A Bayesian-Network-based Approach to Risk Analysis in Runway Excursions

Fernando Calle-Alonso[1], Carlos J. Pérez[2] and Eduardo S. Ayra[3]

[1](*Universidad de Málaga, Spain*)
[2](*Universidad de Extremadura, Spain*)
[3](*Universidad Politécnica de Madrid, Spain*)
(E-mail: fernandocalle@uma.es)

Aircraft accidents are extremely rare in the aviation sector. However, their consequences can be very dramatic. One of the most important problems are runway excursions, when the aircraft exceeds the end (overrun) or the side (veer-off) of the runway. After performing exploratory analysis and hypothesis tests, a Bayesian-network-based approach was considered to provide information from risk scenarios involving landing procedures. The method was applied to a real database containing key variables related to landing operations on three runways. The objective was to analyse the effects over runway overrun excursions of failing to fulfil expert recommendations upon landing. For this purpose, the most influential variables were analysed statistically, and several scenarios were built, leading to a runway ranking based on the risk assessed.

KEYWORDS

1. Aviation. 2. Safety. 3. Bayesian estimation. 4. Risk Minimisation. 5. Runway excursion.

Submitted: 13 March 2018. Accepted: XX February 2019.

1. INTRODUCTION. Aviation has grown from the first commercial flight in 1914 to the present with more than 120,000 landings and 12 million passengers per day globally in 2018, according to the Air Transport Action Group (ATAG, 2018). Forecasts on the increase of commercial flights are optimistic, with the number of passengers expected to double by 2030. Despite the large number of flight operations, the amount of accidents and incidents remains low. However, their consequences can be very dramatic in both losses of lives and economic costs. According to Boeing (2017), there were a total of 1948 accidents in the period 1959-2016 worldwide (388 from 2007 to 2016). 31.98% of those accidents had fatal consequences (15.98% in 2007-2016).

In the last 50 years, there has been a major reduction in the number of accidents. This is due to improvements in technology and risk management. The research and development departments of aviation companies and the international aviation agencies have made an effort to evaluate various risks. Since these cannot be fully eliminated, the main objective is to provide procedures for their efficient and optimal management. It was not until 2006 that the International Civil Aviation Organization (ICAO) published the first risk management guidelines (ICAO, 2006). The proposals were widely accepted by air transport authorities and aviation manufacturers. Two updates were released in 2009 and 2013. However, the number of accidents (especially runway overruns) worsened in the period 1992-2011 with the increase in flight operations (Rosenkrans, 2012).

There are many types of accidents and incidents in aviation see, for example, Evans (2014). Some of the most serious happen during landing. This phase is critical as pilots must consider many factors and make the right decisions in a short time (Gerard, 2006; You et al., 2013;

Wang et al., 2014; Boyd and Stolzer, 2016). According to Boeing (2016), in the last ten years 49% of accidents in commercial flights occurred during the final approach or landing. There are many factors affecting landing manoeuvres. Jenkins and Aaron (2012) point to the following causes as the most relevant in landing accidents: unstabilised approach, tail or crosswind, high speed, long landing or poor use of reverse thrust. Other authors such as Chang et al. (2016), Hunter et al. (2011) and Ahmed et al. (2014) focus on environment, containment runways and weather as risk factors. Rosenkrans (2012) defined some recommendations that should be followed to reduce risks, such as crossing the threshold at the right height at moderate speed, beginning the landing at the specified zone and properly using reverse thrust and other braking systems. The Flight Safety Foundation (FSF) also provides recommendations to reduce the risk of runway excursions (FSF, 2009).

Runway excursions happen during landings when an aircraft lands short (undershoot) exceeds the end (overrun) or the side (veer-off) of the runway (see Wong et al., 2006). In the period 2010-2014, 87 out of 415 accidents with damage were runway excursions, resulting in a total of 174 deaths according to the International Air Transport Association (IATA) (see IATA, 2015). Overruns are especially relevant since they account for 22% of all accidents and incidents in air transportation, and 44% of runway excursions are overruns.

There are many risk factors potentially affecting runway overrun. To discover them, it is necessary to analyse the conditions before landing and the information obtained after an accident occurs. Van Es (2005) analysed data from 400 runway overruns covering 35 years. Later, Van Es et al. (2009) dealt with identifying the main factors in runway overruns and proposed a risk index. Ayres et al. (2013) described a frequency model to determine the location of runway overruns and analysed the influence of several risk factors.

To prevent these risks, it was recommended by ICAO (2013) to build Runway Safety Areas (RSAs), which help to reduce damage in overruns and also in overshoots. In this context, Arnaldo Valdés et al. (2011) proposed a probability-based approach to estimate risks in both overrun and undershoot, offering solutions for the construction of RSAs. Drees et al. (2014) analysed the influence of some aerodynamic variables and proposed a sensitivity analysis method based on simulated data. Benedetto et al. (2014) analysed runway deflection and rutting, trying to reduce the overrun risks with cleared and graded areas. Also, Wilke et al. (2014) provided a risk analysis based on the runway surface. An overrun prevention system based on the alert box defined by Ryan and Brodegard (2003) has been patented.

Frequency models have been widely used for risk assessment in aviation accident analysis. However, they provide a descriptive perspective and do not allow one to relate multiple variables from a probabilistic point of view. Bayesian Belief Networks or simply Bayesian Networks (BNs) (Cowell et al., 2006) constitute an especially interesting probabilistic method for constructing risk scenarios related to aviation accidents.

Bayesian modelling is not very common in aviation literature, but it is very promising. For instance, Feng et al. (2009) applied it for aviation baggage screening, Ronald and Fabian (2014) for aviation delay modelling and Andres et al. (2005) for aviation human-system risk and safety. In particular, BN's have been applied in several settings for the purpose of supporting probabilistic predictions in aviation risk analysis. For example, Luxhøj and Kauffeld (2003) presented a new risk model for aviation systems that evaluates the impact of including new technologies to reduce the negative consequences of accidents. This BN model was implemented in the Probabilistic Decision Support System (PDSS) software package to evaluate risks of new technology interventions. Brooker (2011) considered expert opinions to estimate rare events in a BN framework and Gu (2009) used BNs to analyse the relationships among the main risk factors (helicopter technical dependability measures and environmental features) for helicopter accidents.

The focus of this paper is on risk analysis for prevention of overrun excursions based on a real dataset. Obtaining real data from aviation companies is very difficult (see, for example, Kirkland et al. (2004)) although its availability would clearly be very useful in that it would provide accurate information. We perform a BN model selection to fit the available data. The optimal BN is then used to define several risk scenarios that allow us to assess the influence of some states of the variables involved. This also lets us rank the corresponding runways based on the risk of overrun excursion. This is based on the probability of the event that the remaining distance to the end of the runway is less than 2,500 feet when the velocity is greater or equal to 80 knots. The higher this probability, the higher the risk of runway overrun.

2. METHODS

2.1. *Data available.* A data collection process was performed based on landing operations on three runways. They have similar operational conditions with landing lengths of less than 7,200 feet. Their locations cannot be divulged for confidentiality reasons. The runways will be denoted *RwA*, *RwB*, and *RwC*.

A total of 266 landing operations over a period of 10 months were recorded for the same airline and aircraft type. The main variable of interest was the remaining length (in feet) of the runway when the aircraft has a speed of 80 knots, see, for example, Burin (2011). This variable, denoted *Rw80*, allows us to identify the risk of not finishing within the runway boundaries. According to the FSF guidelines (Burin, 2011), the risk of an accident increases considerably when the aircraft has a speed greater than 80 knots when it has a remaining landing distance of 2,000 feet. The other variables recorded are:

- *Height*: This variable represents the height at threshold in feet over the runway.
- *DiffIV:* This represents the difference between the indicated airspeed and the final approach speed at threshold, measured in knots. The indicated airspeed is read directly from the instruments, whereas the final approach speed represents the airspeed to be maintained down to 50 feet over the runway threshold (plus corrections for wind).
- *Tailwind*: This is the tailwind at threshold, measured in knots.
- *Xwind*: This represents the crosswind at threshold, measured in knots. Xwind and tailwind are of interest specifically for landing overruns. Other problems such as veer off or undershoot would also be affected by head wind but they are not included in this research.
- *Trevf:* This variable describes the time in seconds that the maximum reverse thrust is operated to provide deceleration.
- *Rw*. This identifies the landing runway: *A*, *B*, or *C*.
- *Approach:* This indicates whether the approach is stabilised or not.
- *Abrake*: This represents the autobrake activation, with three possible values: not activated, low, or medium.

It is also important to define how stabilized approach was defined. According to the FSF, IATA and Eurocontrol, an approach is stabilized when all the following criteria are met (see https://www.skybrary.aero/index.php/Stabilised_Approach):

- The aircraft is on the correct flight path.
- Only small changes in heading/pitch are necessary to maintain the correct flight path.
- The airspeed is not more than Velocity of Reference (VREF) + 20kts indicated speed and not less than VREF.
- The aircraft is in the correct landing configuration.
- Sink rate is no greater than 1,000 feet/minute; if an approach requires a sink rate greater than 1000 feet/minute a special briefing should be conducted.
- Power setting is appropriate for the aircraft configuration and is not below the minimum

power for the approach as defined by the operating manual.
- All briefings and checklists have been conducted.

In Section 4, this dataset will be analysed. We next describe the methods used for data analysis.

2.2. *Statistical methods.* We performed a statistical data analysis of the database. Firstly, exploratory techniques were used. In order to analyse the relationships between pairs of variables, different hypothesis tests were considered. A chi-squared test was applied to determine the possible association between the use of autobraking and the runway. A Fisher exact test was applied for the same purpose between approach stabilisation and the runway (as, in this case, the applicability conditions for a chi-squared test were not met). For two quantitative variables, Spearman correlation coefficients and testing were considered to analyse the strength of the association between variables (as the conditions for a Pearson test were not met). Finally, since Analysis of Variance (ANOVA) applicability conditions could not be assumed, a nonparametric alternative was applied. Specifically, the Kruskal-Wallis test (Kruskal and Wallis, 1952) was used to analyse the differences between each quantitative variable for the different runways. Pairwise comparisons were performed using a Mann-Whitney *U* test (Mann and Whitney, 1947) with a Bonferroni correction (Dunn, 1961).

A BN is a directed acyclic graph that defines a joint probability distribution for a set of features. Each node represents a variable and the arcs represent causal dependence between nodes. For each node, the distribution of its variable conditioned on its predecessors in the net is defined, it is called Conditional Probability Distribution (CPD). The complete network defines a joint probability distribution, based on the CPD and the graph relationships. This joint distribution is given by the product of the probabilities of each node conditioned on the value of its predecessors. BNs are specifically defined for modelling uncertain and complex risk domains. As they are defined strictly in terms of probabilities and conditional independence statements, their main use would be to analyse different conditional probabilities scenarios Akhtar and Utne (2014). They can also be applied to prediction problems (see Banghart et al., 2017). One of the main advantages of its use is that it provides both a causal and probabilistic model, thus it is perfect for providing information supported by experts' knowledge combined with conditional probabilities, resulting in a human-interpretable decision tool (Heckerman, 2008).

The BNs take all the variables in the database into consideration conjointly to construct risk scenarios. For this task, quantitative variables were discretised by separating them into non-overlapping intervals. For comparative purposes, six algorithms for BN were considered in a model selection framework: Bayesian Search, Essential Graph Search, Tree Augmented Naive Bayes, Augmented Naive Bayes, Simple Naive Bayes, and Greedy Thick Thinning, all of which are implemented in Genie (ByoungHee, 2014), which was also used for BN calculations. The performance of the algorithms when applied to the current database were compared through five-fold cross-validation using 1,000 repetitions (Kim, 2009). The accuracy and the Area Under the Receiver Operating Characteristic (ROC) Curve (denoted AUC) were considered as goodness-of-fit measures. The ROC curve presents sensitivity in the y coordinate versus one-specificity or false positive rate in the x coordinate. The AUC is a measure of model performance used for classification models that varies in the range [0,1]. When it is close to 1, it means that the algorithm is able to adequately identify the risky cases in which the remaining length of the runway is reduced and becoming a risk for the landing. As we are studying the remaining runway over and above 2,500 feet, this measure is especially interesting. It is one of the most used performance metric combined with accuracy rate.

The algorithm best fitting these data was Greedy Thick Thinning (Dash and Cooper, 2004). This is a graph-structure learning algorithm that searches for a Directed Acyclic Graph (DAG) which maximises the scoring function in the search space of all DAGs containing the finite set of variables. In BN structure learning, a scoring function measures how good a given network matches the data set. In this case, the score was calculated with the Bayesian Dirichlet equivalent uniform (BDeu) criterion. This is based on a Dirichlet distribution with a weakly informative uniform prior, see Cooper and Herskovits (1992) and Silander et al. (2008). The Greedy Thick Thinning algorithm starts with an empty graph at a specific point in the structure space. In order to maximise the Bayesian score, it continues adding neighbouring arcs until no additional arc improves the score. It then starts to remove arcs until a local optimum is achieved. The first process is known as thickening; the second one as thinning. The result is the model that best fits the given data.

The BN produced by this algorithm will be used in the following section to construct risk scenarios through evidence propagation, one of the most powerful features of Bayesian networks. With this technique, probabilities at each node can be updated via two-way propagation of new information throughout the structure. The resulting probabilities will be expressed as percentages. In each scenario, a set of evidences (represented by 100% probability in some categories) is defined according to very risky or very safe situations. We study how these evidences influence the probabilities of runway excursions. Probability estimates for different scenarios are reported in two tables.

3. RESULTS AND DISCUSSION. We shall present first the main exploratory results, and then we shall apply the Bayesian-network-based approach to perform a risk analysis.

3.1. *Exploratory analysis*. Table 1 lists the frequency distributions of the variables describing runway, type of approach, and autobrake activation. Runway *A* is the most frequently used, receiving almost 40% of landing operations. Autobraking was not activated in more than one third of the operations, while unstabilised approaches were rare (2.3%).

Table 1. Frequency distributions for qualitative variables.

Runway	Frequency	%
A	104	39.1
B	80	30.1
C	82	30.8
Type of approach	Frequency	%
Stabilised	260	97.7
Unstabilised	6	2.3
Autobrake activation	Frequency	%
No	95	35.7
Low	56	21.1
Medium	115	43.2

Descriptive statistics of the relevant quantitative variables are presented in Table 2. The global recommendations FSF (2009) for reducing the risk of runway excursions are largely satisfied. For instance, both crosswind and tailwind tend to be under 10 knots; the speed difference tends to be less than 10 knots; the height, less than 50 feet, and there remains a length of runway greater than 2,000 feet when a speed of 80 knots has been reached. However, some landings did not satisfy the recommendations. Specifically, 6% and 0.8% of the landing operations were performed with crosswind and tailwind speeds greater than 10 knots, respectively; 2.3% were unstabilised; 1.9% had speed differences greater than 10 knots; and, finally, 19.9% had heights at threshold above 50 feet.

Figure 1 shows box plots of the quantitative variables. The frequency distributions for maximum reverse thrust, tailwind, crosswind, and height have a right-tailed asymmetry. All variables except maximum reverse thrust present outliers. For Rwy80, the most worrisome outliers are located close to 2,000 feet. The height shows quite a few risky operations above 50 feet, some of them close to 99 feet. Finally, there are more crosswind than tailwind outliers above the maximum recommendation of 10 knots.

Table 2. Descriptive statistics of quantitative variables involved.

	Height	DiffIV	Tailwind	Xwind	Trevf	Rwy80
Mean	31.23	2.53	2.32	5.35	5.36	3602.10
Median	28.00	2.00	1.00	4.00	4.50	3613.46
Std. Dev.	24.53	3.89	2.87	2.98	5.57	540.19
Minimum	1.00	-7.00	0.00	1.00	0.00	1890.77
Maximum	99.00	13.00	12.00	17.00	20.00	5518.40
FSF Recom	<50	<10	<10	<10	>0	>2000

Table 3 presents Spearman correlation coefficients for pairwise variables and their p-values (*p*). Six, out of fifteen, pairwise variable associations, were not significant; the greatest significant correlation coefficient was 0.383 ($p < 0.001$), obtained for speed difference and height. The variable Rwy80 has significant negative correlations with all the other variables, showing that the remaining length of the runway at 80 knots decreases as height, crosswind, tailwind, speed difference or maximum reverse thrust increases. An increase in any of these variables would suggest an increase in the risk of runway excursions. Nevertheless, none of these associations seems strong.

Table 3. Pairwise Spearman correlation coefficients and two-sided p-values.

		Height	DiffIV	Tailwind	Xwind	Trevf	Rwy80
Height	Spearman p	1.000	0.383	-0.171	0.139	0.024	-0.200
		.	0.000	0.005	0.024	0.692	0.001
DiffIV	Spearman p	0.383	1.000	0.034	0.085	0.056	-0.207
		0.000	.	0.580	0.168	0.365	0.001
Tailwind	Spearman p	-0.171	0.034	1.000	0.052	0.138	-0.283
		0.005	0.580	.	0.400	0.025	0.000
Xwind	Spearman p	0.139	0.085	0.052	1.000	0.059	-0.198
		0.024	0.168	0.400	.	0.341	0.001
Trevf	Spearman p	0.024	0.056	0.138	0.059	1.000	-0.162
		0.692	0.365	0.025	0.341	.	0.008
Rwy80	Spearman p	-0.200	-0.207	-0.283	-0.198	-0.162	1.000
		0.001	0.001	0.000	0.001	0.008	.

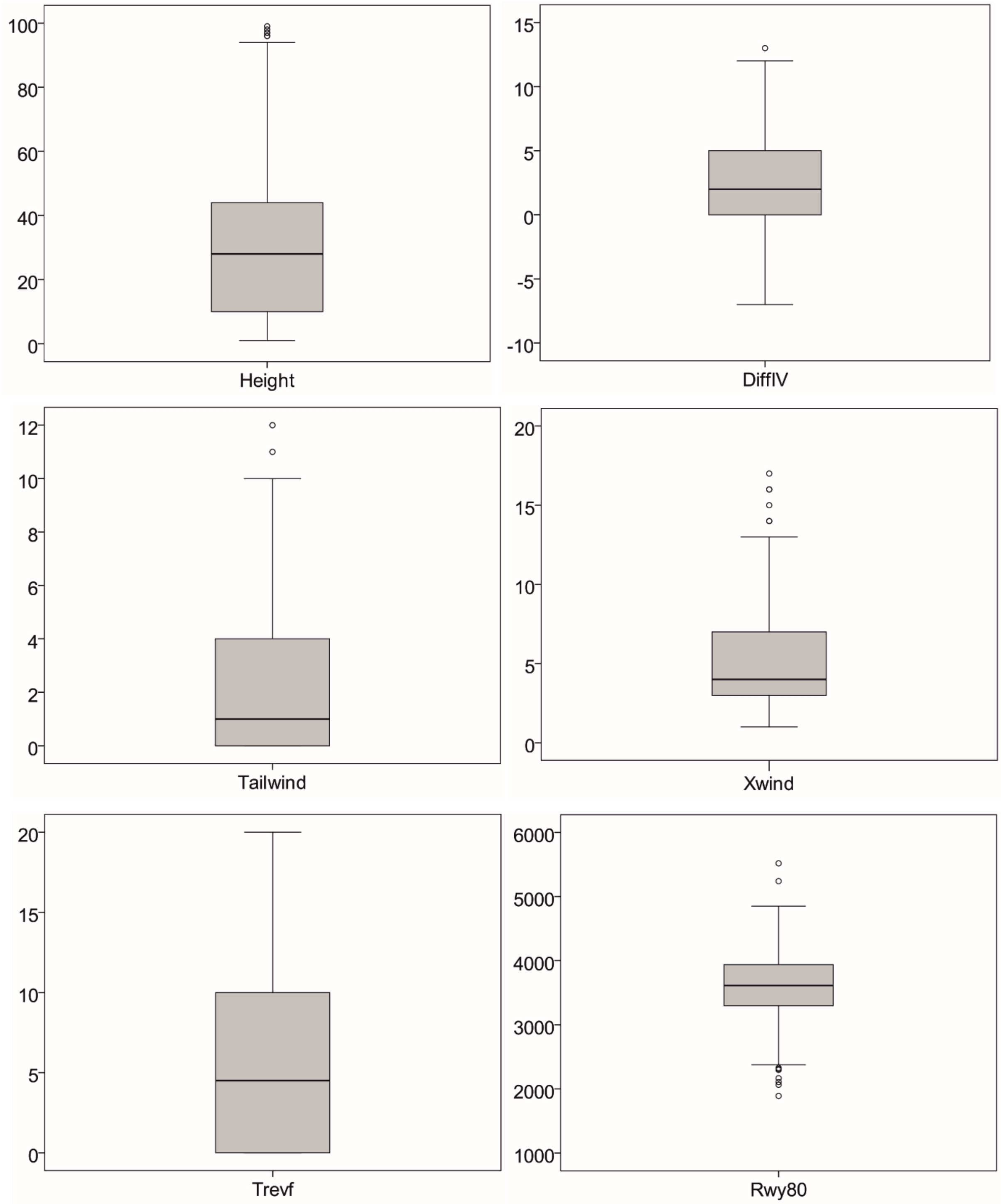

Figure 1. Box plots for: height, DiffIV, Tailwind, Xwind, Trevf, and Rwy80.

Aviation safety experts stress that landing operations are influenced by the characteristics and state of the runway (Daidzic and Shrestha, 2008). Table 4 provides a comparison between stabilised and unstabilised landings depending on the runway used. The highest number of unstabilised landings among the three runways observed occurred in Runway *C*. However, Fisher's exact test showed that there is no significant association between the type of approach and runway (p = 0.674).

Table 4. Type of approach by runway.

	Type of approach					
	Stabilised			Unstabilised		
Runway	Freq.	% Row	% Col.	Freq.	% Row	% Col.
A	102	98.1%	39.2%	2	1.9%	33.3%
B	79	98.8%	30.4%	1	1.3%	16.7%
C	79	96.3%	30.4%	3	3.7%	50.0%

On the other hand, Table 5 presents the contingency table for autobrake activation and runway. Autobraking was used with less intensity on Runway *C* and with greater intensity on Runway *A*. In this case, the chi-squared test showed a significant statistical association between use of autobraking and runway ($p < 0.001$).

Table 5. Autobrake activation by runway.

	Automatic brake								
	No			Low			Medium		
Runway	Freq.	% Row	% Col.	Freq.	% Row	% Col.	Freq.	% Row	% Col.
A	22	21.2%	23.2%	12	11.5%	21.4%	70	67.3%	60.9%
B	31	38.8%	32.6%	22	27.5%	39.3%	27	33.8%	23.5%
C	42	51.2%	44.2%	22	26.8%	39.3%	18	22.0%	15.7%

Table 6 provides descriptive statistics of the quantitative variables by runway. Some differences can be observed. Landing operations performed on Runway *A* have the greatest height, crosswind, and maximum reverse thrust averages. This runway also has the shortest remaining length average when the speed has fallen to 80 knots (although the minimum observation happens in Runway *C*), further evidence that this runway has the greatest excursion risk. Runway *B* seems to be following all the recommendations with more allowance, while Runway *C* only provides more allowance for crosswind and tailwind (both mean values are below those obtained for Runways *A* and *B*). Figure 2 provides box plots of all the variables by runway.

Table 6. Descriptive statistics of the quantitative variables by runway.

	Runway					
	A		B		C	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Height	38.0	26.2	21.4	14.8	32.2	27.0
DiffIV	2.4	3.5	1.7	4.1	3.6	4.0
Tailwind	2.4	3.0	2.8	2.9	1.8	2.7
Xwind	6.5	3.5	5.2	2.8	4.0	1.7
Trevf	5.9	5.4	4.9	5.8	5.1	5.6
Rwy80	3365.3	486.8	3831.3	522.7	3678.9	506.9

Significance tests were applied. Kruskal-Wallis tests were applied to analyse whether there were statistically significant differences in the variables of interest for the three runways. Such differences were observed for all quantitative variables ($p < 0.05$) except for maximum reverse thrust ($p = 0.376$). Pairwise comparisons were carried out for each pair of runways using penalised Mann-Whitney U tests. The height difference was statistically significant between Runways *A* and *B* ($p < 0.001$). There were also differences in speeds at Runways *B* and *C* ($p = 0.024$) (with Runway *C* being the one presenting the greatest speed variations) and in tailwind ($p < 0.05$) for *B* and *C* again (but in this case with Runway *B* being that with the greatest value). With respect to crosswind, there were differences for all pairs of runways ($p < 0.05$). The same was the case for the remaining runway lengths at 80 knots ($p < 0.001$). Runway *A* had the lowest average remaining length and the greatest average crosswind.

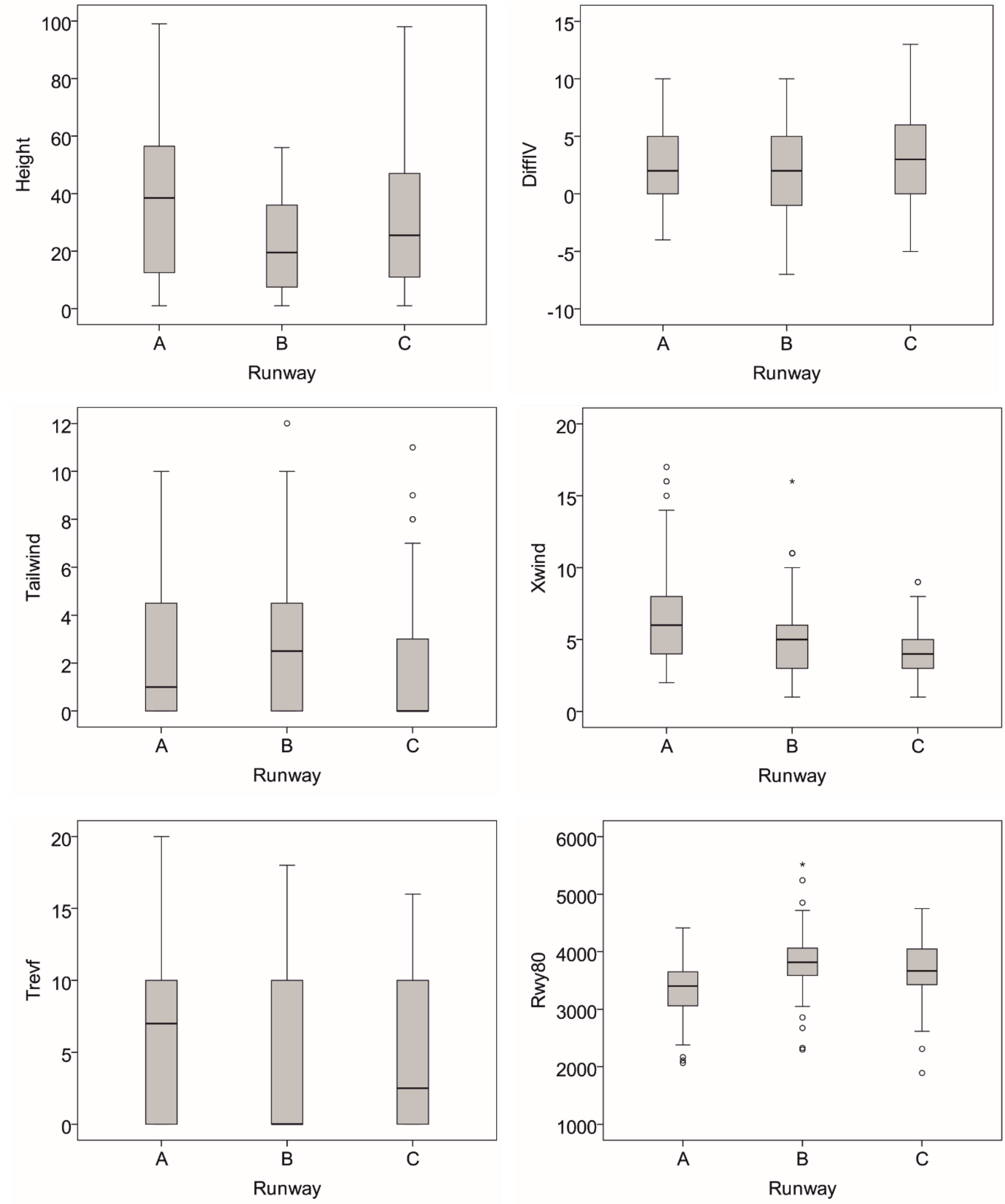

Figure 2. Box plots for the quantitative variables by runway.

3.2. *Bayesian networks for risk assessment.* With this background in mind, we have performed a BN analysis. For comparison of the different BN models, the specifications given in Section 2.2 have been followed (six models; five-fold cross-validation with 1,000 repetitions; accuracy and AUC comparison). We define a binary outcome with only two events to be observed: $Rwy80 < 2500$ and $Rwy80 \geq 2500$. As in this database there are very few cases under the recommendation of the Flight Safety Foundation, we will try to identify landings not only under 2,000 feet, but also close to them (< 2,500). This intends to make the requirement stricter. Table 7 lists the accuracies of the six algorithms after cross-validation. The Greedy Thick Thinning algorithm has the greatest accuracy (0.78), followed by the Tree Augmented Naive Bayes (0.73).

Table 7. Accuracies for the six algorithms evaluated.

Algorithms	Accuracy
Bayesian Search	0.49
Essential Graph Search	0.41
Tree Augmented Naive Bayes	0.73
Augmented Naive Bayes	0.61
Naive Bayes	0.56
Greedy Thick Thinning	0.78

ROC curves present a false positive rate (one-specificity) versus sensitivity. The area under the ROC curve (AUC) is a measure of performance used for classification models. It is the most used performance metric together with accuracy rate.

Figure 3 presents the ROC curves and AUCs for the six algorithms following the cross-validation procedure. Again, the best performance was for the Greedy Thick Thinning algorithm ($AUC = 0.998$), followed by the Tree Augmented Naive Bayes ($AUC = 0.968$).
Given its best results, the Greedy Thick Thinning algorithm was run to construct the final BN used. The space of states for *Rwy80* was then divided into six intervals: < 2,500, [2,500, 3,000), [3,000, 3,500), [3,500, 4,000), [4,000, 4,500), ≥ 4,500.

The estimated marginal probabilities in the BN (in the form of percentages) associated with each state are presented for each node in Figure 4. Relationships are represented by arcs, for which aviation experts checked that the connections represented were meaningful.

We performed a sensitivity analysis to observe if small changes in the predictors have an important effect on the prediction of the feature measuring the runway remaining at 80 knots. In Figure 4, the nodes are shown in a graduated red colour so that the more intense the colour, the greater influence this variable has on the conditional probability of the target node (*Rwy80*): runway, height at threshold, speed difference, autobrake activation and maximum reverse thrust seem the most influential variables. Figure 5 provides a tornado chart with which to analyse the influence of each state on the target variable, showing the ten most relevant states or state combinations. In particular, the figure shows the changes in $P[Rwy80 < 2{,}500]$ when varying the probability of the states shown at each bar by ±20%. Green denotes an increase of 20%, whereas red denotes a decrease of 20%. For instance, considering the top bar, if the probability of landing on Runway *A* increases by 20%, then $P[Rwy80 < 2{,}500]$ rises to 0.125. On the other hand, if the probability of landing on Runway *A* decreases by 20%, then $P[Rwy80 < 2{,}500]$ declines to 0.121. Although, at first sight, this might seem a rather small probability difference, it implies four more accidents per 1,000 operations. Since aviation accident rates tend to be very low, this could be a major point of analysis to bear in mind so as to try to reduce even small risks.

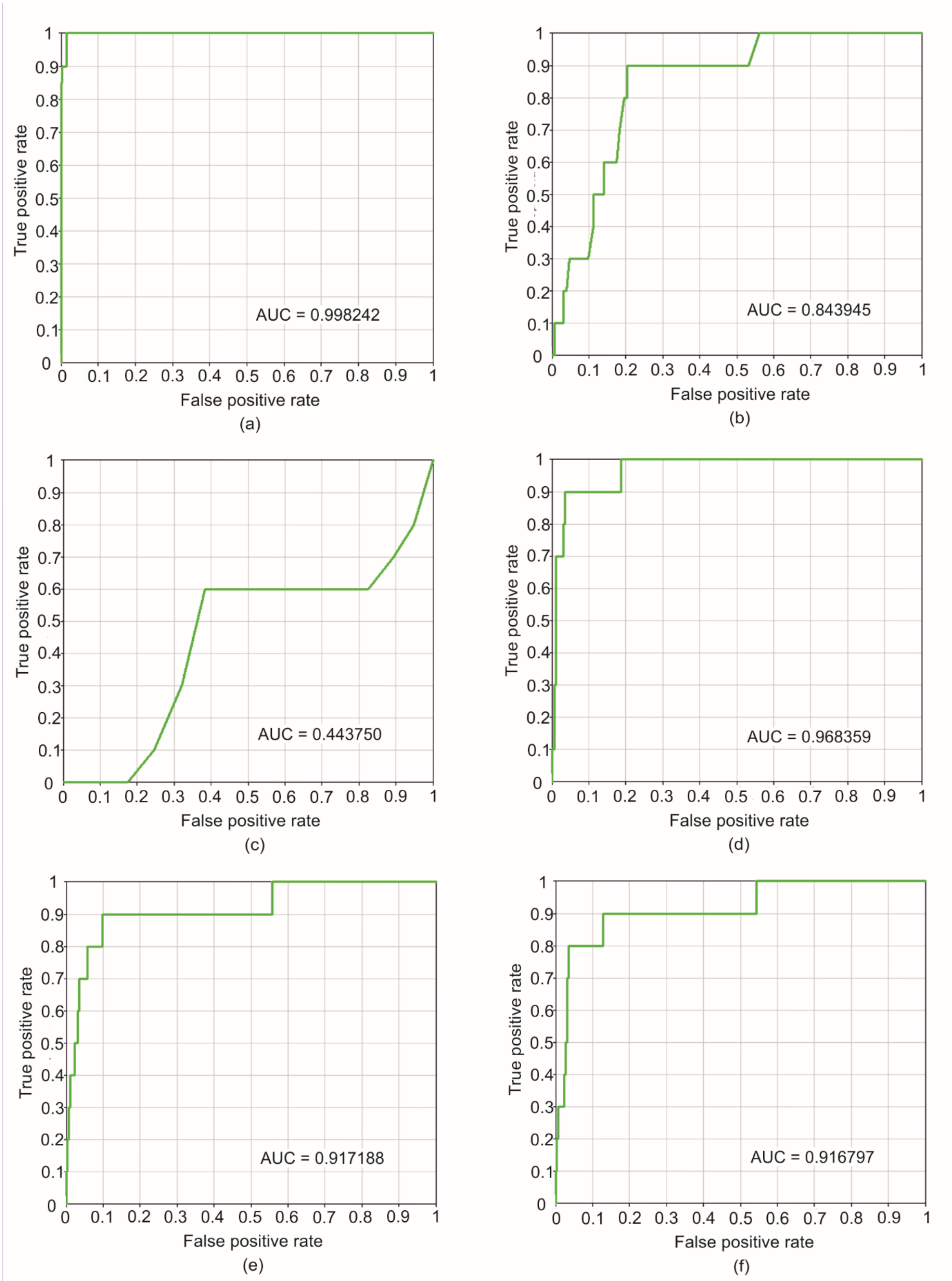

Figure 3. ROC curves and AUCs using different algorithms for Rwy<2,500: (a) Greedy Thick Thinning; (b) Bayesian Search; (c) Essential Graph Search; (d) Tree Augmented Naive Bayes; (e) Augmented Naive Bayes; (f) Naive Bayes.

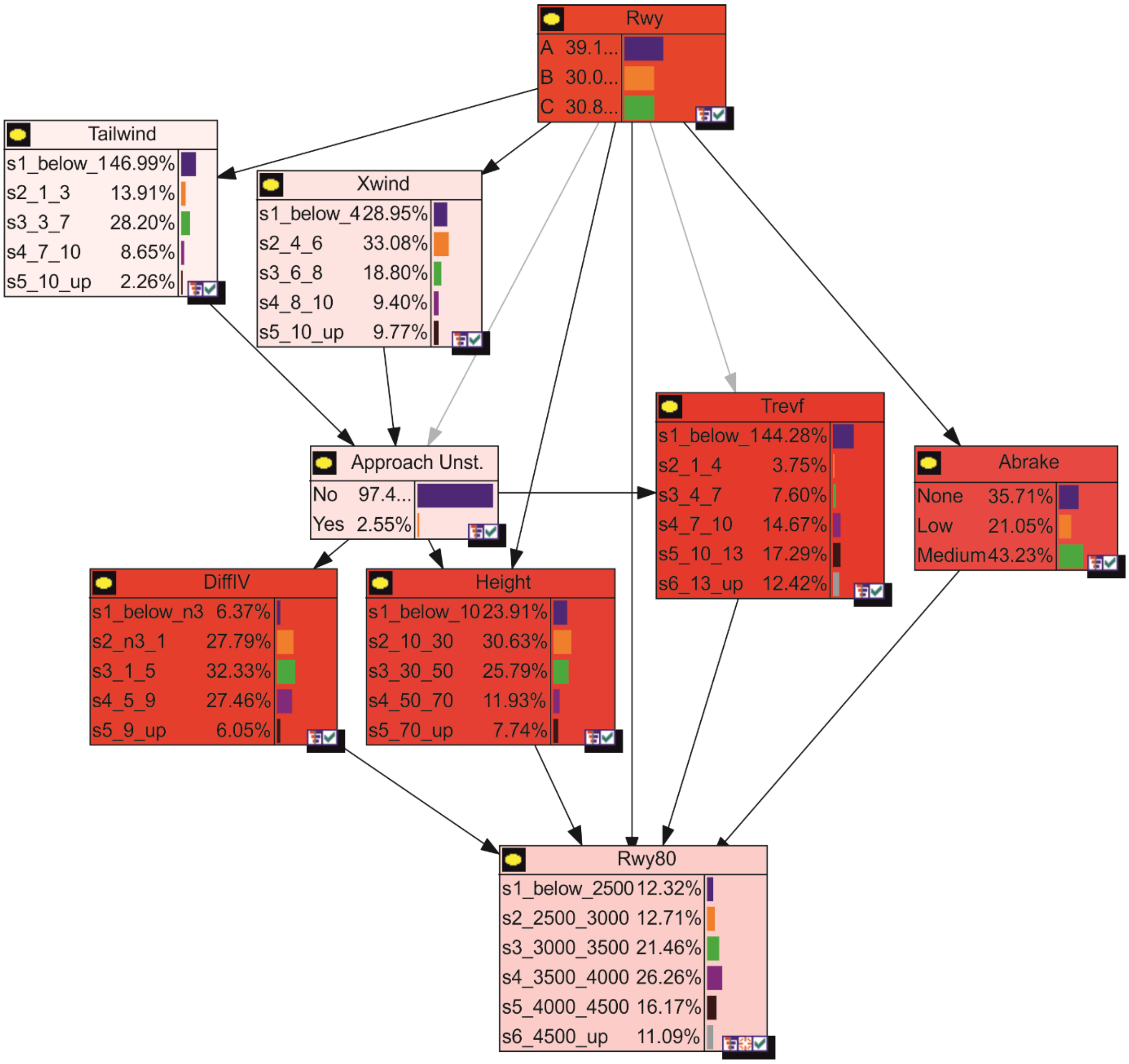

Figure 4. Graphical representation of the BN considered.

Figure 5. Tornado chart for the less than 2,500 feet of runway remaining at 80 knots event.

We then constructed ten risk scenarios using evidence propagation. This allowed us to evaluate the risk of overrun based on hypothetical scenarios with the states considered of the variables of interest. For each scenario, a probability of 1 (100% in percentage terms) is set for one state in one or more nodes.

Firstly, three scenarios were defined to analyse the characteristics of each runway as represented by the data, and two further scenarios were formulated to describe situations with maximum and minimum runway overrun risks. Table 8 presents the results of the inference process based on the BN considered.

Comparing the first three scenarios related to runways, we observe that Runway *A* has the greatest percentages of landings with more than 10 knots of tailwind (2.88%) and crosswind (17.31%) and Runway *C* the lowest. Runway *A* also stands out in unstabilised landings (3.95%), followed by 2.19% for Runway *B* and 1.14% for Runway *C*. The activation time for maximum reverse thrust is similar for the three runways. The use of autobrake systems is very different, however, with Runway *A* requiring the greatest activation of autobraking at medium level (67.31%) and Runway *B* the least (21.95%). The recommendation is that these two braking systems should be used moderately during landing roll. Runway B has the greatest percentage of moderate use of both devices. According to all this information, Runway A could be catalogued as riskier, and landings should be taken more carefully, for example, when situations with high wind are observed.

Speed difference is similar for the three runways. Nevertheless, for differences greater than 9 knots, Runway *A* presents a slightly greater percentage than *B* and *C*. Height at threshold differs considerably among the three runways. Again, Runway *A* has the greatest percentage (32.96%) of landing manoeuvres at threshold above the recommended height of 50 feet. Runway *C* has a percentage of 20.78%, while Runway *B* is the safest in this regard with 1.24%. In this case runway *A* should have stricter recommendations for landing for the speed and the height at the threshold.

With respect to the remaining length of runway at 80 knots, Runway *A* has the greatest percentage (14%) of lengths being less than 2,500 feet, followed by Runway *C* with 11.51%, and finally by Runway *B* with 10.97%. Considering the accumulated percentages for landings with more than 3,500 feet remaining, we observe that Runway *B* would be the safest with 63.96%, followed by Runway *C* with 55.97%, and, finally, Runway *A* with 43.55%.

This analysis by runway is an example of the valuable insights that could be offered to any airport, giving the main risks to be observed during landing. Each different runway has its own characteristics and risks. A customised probability of suffering accidents could be given for any different runway.

We next analyse the scenarios for maximum and minimum risks, according to the FSF's recommendations. Scenario 4 considers conditions not recommended for landing by the FSF. This kind of landing has most probability of occurring at Runway *A* (92.36%), whereas at Runway *C* it would be 7.64%. According to the BN model, these conditions would be rarely met at Runway *B*. The opposite is the case for Scenario 5, where the best landing conditions are considered. In this case, there is a probability of 79.52% that the landing was performed on Runway *B*, followed by Runway *C* (12.98%) and with the lowest probability being for Runway *A* (7.5%).

All the information provided in the scenarios in Table 8 is coherent with the statement that Runway *B* is the safest, followed by Runway *C*, and, finally, with Runway *A* being the least safe. This also agrees with the results obtained in Section 3.1.

Table 8. Scenarios for excursion risk evaluation by runway.

	1	2	3	4	5
Runway *A*	100.00			92.36	7.50
B		100.00		0.00	79.52
C			100.00	7.64	12.98
Tailwind <1	47.12	36.25	57.32		100.00
1-3	16.35	13.75	10.98		
3-7	23.08	38.75	24.39		
7-10	10.58	8.75	6.10		
>10	2.88	2.50	1.22	100.00	
Crosswind					
<4	19.23	30.00	40.24	0.00	31.17
4-6	25.96	33.75	41.46	0.00	33.90
6-8	23.08	18.75	13.41	0.00	18.79
8-10	14.42	7.50	4.88	49.62	7.36
>10	17.31	10.00	0.01	50.38	8.79
Approach					
Stabilised	96.05	97.81	98.86		100.00
Unstabilised	3.95	2.19	1.14	100.00	
Trevf.					
<1	43.88	44.38	44.68	100.00	
1-4	3.69	3.76	3.80		
4-7	7.97	7.50	7.22		100.00
7-10	14.70	14.66	14.64		
10-13	17.28	17.29	17.30		
>13	12.48	12.40	12.36		
Abrake					
No	21.15	38.75	51.22		
Low	11.54	27.50	26.83		100.00
Medium	67.31	33.75	21.95	100.00	
DiffIV					
<-3	6.28	6.40	6.46		
-3-1	27.63	27.83	27.95		100.00
1-5	32.35	32.33	32.32		
5-9	27.55	27.44	27.38		
>9	6.20	6.01	5.89	100.00	
Height <10	19.87	33.14	20.02	92.36	
10-30	21.75	33.43	39.17	7.64	
30-50	25.43	32.18	20.02	0.00	100.00
50-70	21.66	1.24	10.01	0.00	
>70	11.30	0.01	10.77	0.00	
Rwy80					
<2500	14.00	10.97	11.51	100.00	
2500-3000	15.01	11.56	10.91		
3000-3500	27.46	13.51	21.60		
3500-4000	21.28	32.38	26.61		100.00
4000-4500	13.40	18.18	17.71		
>4500	8.87	13.40	11.65		

Five additional relevant scenarios not compliant with some of the most common FSF recommendations are presented in Table 9. Scenario 6 analyses the probability of an unstabilised approach. Runway *A* has the greatest percentage of unstabilised approaches (60.48%), whereas Runway *B* has almost double the percentage of Runway *C* (25.79% vs 13.79%). When the approach is unstabilised, 56.31% of the landings are performed with a crosswind greater than 10 knots (much greater than in the rest of the scenarios). Also, the speed difference is greater than 9 knots for 16.67% of the landings, a larger value than for the other scenarios. Autobrake systems are mainly used at a medium level (52.43%). Finally, 14.98% of the landings with an unstabilised approach had a remaining runway length at 80 knots of less than 2,500 feet, showing the risk of unstabilised approaches for a runway overrun.

In Scenario 7, tailwind speed was considered to be greater than 10 knots. This situation occurs above all on Runway A (50%), followed by Runway B (33.33%). Although crosswind risk probabilities are not greatly affected by tailwinds greater than 10 knots, they do seem to be related to the stabilisation of the approach (11.26% of the approaches are unstabilised in this case). This scenario also shows the need to use autobraking at a medium level (for which the percentage is relatively high, 48.56%). A remaining length at 80 knots of less than 2,500 feet seems to be not especially affected by high tailwind speeds in comparison with the other scenarios.

Scenario 8 considers a case in which the maximum reverse thrust is used briefly (between 1 and 4 seconds). This almost always occurs for stabilised approaches (99.9%). The probability of initiating landing with a height above 50 feet is 19.94%, the greatest value in these scenarios. Finally, the probability of having less than 2,500 feet of remaining length on the runways at 80 knots is 14.79%, one of the greatest probabilities of the scenarios analysed.

A large difference between recommended and actual speeds can lead to considerable risk during landing. Scenario 9 considers the consequences associated with a speed difference larger than 9 knots. Again, this risky situation most probably occurs at Runway *A* (40.08%), followed by Runways *B* and *C* with 29.88% and 30.04%, respectively. Speed differences greater than 9 knots occur with a probability of 7.03% in unstabilised approaches. The worst consequence is that the probability of landing with a remaining runway length of less than 2,500 feet at 80 knots is 15.05%.

Scenario 10 is related to heights greater than 70 feet at threshold. This kind of situation appears to be almost totally improbable for the landings on Runway *B*, whereas the probabilities for Runways *A* and *C* are 57.1% and 42.9%, respectively. In this case, a landing that is initiated at a greater than normal height seems to be not associated with other negative factors. Moreover, the probability of a remaining runway length at 80 knots that is less than 2,500 feet is the lowest (12.52%) of the risk scenarios in Table 9.

Table 9. Scenarios to evaluate runway excursion risk that are non-compliant with FSF recommendations.

	6	7	8	9	10
Runway *A*	60.48	50.00	38.54	40.08	57.10
B	25.79	33.33	30.19	29.88	0.00
C	13.73	16.67	31.27	30.04	42.90
Tailwind <1	40.28		47.17	46.68	52.25
1-3	19.85		13.75	14.18	13.58
3-7	0.00		28.93	26.90	23.56
7-10	29.92		8.09	9.62	8.61
>10	9.95	100.00	2.05	2.61	2.00
Crosswind					
<4	0.00	26.32	29.71	27.62	27.67
4-6	28.42	31.14	33.20	32.87	33.89
6-8	0.00	20.02	19.29	17.93	19.13
8-10	15.27	10.52	9.24	9.67	10.63
>10	56.31	11.99	8.56	11.91	8.69
Approach					
Stabilised		88.74	99.99	92.97	96.98
Unstabilised	100.00	11.26	0.01	7.03	3.02
Trevf					
<1	16.67	41.81		43.01	44.14
1-4	0.00	3.41	100.00	3.58	3.73
4-7	33.33	9.90		8.78	7.72
7-10	16.67	14.85		14.76	14.68
10-13	16.67	17.24		17.26	17.29
>13	16.67	12.80		12.61	12.44
Abrake					
No	29.82	32.03	35.87	35.44	34.05
Low	17.75	19.41	21.14	20.90	18.10
Medium	52.43	48.56	42.99	43.66	47.85
DiffIV					
<-3	0.00	5.80	6.54		6.34
-3-1	16.67	26.79	28.08		27.73
1-5	33.33	32.42	32.31		32.34
5-9	33.33	27.99	27.31		27.49
>9	16.67	7.00	5.77	100.00	6.10
Height <10	56.03	27.64	23.07	25.38	
10-30	34.82	29.54	30.52	30.83	
30-50	0.00	24.44	26.47	24.61	
50-70	0.00	11.52	12.24	11.38	
>70	9.15	6.87	7.70	7.80	100.00
Rwy80					
<2500	14.98	12.73	14.79	15.05	12.52
2500-3000	15.67	13.35	16.15	15.89	17.60
3000-3500	19.82	21.52	22.34	19.81	19.29
3500-4000	20.36	25.51	16.68	18.73	19.38
4000-4500	16.52	15.84	15.24	15.48	19.60
>4500	12.65	11.05	14.79	15.05	11.62

In summary, on one hand, an unstabilised approach and a large speed difference constitute two major factors for unsafe landings. When an approach is unstabilised, it is typically accompanied by some non-recommended conditions, for example, high crosswind and tailwind speeds, or a large speed difference. Both unstabilised approaches and high-speed differences occur with greater frequency at Runway *A*. On the other hand, for the probability of having a remaining runway length of less than 2,500 feet at 80 knots, the least dangerous factors are tailwind speed and height. Having under control all these aspects can lead to a safer context, reducing the runway excursion incident rates.

Finally, evidence propagation in the BN has allowed us to define risk scenarios that provide information about hypothetical situations that could arise in real landings. The method is extensible to other aviation datasets. Indeed, in general, the more data, the better the information that would be obtained.

4. CONCLUSIONS.

This paper explores the risk for runway overrun excursions under several scenarios. Data from three real runways were collected and analysed with a BN model, as an illustration of a risk evaluation that could be performed for all runways. The BN method using evidence propagation facilitated the analysis of the most common runway excursion risk scenarios.

This research revealed the relations among features measured when landing. Three runways were classified in accordance to their potential risks, and early risk disclosure was provided. The generic recommendations established by safety organisations can be revisited for each runway using the results of this kind of analysis. For example, Runway *A* is less safe than the others, and recommendations for landing should be stricter. This kind of analysis can be performed for more runways in different airports, providing a classification according to their runway overrun risk. This tool could be used in aviation operations for runway classification that allows pilots to have a runway excursion probability when landing.

The proposed approach can provide insights into why runway excursions occur and potentially lead to reductions in accident rates. The runway risk ranking could be useful for authorities to mitigate the amount of excursions, improving the weaknesses observed at each runway. With the BN graph, the scenarios given and the sensitivity analysis, the relations and influence among the main features is exposed. These results could be extended if more airports open their databases. This study remains open in case further data can be obtained in the future. New features such as touchdown point, landing procedure, weather situation or runway contamination would be especially important. The results obtained cover different kinds of risk, offering a comprehensive safety framework that would be applicable to all runways.

FINANCIAL SUPPORT

This research was supported by the Agencia Estatal de Investigación - Spain (Projects MTM2014-56949-C3-3-R and MTM2017-86875-C3-2-R); the Gobierno de Extremadura - Spain (Projects GR15106 and GR18108); and the European Union (European Regional Development Funds).

REFERENCES

Ahmed, M. M., Abdel-Aty, M., Lee, J. and Yu, R. (2014). Real-time assessment of fog-related crashes using airport weather data: A feasibility analysis. *Accident Analysis & Prevention*, **72**, 309 – 317.

Akhtar, M. J. and Utne, I. B. (2014). Human fatigue’s effect on the risk of maritime groundings

- A Bayesian Network modeling approach. *Safety science*, **62**, 427-440.
Andres, D. M., Luxhoj, J. T., and Coit, D. W. (2005). Modelling of human-system risk and safety: aviation case studies as exemplars. *Human Factors and Aerospace Safety*, **5**, 137-167.
Arnaldo Valdes, R. M., Gomez Comendador, F., Mijares Gordun, L. and Saez Nieto, F. (2011). The development of probabilistic models to estimate accident risk (due to runway overrun and landing undershoot) applicable to the design and construction of runway safety areas. *Safety Science*, **49**, 633–650.
ATAG Air Transport Action Group. (2018). Aviation Benefits Beyond Borders. Technical report. https://aviationbenefits.org/media/166344/abbb18_full-report_web.pdf Accessed 11 February 2019.
Ayres, M., Shirazi, H., Carvalho, R., Hall, J., Speir, R., Arambula, E., David, R., Gadzinski, J., Caves, R., Wong, D. and Pitfield, D. (2013). Modelling the location and consequences of aircraft accidents. *Safety Science*, **51**, 178–186.
Banghart, M., Bian, L., Strawderman, L. and Babski-Reeves, K. (2017). Risk assessment on the EA-6B aircraft utilizing Bayesian networks. *Quality Engineering*, **29**, 499-511.
Benedetto, A., D'amico, F. and Tosti, F. (2014). Improving safety of runway overrun through the correct numerical evaluation of rutting in cleared and graded areas. *Safety Science*, **62**, 326–338.
Boeing (2017). *Statistical summary of commercial jet airplane accidents: Worldwide operations 1959-2016*. Technical Report. https://www.skybrary.aero/bookshelf/books/4239.pdf Accessed 11 February 2019
Boyd, D. D. and Stolzer, A. (2016). Accident-precipitating factors for crashes in turbine-powered general aviation aircraft. *Accident Analysis & Prevention*, **86**, 209–216.
Brooker, P. (2011). Experts, Bayesian Belief Networks, rare events and aviation risk estimates. *Safety Science*, **49**, 1142–1155.
Burin, J. M. (2011). Keys to a safe arrival. *AeroSafety World*, **6**, 14–17.
Byoung-Hee, K. (2014). Bayesian networks. *Machine Learning*, **10**, 601B.
Chang, Y.-H., Yang, H.-H. and Hsiao, Y.-J. (2016). Human risk factors associated with pilots in runway excursions. *Accident Analysis & Prevention*, **94**, 227–237.
Cooper, G. F. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, **9**, 309–347.
Cowell, R. G., Dawid, P., Lauritzen, S. L. and Spiegelhalter, D. J. (2006). *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media.
Daidzic, N. E. and Shrestha, J. (2008). Airplane landing performance on contaminated runways in adverse conditions. *Journal of Aircraft*, **45**, 2131– 2144.
Dash, D. and Cooper, G. F. (2004). Model averaging for prediction with discrete Bayesian networks. *The Journal of Machine Learning Research*, **5**, 1177–1203.
Drees, L., Wang, C. and Holzapfel, F. (2014). Using subset simulation to quantify stakeholder contribution to runway overrun. Proceedings of Probabilistic Safety Assessment and Management PSAM 12, Honolulu, Hawaii.
Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, **56**, 52–64.
Evans, J. K. (2014). *Frequency of Specific Categories of Aviation Accidents and Incidents During 2001-2010*. Technical Report 218184. NASA. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140012710.pdf Accessed 11 February 2019.
Feng, Q., Sahin, H. and Karson, M. J. (2009). Bayesian analysis models for aviation baggage screening. *IIE Transactions*, **41**, 995-1006.

FSF. (2009). *Runway Safety Initiative. Reducing the risk of runway excursions*. Technical Report. Flight Safety Foundation. https://flightsafety.org/files/RERR/fsf-runway-excursions-report.pdf Accessed 11 February 2019.
Gerard, V. E. (2006). When a runway is not long enough to land on. *Aviation Safety Letters*, **1**, 27–28.
Gu, Y. (2009). *Risk Influence Modeling for Helicopter Safety*. Ph.D. thesis Norwegian University of Science.
Heckerman, D. (2008). *A tutorial on learning with Bayesian networks*. Springer.
Hunter, D. R., Martinussen, M., Wiggins, M. and O'Hare, D. (2011). Situational and personal characteristics associated with adverse weather encounters by pilots. *Accident Analysis & Prevention*, **43**, 176–186.
IATA. (2015). *Runway Safety Accident Analysis Report*. Technical Report. International Air Transport Association. https://www.iata.org/whatwedo/safety/runway-safety/Documents/RSAR-1st-2015-final-version.pdf Accessed 11 February 2019
ICAO. (2006). *Aircraft Operations, Volume I, Flight Procedures (Doc 8168 OPS/611)*. Manual. International Civil Aviation Organization. http://www.chcheli.com/sites/default/files/icao_doc_8168_vol_1.pdf Accessed 11 February 2019.
Jenkins, M. and Aaron, R. F. (2012). Reducing runway landing overruns. *Aero Magazine*, **3**, 14–19.
Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational statistics & data analysis*, **53**, 3735-3745.
Kirkland, I., Caves, R., Humphreys, I. and Pitfield, D. (2004). An improved methodology for assessing risk in aircraft operations at airports, applied to runway overruns. *Safety Science*, **42**, 891–905.
Kruskal, W. H. and Wallis, W. A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **47**, 583–621.
Luxhøj, J. T. and Kauffeld, K. (2003). Evaluating the effect of technology insertion into the national airspace system. *The Rutgers Scholar*, **5**.
Mann, H. B. and Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, **18**, 50–60.
Ronald, W. and Fabian, N. (2014). Bayesian model averaging: an application to the determinants of airport departure delay in Uganda. *American Journal of Theoretical and Applied Statistics*, **3**, 1-5.
Rosenkrans, W. (2012). *Overrun Breakdown*. Report. Flight Safety Foundation. https://flightsafety.org/asw-article/overrun-breakdown/ Accessed 11 February 2019.
Ryan, D. and Brodegard, W. (2003). Runway overrun monitor and method for monitoring runway overruns. Google Patents US App. 10/442,147.
Silander, T., Roos, T., Kontkanen, P. and Myllymaki, P. (2008). Factorized normalized maximum likelihood criterion for learning Bayesian network structures. Proceedings of the 4th European workshop on probabilistic graphical models (PGM-08), Hirtshals, Denmark.
Van Es, G. (2005). *Running out of runway: Analysis of 35 years of landing overrun accidents*. Technical Report. 498 National Aerospace Laboratory NLR. https://skybrary.aero/bookshelf/books/1146.pdf Accessed 11 February 2019
Van Es, G., Tritschler, K. and Tauss, M. (2009). *Development of a landing overrun risk index*. Technical Report. 280 National Aerospace Laboratory NLR. https://reports.nlr.nl/xmlui/bitstream/handle/10921/235/TP-2009-280.pdf Accessed 11 February 2019.
Wang, L., Wu, C. and Sun, R. (2014). An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents. *Reliability Engineering & System Safety*,

127, 86–96.
Wilke, S., Majumdar, A. and Ochieng, W. Y. (2014). Airport surface operations: A holistic framework for operations modeling and risk management. *Safety Science*, **63**, 18–33.
Wong, D. K., Pitfield, D. E., Caves, R. E. and Appleyard, A. J. (2006). Quantifying and characterising aviation accident risk factors. *Journal of Air Transport Management*, **12**, 352-357.
You, X., Ji, M. and Han, H. (2013). The effects of risk perception and flight experience on airline pilot's locus of control with regard to safety operation behaviors. *Accident Analysis & Prevention*, **57**, 131–139.