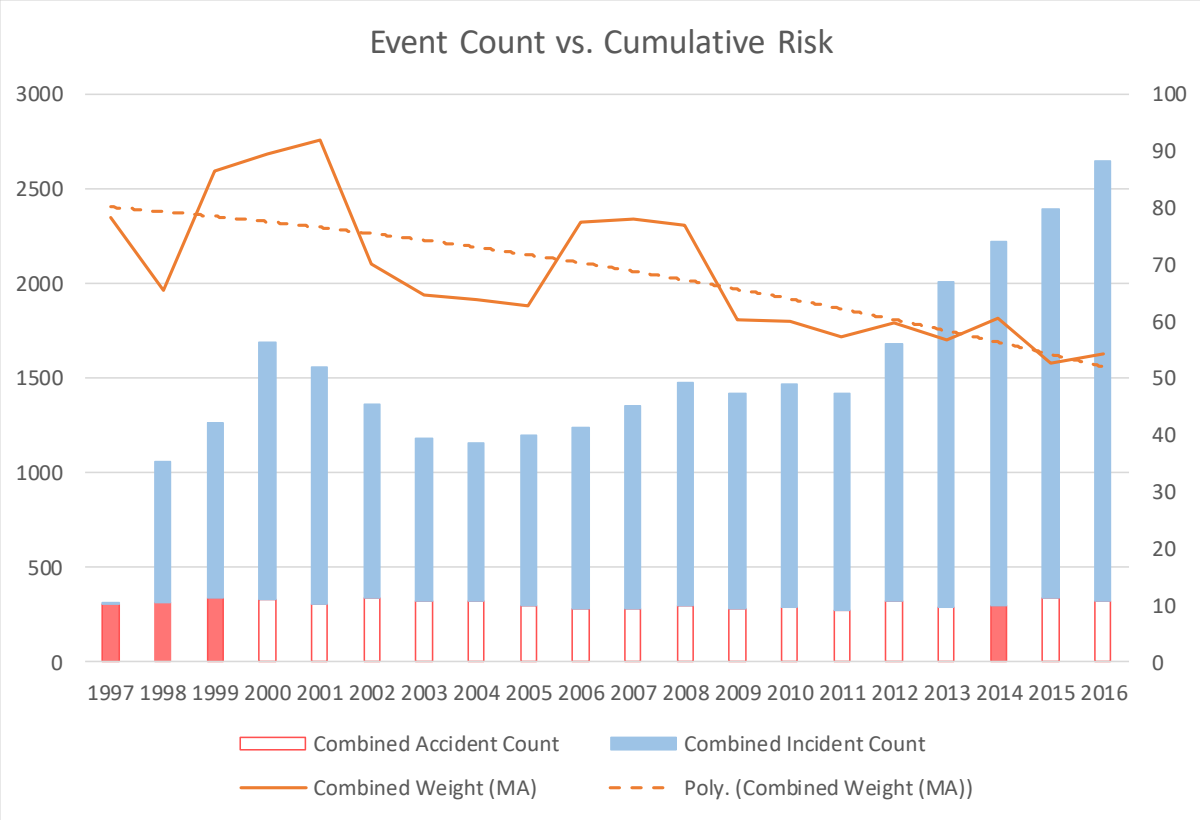

Figure 7: Moving Averages of Commercial and Non-Commercial Weights

5.2. COMMERICIAL VS NON-COMMERICAL RUNWAY SAFETY PERFORMANCE

The safety performance of commercial and non-commercial operations is known to be vastly different.[7] For various reasons (e.g., pilot proficiency and aircraft design), commercial operations are considerably safer than non-commercial operations. Therefore, the safety performance of the two environments should be considered independently.

Figure 8: Moving Average of Commercial Weights shows moving average of the normalized RSM severity index since FY 1997 for commercial operations. The figure clearly shows that although the number of operations has slightly decreased, the overall severity has significantly decreased. The dotted curve is an exponential fit that shows the average downward trend in overall severity.

[7] "Commercial" is defined as flight operations including FAR Part 121, 129, and 135. All other operations are considered non-commercial.

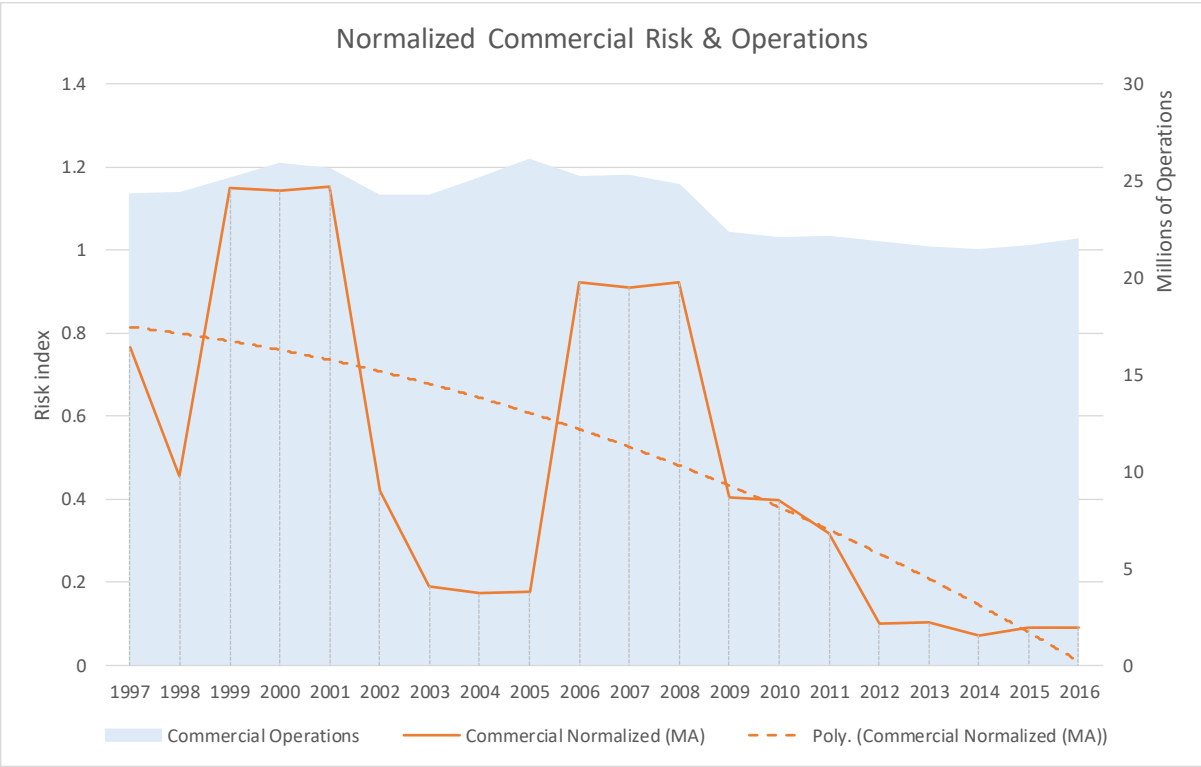

Figure 8: Moving Average of Commercial Weights

Figure 9: Moving Average of Non-Commercial Weights shows moving averages of the normalized RSM severity index for non-commercial operations. In contrast to commercial safety, overall non-commercial safety performance declined from FY 2009 to FY 2014. This decline can be partially attributed to the slight decrease in the number of accidents and their severities, which is disproportionate to the significant decrease in the number of operations. Additionally, RE tracking has increased since 2013, which may also have contributed to the recent rise in the non-commercial RSM severity index.

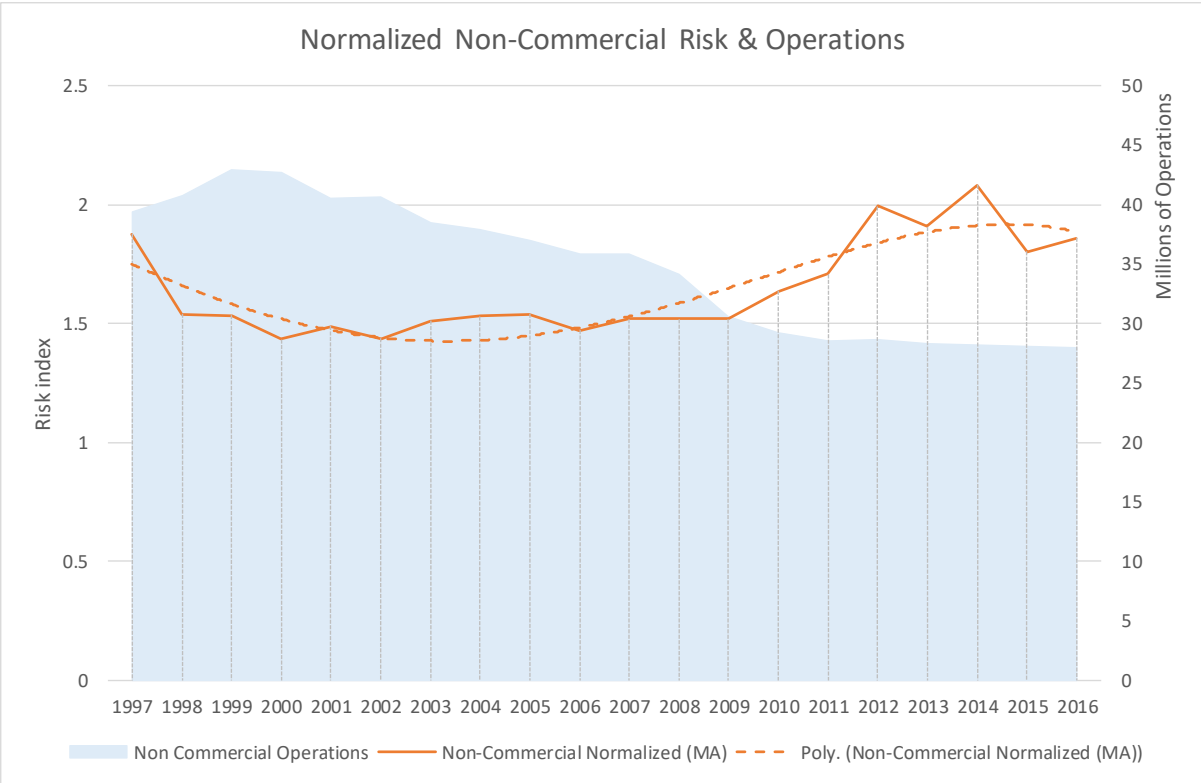

Figure 9: Moving Average of Non-Commercial Weights

6. CONCLUSION

This analysis followed a systematic approach to develop a RSM for evaluating the safety performance of the NAS in the runway environment. All relevant datasets were incorporated, appropriate assumptions were applied, and a quantitative technique was used to assign severity measures to runway safety events proportional to their proximity to a fatal accident. The RSM combines two quantitative measures: a penalty term, which is proportional to the extent of undesired outcomes in an event; and a credit term, which is proportional to mitigating the outcome. The severity weights can be applied to analyze the overall performance of the NAS over time in two ways: by considering only the worst outcome, and by aggregating all undesired outcomes. The RSM employs aggregation by accounting for all possible outcomes. Because the severity of events was measured against a fatal accident, the RSM is sensitive to events involving fatal injuries.

REFERENCES

Read, J. Pfahringer, B. Holmes, G. (2008): *Multi-label Classification Using Ensembles of Pruned Sets.* Data Mining, Eighth IEEE International Conference. Retrieved from https://pdfs.semanticscholar.org/6c2e/32948af0134f0c3d85e8c262345568912208.pdf

Federal Aviation Administration (Oct. 5, 2010): *Fact Sheet – Airport Surface Detection Equipment, Model X (ASDE-X,* Retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=6296

Federal Aviation Administration (Oct. 5, 2010): *Fact Sheet – Engineered Material Arresting System (EMAS),* Retrieved from https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=13754

Friedman, J. (2001): Greedy Function Approximation: A Gradient Boosting Machine. Institute of Mathematical Statistics, Retrieved from http://www.jstor.org/stable/2699986

Edwards, L. (2009): Accidents vs. Incidents, Retrieved from http://www.aviationpros.com/article/10374517/accidents-vs-incidents