Transportation Research Part C: Emerging Technologies, Volume 128, July 2021, Article number 103180
DOI: 10.1016/j.trc.2021.103180

Estimating runway veer-off risk using a Bayesian network with flight data

David J. Barry

Abstract

Risk assessments in airline operations are mostly qualitative, despite abundant data from programmes such as flight data monitoring (FDM) and flight operations quality assurance (FOQA). In this paper, features relating to runway excursion causal factors are extracted from flight data from over 310,448 flights from Airbus A320 series aircraft flown on a European network. The data is combined with meteorological data to provide additional features.

Bayesian networks are then learnt from the feature set, and two network learning algorithms are compared, Bayesian Search and Greedy Thick Thinning (GTT). Cross-validation of the resulting networks shows both algorithms produce similarly performing networks, and a subjective analysis concludes that the GTT algorithm is marginally preferred.

The resulting networks produce relative probabilities, which airlines can use to quantitatively assess runway veer-off risk under different scenarios, such as different meteorological conditions and unstable approaches.

This paper's main finding is that by utilising existing data sources, such as FDM and weather databases, airlines can create and use Bayesian networks alongside their existing qualitative risk assessment methods to provide quantitative risk assessment and understand the effect of different conditions on those risks. This is not possible with current methods in use by airlines.

The method described can be extended to other operational safety risks, such as runway overrun.

2 Introduction

There is a scarcity of quantitative risk assessment (QRA) methods used in the airline operations domain. In other areas of aviation, such as airworthiness and design, quantitative methods, such as Fault Tree Analysis (FTA), are routinely used, however this is not the case in the operations domain. Some methods give the allusion of being quantitative, providing the analyst with a ‘quantitative’ score, but these are usually numerical values applied to qualitative measures (e.g. risk matrices, bow ie diagrams) (Hubbard, 2009). This can be dangerous, as it provides airline practitioners and management with false comfort, believing risk has been assessed with scientific rigour, when in fact, the result has relied upon subjective judgement.

The ICAO Safety Management Manual (International Civil Aviation Organisation, 2018) is the basis on which most safety management practices in airlines are founded. The guidance it contains is adopted by national regulators and used by inspectors to determine the quality and effectiveness of an airline’s safety management system (SMS).

LIKELIHOOD	MEANING	VALUE
FREQUENT	Likely to occur many times (has occurred frequently)	5
OCCASIONAL	Likely to occur sometimes (has occurred infrequently)	4
REMOTE	Unlikely to occur, but possible (has occurred rarely)	3
IMPROBABLE	Very unlikely to occur (not known to have occurred)	2
EXTREMELY IMPROBABLE	Almost inconceivable that the event will occur	1

Table 1. Safety risk probability table from ICAO SMM 4th Edition

ICAO defines safety risk as "*The predicted probability and severity of the consequences or outcomes of a hazard*", hence estimating the probability of consequences is fundamental to risk assessment. Table 1 shows the ICAO guidance for assessing probabilities. Note the qualitative definitions for "Likelihood". Safety analysts in airlines use these definitions to rank risks and thereby prioritise resources towards mitigating those risks. In a large airline, the analyst will have the advantage of having plenty of historical data to call upon while reviewing a safety incident and trying to determine the risk likelihood, but problems arise due to the highly subjective nature of the likelihood meanings. Where, for example, can the line be drawn between "has occurred infrequently" and "has occurred rarely"? Hubbard (2009) explains how bias and overconfidence can affect how risks are rated using terms like these, leading to poor risk assessment.

Another problem in the large airline scenario is that it is necessary to have several analysts performing the categorisation, and it can be challenging to maintain consistency between individuals. The problem is worse for smaller airlines that do not have vast amounts of historical data. Compare a sizeable low-cost airline with, say 400 aircraft, and a smaller airline operating five aircraft. It would take the smaller airline 80 years to gain operational experience and generate the safety data the larger airline achieved in one year. The result is that for a given risk, the analyst in the small airline might choose "not known to have occurred", whereas the analyst in the large airline might choose "has occurred frequently".

Wholly qualitative assessment would be perfectly acceptable if sources of quantitative data were not available, however airlines have rich sources of data such as safety reports, audit findings, investigation findings and, significantly, recorded flight data.

Modern aircraft can record thousands of parameters on quick access recorders (QAR) throughout every flight. The parameter set, known as the data frame, consists of measurements and condition status from a wide variety of sensors and systems, such as:

- Flight control positions
- Control surface positions
- Aircraft attitude
- Airspeed, altitude, vertical speed
- Geographic position
- Engine thrust, temperatures, pressures and vibrations
- Autopilot modes and selections
- Accelerations (normal, longitudinal and lateral)

The recorded data feeds into flight data monitoring (FDM) programmes which identify measures and events from each flight. This, in turn, supplies safety knowledge into the operator's safety management system (SMS).

At the core of safety management systems is risk assessment and management, and in 2011 the UK CAA, working with industry, published a list of the seven most significant risks facing air transport (Civil Aviation Authority, 2011):

- Loss of control
- Runway overrun or excursion (veer-off)
- Controlled flight into terrain
- Runway incursion and ground collision
- Airborne conflict
- Ground handling operations
- Airborne and post-crash fire

Many of the individual events and measurements from FDM can be mapped on to some of these risks as precursors. For example, a high angle of bank event may be mapped to the loss of control risk. Usually, the individual events have a basic risk weighting, commonly labelled low, medium or high, depending on the magnitude of exceedance. An airline can therefore study the prevalence of 'high' risk-weighted events over time or at different airports; however this method does little to help an airline quantify actual risk or be able to answer questions such as:

- What is this airline's risk of a runway veer-off?
- At which airport is a runway veer-off incident most likely?

Answers to questions like these are much more helpful in managing risk than simple precursor trends. However, providing answers can be complicated, even with the abundant data available today. This paper sets out a method using flight data and other common sources of data available to airlines to allow QRAs to be generated alongside the existing commonly used qualitative assessments. The method allows airlines to exploit the data they have available and properly quantify the common risks they encounter. Once established, the method can be used routinely, with little additional effort, to provide contemporary risk assessments.

3 Literature review

3.1 Regulatory framework

By the time FDM became an ICAO mandatory practice in 2005, many operators and airlines already had mature programmes in place. Accordingly, the regulations which were introduced reflected current practice, and so many airlines did not have to make changes to existing programmes. However, those operators without existing FDM programmes had to invest in FDM hardware and software or opt to outsource their programmes to service providers.

The ICAO regulations, enshrined in Annex 6 (International Civil Aviation Organisation, 2010), state that:

> *3.3.5* ***Recommendation****. An operator of an aeroplane of a certificated take-off mass in excess of 20 000 kg should establish and maintain a flight data analysis programme as part of its safety management system.*
>
> *3.3.6 An operator of an aeroplane of a maximum certificated take-off mass in excess of 27 000 kg shall establish and maintain a flight data analysis programme as part of its safety management system.*

This means that aircraft such as the BAe Jetstream 41, Saab 340 and ATR42-500 fall outside the recommendation. However, it is a *recommended practice* for types such as the Embraer ERJ 145, ATR 72-600 and Bombardier CRJ200, whereas it is *mandatory* for Bombardier Q400, Embraer 170, Avro RJ-85 and larger aircraft.

The ICAO requirements are reflected in European Commission Regulation No 859/2008 (European Aviation Safety Agency, 2008), specifically in OPS 1.037, which states:

> *4 . a flight data monitoring programme for those aeroplanes in excess of 27 000 kg MCTOM. Flight data monitoring (FDM) is the pro-active use of digital flight data from routine operations to improve aviation safety. The flight data monitoring programme shall be non-punitive and contain adequate safeguards to protect the source(s) of the data;*

In the "*Acceptable Means of Compliance and Guidance Material*" to the regulation (European Aviation Safety Agency, 2014), the overall intentions of FDM with regard to risk are stated:

> *Part ORO.AOC.130*
>
> *An FDM programme should allow an operator to:*
>
> *(1) identify areas of operational risk and quantify current safety margins;*
>
> *(2) identify* ***and quantify operational risks*** *by highlighting occurrences of non-standard, unusual or unsafe circumstances;* [emphasis added]

These two points are especially interesting because they require operators to quantify operational risks. FDM is very good at "*…highlighting occurrences of non-standard, unusual or unsafe circumstances",* however difficulty arises when trying to translate information on those occurrences to a measure of risk. For example, how does the airline analyst translate the typical 59 events listed in CAP 739 (Civil Aviation Authority, 2013) (e.g. Approach speed high below 500 ft AAL, High normal acceleration at landing) to the risks identified in the CAA "Significant Seven" (Civil Aviation Authority, 2011)? It is this *translation* that is the novelty of the work described in this paper.

The ICAO and EU regulations then filter down to the national level and in the UK are embodied in the "Air Navigation Order" (Civil Aviation Authority, 2009) Article 94. In addition, the UK CAA has produced extensive guidance material in CAP 739 (Civil Aviation Authority, 2013), providing aircraft operators with nearly 200 pages of detail of how FDM programmes can be implemented. It is a useful resource for airlines, analysts and safety managers embarking on FDM.

The ICAO regulations are reflected at a national level around the globe, however the US FAA is a notable exception. In the US, Flight Operations Quality Assurance (FOQA), as it is called there, is a voluntary safety programme and is described comprehensively in FAA Advisory Circular 120-82 (Federal Aviation Administration, 2004). Despite being voluntary, it has been widely adopted, with many large US carriers having implemented FOQA well before the ICAO regulations were established in 2005.

3.2 Risk assessment and risk management methods used in airline operations

3.2.1 Risk matrices

The use of risk matrices in flight operations safety is commonplace and follows ICAO's guidance in the Safety Management Manual (International Civil Aviation Organisation, 2018). In the latest edition (fourth), ICAO defines safety risk probabilities, as shown previously in Table 1, using qualitative descriptions. In a note to the original table, ICAO acknowledges that "…*organizations might include both qualitative and quantitative criteria*", but it is only the qualitative criteria for which examples are provided.

In contrast to later editions, the first edition of the Safety Management Manual (International Civil Aviation Organisation, 2006) did refer to quantifying risk probabilities in the table shown in Table 2. It would be impossible for ICAO to provide similar quantitative definitions for all sizes of airline operation, but at least it gave operators some idea of how to quantify likelihood and even perhaps produce their own definitions.

LIKELIHOOD	MEANING	QUANTITATIVE DEFINITION
FREQUENT	May occur once or several times during operational life.	1 to 10^{-3} per flight hour
REASONABLY PROBABLE	May occur once during total operational life of one system.	10^{-3} to 10^{-5} per flight hour
REMOTE	Unlikely to occur during the total operational life of each system but may occur several times when considering several systems of the same type.	10^{-5} to 10^{-7} per flight hour
EXTREMELY REMOTE	Unlikely to occur when considering several systems of the same type, but nevertheless has to be considered as being possible.	10^{-7} to 10^{-9} per flight hour
EXTREMELY IMPROBABLE	Should virtually never occur in the whole fleet life.	< 10^{-9} per flight hour

Table 2. Safety risk probability table from ICAO SMM 1st Edition

Note in Table 1 that a "Value" is given. Through a risk matrix (Figure 1), this value is combined with another value for the potential severity of an undesirable outcome to provide a risk index. In this case, the risk index is an alpha-numeric category, but in some matrices, numerical values are multiplied to provide a numerical "score". This may lead some safety managers to believe they are

actually quantifying risk because a numerical value is the result. However, as the index is based solely on qualitative assessments, it is impossible to produce a genuinely quantitative result.

Safety Risk	Severity				
Probability	Catastrophic A	Hazardous B	Major C	Minor D	Negligible E
Frequent - 5	5A	5B	5C	5D	5E
Occasional - 4	4A	4B	4C	4D	4E
Remote - 3	3A	3B	3C	3D	3E
Improbable - 2	2A	2B	2C	2D	2E
Extremely improbably - 1	1A	1B	1C	1D	1E

Figure 1: Typical 5x5 risk matrix. Source: ICAO Safety Management Manual, 4th Edition

Hubbard (2009) asserts that this type of risk assessment could be worse than doing no risk assessment at all because it can give management undue reassurance that something robust is being done. Anthony Cox (2008) provided a thorough examination of the weaknesses of risk matrices and the resulting risk assessments and cautions that "*Risk matrices do not necessarily support good (e.g. better than random) risk management decisions…*".

3.2.2 Bow tie diagrams

Bow tie diagrams are a useful way to communicate risk and they have been adopted by some airlines as part of their risk management processes. The concept is relatively simple and follows the principles of layered security made popular by the "Swiss-cheese" (Reason, 1997) accident causation model. At the centre of the bow tie (Figure 2) is a *hazard* connected to an *event*. This event is the undesirable situation occurring due to the release of the hazard.

Figure 2: Generic bow tie

On the far left of the bow tie are *threats* that can cause the hazard to be released into an event. To the right of the threats are *barriers* or *controls* to stop the threat releasing the hazard. Similarly to the Swiss-cheese model, these barriers may have different degrees of effectiveness. To the far right are *consequences*; these are the potential outcomes if the undesired event develops and represent some form of loss (e.g. injury, equipment damage, reputation damage). To the left of the consequences are recovery measures, i.e. further barriers or controls to try to prevent or minimise the undesirable consequences, given that the event has occurred. An example might be a runway

overrun area to minimise aircraft damage in the event of an excursion or an emergency response plan.

The barriers on both sides of the bow tie could be degraded in certain circumstances (e.g. a power-cut might render a warning system useless), and these are called *escalation factors*.

The UK CAA has produced a series of bow tie diagrams based on their "Significant Seven" risks (see CAA n.d.) and they are now widely used in airlines to help manage risk. The diagrams are an excellent way of communicating how risks are being managed and therefore are a valuable tool for demonstrating risk management to a regulator, for example, during an audit. A bow tie can be thought of as a Fault Tree and Event Tree, coupled together to represent the left and right-hand sides respectively. As such, it is possible to use quantitative information to produce bow ties, however those produced in airline operational environments are usually qualitative with little, if any, information relating to probabilities, even where data exists (e.g. FDM) that could provide this information.

3.2.3 ARMS methodology

The Airline Risk Management Solutions (ARMS) working group has produced a methodology (Aviation Risk Management Solutions, 2010) which, like bow ties, is based on barrier principles of safety management. The methodology comprises three main elements. Firstly, new individual safety events are classified – the "Event Risk Classification" (ERC). This is "…*a quick initial estimate on the risk inherent in the event*", which results in a subjective numerical value being assigned to the risk "….*which can be used in quantitative risk analysis".* It could be argued that the use of "quantitative" here is misleading; a numerical value is used, but it could just as well be a character, A, B or C. The values are later summed to produce a cumulative ERC score, but this could as easily be counts of A, B or C, hence it is not strictly *quantitative* as it depends on an initial qualitative assessment.

The second element is the analysis of a safety event database, comprising ERC-scored events, to produce safety trends and identify Safety Issues.

The third element is the "Safety Issue Risk Assessment" (SIRA). The Safety Issues identified are risk assessed and assigned a risk value. The intention then is that the risk assessment then leads to safety improvement actions. Safety Assessments can also be carried out using similar methods in advance of conducting a new activity.

The SIRA risk assessments are "*calculated*" using the following factors, quoted from Aviation Risk Management Solutions (2010):

- *Frequency/probability of the so-called Triggering Event*
- *Effectiveness of the Avoidance Barriers*
- *Effectiveness of the Recovery Barriers*
- *Severity of the (most probable) accident outcome*

The methodology is in use at several large airlines that were instrumental in its development, and one of the objectives was to keep it *"…easy to use and not create an unreasonable workload*", which it admirably achieves. However, while the attempt to quantify barrier effectiveness and the probabilities of "*triggering events*" is to be welcomed, the method still relies on some subjective assessment. The language used may mislead the user into believing the method is more quantitative in nature than it really is. Also, while the method has been broadly accepted by the operational community, literature and research on its effectiveness is sparse to non-existent.

Other areas of the aviation industry are well used to quantifying risk. In design and airworthiness, it is common to use techniques that do not rely solely on qualitative analysis.

3.3 Research based on flight data and FDM

Considering the length of time flight data and flight data monitoring has been around, there is surprisingly little published research on the topic. This is perhaps a symptom of the data being regarded as sensitive, and most FDM programmes have restrictions on the use and divulgence of FDM data. Prior to FDM becoming a mandatory practice, airlines would have to negotiate the introduction of FDM with pilot union groups who were afraid that the data could be misused. Usually, the negotiations resulted in quite strict protocols allowing only a limited number of "approved" individuals to have full access to identifiable[1] data. Despite FDM becoming a mandatory practice in 2005, this practice persists, as reflected in the CAP 739 guidance (Civil Aviation Authority, 2013), making it difficult for researchers to get access to data. Hopefully, this will change over time, and the vast amounts of data can be made more available, allowing deeper mining of the safety information within.

Fortunately, some researchers have acquired data and used it to find new methods for gaining safety insights. Several studies have focussed on addressing an inherent weakness in FDM; the reliance on predefined events and measures to detect anomalies. FDM has traditionally relied on exceedance events being defined in response to safety incidents or anticipated issues, and this has left gaps in the oversight it provides. A good example was an incident during a landing at Gibraltar in 2002 (Air Accidents Investigation Branch, 2014). Gibraltar can be a reasonably challenging airport to operate into, with a relatively short runway, and aircraft can experience turbulence off the nearby terrain during the final approach. On the 22nd May 2002, a Boeing 757 landed at Gibraltar and before the nosewheel touched down, the commander applied full nose-down elevator resulting in an excessive de-rotation. This caused the nosewheel to impact the ground hard, and the aircraft was significantly damaged. Through using historical flight data, the subsequent investigation found that the commander had developed a habit of applying nosewheel down input during landings, against the advice of the aircraft type's training manual. During the incident landing, the commander applied the input earlier than usual, resulting in the hard impact and damage. Had there been an FDM exceedance event in place to monitor excessive de-rotation, the operator may have been aware of the developing technique and could have intervened. Soon after publication of the report, FDM software and service providers implemented de-rotation-at-landing events into their systems.

The example shows how reliant FDM is on previous incidents; where a previous incident has not occurred, or no one has had the foresight to envisage a type of incident, it is unlikely that an FDM event will exist. This has led researchers to explore ways to avoid the predefinition of events and find ways to detect anomalies in data automatically.

Mugtussidis (2000) ran into the flight data access problem, having just four flights available. However, knowing the limitations of flight data, Mugtussidis proposed a sensible approach to flight data anomaly detection by limiting the analysis to twenty or so features rather than every parameter recorded. A Bayesian classifier was then proposed, which would aim to identify which flight a particular data point originated. If this were possible, it would be concluded that the data point was unusual. Unfortunately, the lack of data available severely restricted the opportunity to validate the approach.

Smart (2011) worked with an FDM service provider to gain access to a large dataset. Smart compared the performance of three one-class classifiers: Support Vector Machine (SVM), Mixture of

[1] i.e. where flight number and flight date are visible, providing the possibility of pilot identification

Gaussian, and K-means. It was found that the SVM performed best in classifying approaches to landing as abnormal and was also able to quantify the abnormality of the approach. Smart found that the SVM outperformed traditional event-based FDM. In common with Mugtussidis, Smart reduced the dimensionality of the flight data by selecting a limited set of features.

Li et al. (2011) used principal component analysis to reduce data dimensionality from 365 Boeing B777 flights before using a density-based clustering algorithm to identify anomalous flights. The algorithm, called DBSCAN, does not require an estimate of the number of clusters present in advance and is not sensitive to the shape of clusters. The results were promising, with abnormalities in flights identified such as low altitude on approach with high power, late alignment with runway, speed high. Traditional FDM exceedance monitoring would likely have also detected these issues, however the strength of this approach is that it did not require specific algorithms to be predefined for each anomaly found.

Clustering methods in flight data anomaly detection were further developed by Das et al. (2012) and the closely related work of Li et al. (2015). Li et al. developed an algorithm called cluster-based anomaly detection (ClusterAD) and compared it alongside multiple kernel anomaly detection (MKAD) and traditional FDM exceedance detection. The researchers had access to a good-sized dataset consisting of over 25,000 flights and took a subset of available recorded parameters for the period of 6NM on the approach to touchdown for comparison. The researchers found overlap in the anomalies detected by the three methods, and each method identified anomalies which the others did not. This suggests that the clustering methods should be used alongside exceedance detection rather than supersede it. An interesting difference in the approaches of exceedance detection and the clustering methods was highlighted by Das et al. Neither of the clustering methods detected short-term transgressions or deviations, whereas the exceedance detection did. Exceedance detection algorithms often have simple logic: if airspeed is greater than X and altitude less than Y, for duration Z, then event is true. If this deviation is corrected, an exceedance event is still triggered, even if the remainder of the flight is normal. Clustering methods can find longer-term anomalies, but because, for short-term deviations, the amount of normal values outweighs the number of anomalous values, no anomaly is found. This again highlights how clustering could be used to complement, rather than replace, exceedance detection.

The work of Li et al. (2015) was further developed in Li et al. (2016), where Gaussian Mixture Model based clustering was used to detect anomalous flights. Gaussian Mixture Models (GMM) are often used to represent subpopulations within a larger population and, therefore, can be applied to aircraft flight data. Within a number of flights, there will be variations in how aircraft are flown, depending on the conditions present. For example, an approach to landing will be flown differently in marginal visibility conditions (i.e. a precision approach) compared to one flown in good visibility (i.e. a visual approach). The approaches will differ in terms of timings of configuration changes, heading relative to runway, height relative to runway and so on. A GMM can represent these differences as different types of approaches within the overall population of flights. This allows the potential to identify anomalous subpopulations and anomalous flights that do not belong to any population. Li et al. transformed the flight data into a number of vectors, each relating to a flight parameter, of equal length and normalised the values. Imagining this as a large table with each column representing a specific flight parameter from a specific flight, each row will represent the population of all parameters from all flights at a particular time or distance from, say, touchdown. This row vector can then be used in cluster analysis. Li et al. carried out the analysis on 10,528 Airbus A320 flights and used four domain subject experts to validate the results. They concluded that the approach could identify flights with elevated risk and outperform existing methods for detecting "known unsafe events". The authors acknowledge that further work is required to determine performance for detecting unknown issues.

Other research focussing on anomaly detection has included Gorinevsky, Matthews and Martin (2012), Mendes (2012) and Nanduri and Sherry (2016), all of whom took different approaches.

Gorinevsky, Matthews and Martin (2012) used the residuals from a regression model to feed a multivariate statistical process control to identify anomalies. The regression model was trained using flight performance coefficients derived from FDM data. Care had to be taken to not include anomalies in the model, as would happen if the entire dataset was used for training. However, if only a small subset of the dataset was used for training, the normal variability of day-to-day aircraft operations would lead to spurious anomalies. A novel three-level regression approach was taken to address those issues. Also worth noting from Gorinevsky, Matthews and Martin is the highly efficient data processing they were able to achieve. Their method scanned for anomalies in 500,000 flights in just 10 hours. The anomalies found were generally related to individual flight data parameters, such as discrepancies in control surface position and sensor problems, rather than operational safety issues. Nonetheless, the anomalies detected could potentially result in safety issues and are, therefore, valuable knowledge.

Similarly to Smart (2011), Mendes (2012) used support vector machines (SVM) to highlight anomalies in aircraft autoland performance using recorded flight data. Principal component analysis was used on a training set of known good autolands, and then an SVM was used to classify the autoland's performance. Mendes' approach was able to filter 111 abnormal autolands from a set of 629, thus significantly reducing the manual effort required to investigate autoland performance.

The anomaly detection methods described so far have relied on some form of dimension reduction to make the method viable, however Nanduri and Sherry (2016) avoid this problem by using recurrent neural networks (RNN).

Standard artificial neural networks are inspired by biological nervous systems, with nodes (neurons) passing signals over connections. Each neuron can process the information sent to it and provide an output based on the input. Typically, there are multiple layers of neurons, with the input and output layers being separated by hidden layers. Artificial neural networks are very good at machine learning tasks and are widely used in fields such as image recognition and medical diagnosis. Their ability to "learn" without prior knowledge makes them extremely useful, and they can be highly accurate; however, the resulting classification decisions can be difficult, if not impossible, to trace back due to the opacity of the network. This is a problem for applications such as credit applications, where decisions can be appealed, however for classification of normal vs anomalous in the safety domain, it is usually not a problem, as the classifier is often just a means of directing an analyst to investigate potentially anomalous data.

Recurrent neural networks, as used by Nanduri and Sherry, allow the output from hidden layer neurons to be fed back as inputs to other neurons, thus allowing the use of past history, which is desirable when using time series data such as recorded flight data. They modified the typical RNN architecture by making use of long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Units (GRU) (Cho et al., 2014). Nanduri and Sherry claim that recurrent neural networks using LSTM and GRU do not share the limitations of MKAD or ClusterAD as described in Li et al. (2015) because "…*they* [RNN] *are capable of handling multivariate sequential input data without any modifications and treat it as time series data"*, hence potentially ideal for recorded flight data. Unfortunately, despite the existence in the United States of the Distributed FOQA Archive (Chidester, 2007) containing some 2 million flights, restrictions on the archive's use prevented the researchers from using it, so they had to simulate data for 500 flights (Nanduri and Sherry, 2016b).

Anomalies were introduced into the simulated data, and the anomaly detection performance of MKAD, LSTM and GRU were compared, finding that LSTM and GRU RNNs significantly outperformed MKAD.

While much of the literature on the use of FDM data focusses on anomaly detection, some researchers have used it to address specific operational problems. Haverdings and Chan (2009) used quick access recorder (QAR) data to study low-level windshear and turbulence, both significant aviation hazards, for aircraft operating at Hong Kong International Airport. Wang, Ren and Wu

(2018) also used QAR data to examine how pilots' landing techniques affected landing safety. They used an analysis of variance on 293 flights to determine the characteristics of landings and then two regression models to examine correlations between pilot operations and landing outcome. Unsurprisingly it was concluded that pilot inputs did effect touchdown distance and vertical acceleration at touchdown, but possibly of more value were their findings around pitch and thrust control technique, i.e. a positive and steady pitch up input, coupled with a relatively slow reduction in thrust lead to better landing outcomes.

Finding new methods to help detect anomalies is valuable research because it helps identify hazards that may otherwise remain hidden and unknown. Likewise, for other research looking at specific operational problems. Hazard identification is central to safety management, and even with the millions of hours of operational experience accumulated by the World's airlines, new hazards are still being identified thanks to anomaly detection and other methods. However, the hazards causing aircraft accidents today are usually well known and documented by individual airlines and the industry in general. At an industry level, the risk these hazards represent is quite easy to quantify by looking at accident statistics and operational data, but at an individual airline level, this risk quantification is much more difficult because, in most cases, the airline has not experienced a loss. Therefore, the airline must use accident precursor (i.e. causal factor) data to quantify the level of risk present in the operation, and it would seem that FDM data would be an obvious source. Indeed, FDM is very good at quantifying the prevalence of those causal factors or precursors in the operation, but how is that prevalence translated into a risk? How does landing fast with a crosswind affect the risk of a landing veer-off, for example? At which airports do those risk factors combine to regularly increase the risk of a runway veer-off? As it stands, FDM is not being used to help provide this risk quantification, and there is sparse to non-existent literature on the topic. Fortunately, there is literature on a method that could potentially make use of FDM data.

3.4 Bayesian networks and risk assessment

Literature (Calle-Alonso, Pérez and Ayra, 2019; Fenton and Neil, 2013) indicates that Bayesian networks are being used for risk assessment, albeit not commonly in airline operations safety. Bayesian networks can exploit mixtures of objective information (such as flight data) and subjective information, where no hard data is available.

In this paper, a method is developed to create a Bayesian network that makes use of flight data to quantify risk, and the resulting method is demonstrated using one of the seven significant risks: runway excursion. The paper compares two network learning algorithms, Bayesian Search and Greedy Thick Thinning, to assess the most appropriate for learning networks from flight data.

3.5 Runway veer-off events

Runway excursion incidents and accidents occur relatively frequently compared to other types of incidents, accounting for around a quarter (24%) of all incidents and accidents in air transport (Australian Transport Safety Bureau, 2008). The literature shows that a large proportion (over 40%) of these are landing veer-off events (Flight Safety Foundation, 2009), hence it is the risk chosen to demonstrate the method.

In a detailed assessment of veer-off events, Moretti et al. (2018) calculated a frequency of approximately 8.5 events per year, equivalent to 1 event per 7 million flights. The consequences of a veer-off can result in aircraft hull losses and, in some cases, fatalities: the 32 veer-offs examined by Moretti et al. resulted in 383 deaths.

3.6 Risk assessment of veer-off events

Moretti et al. (2018) looked at veer-off events from the airport's perspective and how infrastructure and airport layout can influence veer-off risk. This paper takes the perspective of the aircraft operator (e.g. an airline) and considers the risk of veer-off in relation to the operational performance of flights conducted by the operator. Current methods rely on identifying veer-off precursors and their prevalence. The best practice guidance from the European Operators Flight Data Monitoring Working Group B (2020) lists precursors which should be monitored for in a typical FDM programme. The list includes precursors such as:

- Crosswind
- Thrust asymmetry
- Poor visibility
- Unstable approach

It is, therefore, relatively easy for an aircraft operator to know the frequency of encountering these precursors, but a difficulty lies in the translation of that information into a risk assessment of a veer-off event. The basic precursor frequencies used today do not enhance our understanding of, say, the relative risk of landing from an unstable approach, or much more an aircraft is at risk if a landing is made in a crosswind in low visibility. This paper addresses that problem and provides a method whereby aircraft operators can quantify the relative risks under different conditions.

4 Bayesian network learnt from flight data

It is possible to derive a Bayesian network for risk assessment based on known accident causal factors. Accident investigation reports can be reliable sources of information on causal factors as they are typically produced by government bodies, following international protocols, and investigators are usually well-trained experts in their field. However, the quality of reports is not entirely consistent across different investigation bodies, and there can be significant differences in the methodologies and models used in the investigation process. Whilst efforts are made to avoid bias, it will inevitably occur in investigations due to, for example, the organisational culture of the investigating body and the previous experiences of the investigators involved.

A possible solution to these problems is to take a more naïve approach and not assume the network structure is simply all causal factors directed towards the outcome. Instead, the data itself may be able to inform the structure of the network. Methods to discover network structure from data have been around since the early 1990s (Buntine, 1991; Cooper and Herskovits, 1992; Spiegelhalter et al., 1993), and generally, the most popular methods use constraint-based or score-based algorithms. Constraint-based algorithms use conditional independence tests to discover which variables should be connected and those which should not. Examples are the PC, Grow-shrink and Incremental Association algorithms. Score-based algorithms attempt to find the network which has the highest probability, given the data. Examples are hill-climbing and simulated annealing. A full description of constraint-based and score-based algorithms can be found in Darwiche (2009) and Scutari and Denis (2015).

4.1 Flight data used in this research

The flight data used in this research was recorded via the integrated quick access recorder within the data management unit (DMU) on Airbus A319, A320 and A321 aircraft. All aircraft were configured to record a near-identical data frame consisting of around 370 recorded parameters.

The flight data was converted from its original raw format into R Object files (.rds). Any duplicate flights were removed and the quality of each flight assessed by checking for non-continuous data and missing take-off or landing points. This resulted in 310,448 flights being available.

The recorded flights were from Europe-based aircraft flying on a mostly European network, for a single airline and standard operating procedures were shared across the aircraft types.

Flight data, such as that used in FDM programmes, is a valuable source of operational information, but it is not without limitation. When using data to learn a Bayesian network for risk assessment, ideally there would be instances in the data of the risk event actually occurring (in this case, runway excursions), but despite large datasets, an example event may not be present. This is not surprising, as while runway veer-offs occur relatively frequently compared to other accidents, the likelihood of one occurring to a specific airline is low. It is, therefore, necessary to use an alternative outcome, the most obvious being a lateral deviation from the runway centreline. Ideally, either the aircraft's recorded GPS position or the localiser[2] deviation would be used during the landing roll. Unfortunately, both parameters' recorded resolution is too poor in this data set to be of any practicable use, so a proxy is needed.

The parameter 'lateral acceleration' is recorded at good resolution, a useable frequency (4Hz), and in the absence of the parameters described above, can be used as a proxy for lateral deviation. In most cases where an aircraft has deviated from the runway centreline during the landing roll, there will likely be disturbances in lateral acceleration as the deviation occurs and is recovered. Figure 3 shows a comparison between lateral acceleration values of a landing in a crosswind (a known veer-off causal factor) and a landing in benign conditions.

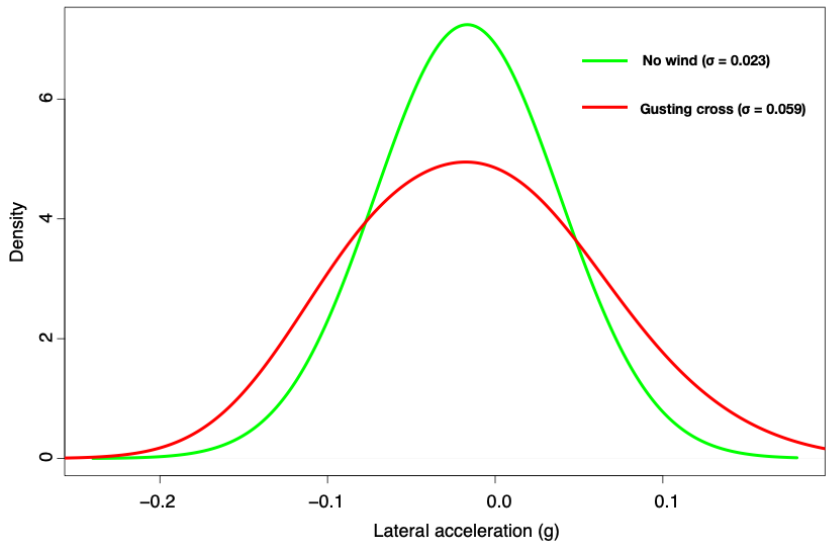

Figure 3: Density plots for lateral acceleration during the landing roll in crosswind and benign conditions

[2] A localiser provides lateral directional guidance for landing aircraft.

The density plots show higher variance in the parameter during a landing with a gusting crosswind, as expected. The standard deviation of lateral acceleration during the landing roll is used in this research as a proxy for a lateral deviation measure. This is supplemented by the standard deviation of the rudder deflection during the landing roll.

4.2 Data preparation for learning Bayesian network structure

Runway veer-off events occur after an aircraft has landed, however the lead up to the landing, known as the 'approach' phase, can influence the landing outcome. The approach phase is the period before landing where the aircraft is being correctly positioned, both laterally and vertically, so that a successful landing can be made. During the approach, the aircraft is configured for landing with high-lift devices (i.e. flaps, slats) and landing gear deployed, and energy is managed so that the aircraft arrives at the runway at the correct height and speed. Due to the influence of the approach phase on landing outcome, features from both the approach and landing are used.

In total, 12 features relating to runway veer-off causal factors are selected from each of the 310,448 flights to explore whether a meaningful and useful Bayesian network could be learnt from the data. The features were chosen because they are cited veer-off causal factors (Post, 2015), and they are widely available in airborne recorded flight data and weather archives: they are features readily available to a typical aircraft operator. Other features could be added if desired, for example, if the operator had more than one fleet, they may wish to use aircraft type as a feature, or perhaps features such as the experience of the handling pilot if that data is available and linked to flight data. In this case, the three aircraft types (A319, A320 and A321) are considered to be so similar that their differences would have little effect on veer-off risk, but they could be split and used as an additional 13th feature if desired.

The 12 selected features are:

- Recorded crosswind velocity component at landing – *Crosswind*
- Runway wet, derived from METAR (aviation routine weather report) reports – *Runway wet*
- Actual airspeed difference from target airspeed at 50ft radio height (indicative of an unstable approach) – *High speed*
- Maximum recorded normal acceleration during landing – *Landing g*
- Difference between heading at touchdown and runway heading – *Heading deviation*
- Glideslope deviation at 150ft (indicative of an unstable approach) – *Glideslope deviation*
- Standard deviation of lateral acceleration during the landing roll – *Lateral g s.d.*
- Duration of asymmetric thrust from touchdown -5 seconds to the end of the landing roll – *Asymmetric thrust*
- Gusting winds present (true/false), derived from METAR reports – *Crosswind gust*
- Visibility, derived from METAR reports – *Visibility*
- Airport ID – *Arrival airport*
- Standard deviation of rudder deflection during landing roll – *Rudder s.d.*

Any incomplete records were omitted from the dataset, thus avoiding potential problems with network learning algorithms that cannot cope with missing values. The 12 features from the remaining dataset (287,039 flights) were imported into GeNIe (BayesFusion LLC, 2019).

4.2.1 Addition of METAR data

The *weatherData* (Narasimhan, 2017) package in R was used to match landings to historical METAR observations. The observation closest to the landing time was used for wind and visibility. If either the observation closest to the landing time or the one before had a precipitation event, the runway was deemed to be wet.

4.2.2 Learning algorithms

There are many algorithmic schemes for learning Bayesian network structure from data. Beretta et al. (2018) conducted a quantitative comparison of some methods, finding that the choice of method can influence network accuracy and the tendency for under or over-fitting to occur. Tonda, Spritzer and Lutton (2014) describe structure learning methods, including heuristic algorithms, evolutionary approaches and memetic algorithms. The choice of algorithms used here is influenced by the availability of commercial implementations of them, as the general aim is to describe a method for risk assessment that can be used in an operational airline environment.

According to Tonda, Spritzer and Lutton (2014), two of the best heuristic algorithms are Bayesian Search (BS) and Greedy Thick Thinning (GTT), both of which are available in the commercial product GeNIe. The Bayesian Search algorithm is a hill-climbing procedure with random restarts, introduced by Cooper and Herskovits (1992). It is stochastic, so results will vary depending on the starting position of the algorithm.

The Greedy Thick Thinning algorithm, described by Cheng, Bell and Liu (1997), is based on the Bayesian Search approach and repeatedly adds arcs (thickening) between nodes and then removes them (thinning) to maximise the marginal likelihood of the observed data in each phase. The algorithm stops when no arc addition or removal results in an increase in the likelihood. GTT is deterministic, so the same results are derived for a given input.

Both BS and GTT return high-quality results in negligible time (Tonda, Spritzer and Lutton, 2014), making them good candidates for operational risk assessment, hence it is these two algorithms that are compared.

4.2.3 Discretisation of data

Both algorithms require the continuous variable to be discretised, and from the 12 extracted features, eight are continuous and require discretising. They are:

- *Crosswind* – discretised using Uniform Counts into four bins
- *High speed* - discretised using Uniform Counts into four bins
- *Landing g* - discretised using Uniform Counts into four bins
- *Heading deviation* - discretised using Uniform Counts into four bins
- *Glideslope deviation* - discretised using Uniform Counts into four bins initially, however, due to the high kurtosis of the empirical distribution, the threshold for the highest bin led to the inclusion of observations that were not significantly different from the mean. Therefore the threshold was manually moved to the right to 0.225 to mirror the lower bin threshold (-0.225). Even after this change, the bin still contained a good number of observations (43,553).
- *Lateral g s.d.* - discretised using Uniform Counts into four bins

- *Visibility* - due to this variable's high skewness, Uniform Counts resulted in the data being divided into just two bins. Uniform Widths was used instead, splitting the data into four bins with thresholds at 2.5km, 5km and 7.5km.

  Rudder s.d. - discretised using Uniform Counts into four bins

4.2.4 Background knowledge

The BS and GTT algorithms allow background knowledge to be specified before network structure learning. It allows the temporal relationship between variables to be specified, the result being that arcs from later to earlier variables are prohibited.

Temporal tier 1	**Temporal tier 2**	**Temporal tier 3**	**Temporal tier 4**	**Temporal tier 5**
Arrival airport	*Crosswind gust*	*Glideslope deviation*	*Landing g*	*Lateral g s.d.*
	Crosswind	*Asymmetric thrust*		*Rudder s.d.*
	Visibility	*High speed*		*Heading deviation*
	Runway wet			

Table 3: Temporal tiers for background knowledge

Table 3 shows the assignment of variables to each temporal tier. The assignment is based on the general order of occurrence of each feature, i.e. the approach to the specific airport (*Arrival airport*) comes first, followed by the environmental conditions at that airport, followed by features of the final approach, followed by a feature of the landing, followed by features from the landing rollout.

5 Results

5.1 Bayesian Search

GeNIe allows for several algorithm parameters to be changed. Most were left at their default values, however '*Iterations*' was set to 100; this represents the number of restarts of the algorithm allowed and the developers of GeNIe suggest making this as large as practicable in terms of processing time. The other default parameter which was changed was '*Max Time (seconds)*', which is the maximum time the algorithm will run. Due to the large dataset, this limit was set to 600s as a starting point. The sample size, which controls how fast the data changes the posterior probability of the network (Spirtes and Meek, 1995), was left at the default value of 50.

5.1.1 Bayesian Search: sample size = 50

The learning algorithm was executed, and the resulting network is shown in Figure 4.

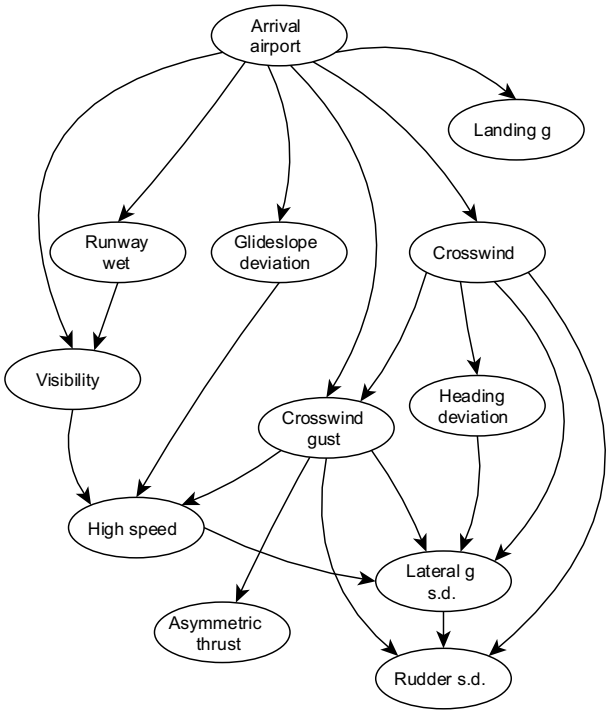

Figure 4: Learnt network using Bayesian Search algorithm

The accompanying metrics reported by GeNIe were:

Best score in iteration 52: -3.50026e+06

EM Log-Likelihood[3]*: -3.49206e+06*

[3] EM Log Likelihood is a metric based on the expectation-maximisation algorithm. It can be used to compare networks learnt with different parameters. Higher values indicate that the network is more consistent with data it was learnt from.

These metrics can be used to compare the network's performance in representing the data, i.e. the learning parameters can be altered and the resulting networks compared.

5.1.1.1 Qualitative evaluation

It is interesting to note that neither *Landing g* nor *Asymmetric thrust* affects any other parameters (i.e. they have no children nodes) according to the Bayesian Search algorithm. It appears that *Landing g* is influenced by the arrival airport (*Arrival airport*), which is reasonable, however, *Asymmetric thrust* has *Crosswind gust* (gusting crosswind present) as a parent, which does not make much sense in causal terms given that asymmetric thrust should not be a technique used on this aircraft type in crosswind conditions. However, it was found that gusting crosswinds were present in 4.6% of landings with asymmetric thrust, compared to 2.8% in those where it was not present. This suggests that there might be different techniques for thrust management being used by some pilots in some circumstances.

Another feature of this network is that it introduces an arc between *Visibility* and *High speed*. The effect of visibility on speed might not be immediately apparent, but it can be seen by introducing evidence at the visibility node and seeing how the speed probabilities change. Setting evidence at the *Visibility* node to the lowest visibility, the probability of *High speed* being high reduces. This would be expected because low visibility approaches are more often flown automatically, usually resulting in less speed deviation from the target.

There is an arc to *Visibility* from *Runway wet*, which in turn has an arc from *Crosswind gust*, which could indicate that poor weather, such as gusty and wet conditions, are influencing the *High speed* node. Approaches are flown faster in gusty conditions, and more variability around the target speed would be expected.

Elsewhere, the learnt network follows a mostly logical and causal flow:

- Weather-related nodes are dependent on the arrival airport – some airports are likely to experience more inclement weather than others.
- *High speed* has *Crosswind gust* and *Crosswind* as parents as expected, and also *Glideslope deviation*. Approaches that are high are quite often fast too.
- *Heading deviation* has an arc to *Lateral g s.d.*, which in turn has an arc to *Rudder s.d.*.

5.1.1.2 Quantitative evaluation

Table 4 shows the strength of influence between the nodes, sorted by the average distance '*…between the various conditional probability distributions over the child node conditional on the states of the parent node*' (BayesFusion LLC, 2019):

635

Parent	Child	Average distance
Runway wet	*Visibility*	0.812903
Lateral g s.d.	*Rudder s.d.*	0.80508
Arrival airport	*Glideslope deviation*	0.744525
Arrival airport	*Crosswind*	0.723703
Crosswind gust	*Runway wet*	0.711269
Arrival airport	*Landing g*	0.700233
Crosswind gust	*High speed*	0.698689
Crosswind	*Rudder s.d.*	0.691147
Crosswind	*Lateral g s.d.*	0.657746
Crosswind	*Heading deviation*	0.64943
Crosswind gust	*Lateral g s.d.*	0.638561
Arrival airport	*Visibility*	0.615769
Heading deviation	*Lateral g s.d.*	0.60204
Arrival airport	*Crosswind gust*	0.543608
Crosswind	*High speed*	0.508641
Visibility	*High speed*	0.505338
Crosswind gust	*Rudder s.d.*	0.436036
Glideslope deviation	*High speed*	0.408371
Arrival airport	*Runway wet*	0.375158
High speed	*Lateral g s.d.*	0.369542
Crosswind	*Crosswind gust*	0.274417
Crosswind gust	*Asymmetric thrust*	0.0442262

Table 4: Strength of influence between nodes, sample size = 50

Note: The table is presented here to show the relative strengths of the arcs based on the J-Divergence distance as advocated by Koiter (2006); for an explanation of the derivation of the values, the reader is directed to Koiter (2006).

Values closer to 1 represent a strong influence between nodes, whereas values towards zero represent weak influence.

The arc between *Crosswind gust* and *Asymmetric thrust* is weak, and *Runway wet* to *Visibility* and *Lateral g s.d.* to *Rudder s.d.* are strong influences. Relatively strong influences are also shown between *Crosswind* and *Rudder s.d.*, *Lateral g s.d.* and *Heading deviation,* which is intuitive.

5.1.2 Bayesian Search: sample size = 500

To investigate the effect of sample size, the network learning was carried out again on the same dataset, with the same GeNIe parameters except that the sample size was set to 500. The resulting network is shown in Figure 5.

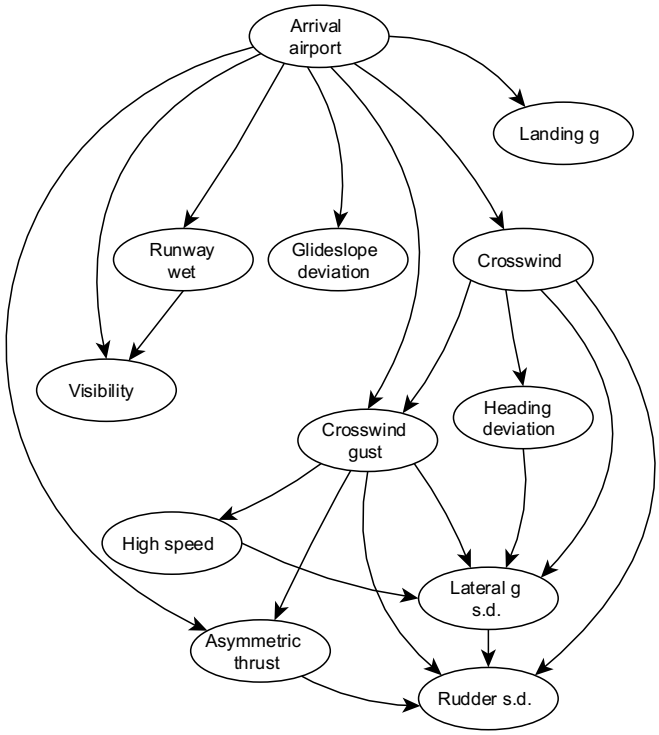

Figure 5: Bayesian Search learnt network, sample size = 500

The accompanying metrics reported by GeNIe were:

Best score in iteration 92: -3.49836e+06

EM Log Likelihood: -3.48774e+06

5.1.2.1 Qualitative evaluation

The resulting network is slightly simpler than the previous example, with 20 arcs compared to 22. The EM Log-Likelihood score has reduced (closer to zero is an improvement). The arc between *Gslope50* and *High speed* has now gone, but a new arc between *Asymmetric thrust* and *Rudder s.d.* has been introduced. The arc between *Visibility* and *High speed* has also gone. Overall, this network seems to be a slight improvement over the previous example as it retains expected causal links whilst being a simpler structure.

5.1.2.2 Quantitative evaluation

Calculating the strengths of the arcs gives:

Parent	Child	Average distance
Crosswind gust	*Asymmetric thrust*	0.873009
Runway wet	*Visibility*	0.812903
Arrival airport	*Glideslope deviation*	0.744525
Arrival airport	*Crosswind*	0.723703
Arrival airport	*Landing g*	0.700233
Arrival airport	*High speed*	0.674521
Crosswind	*Heading deviation*	0.64943
Arrival airport	*Visibility*	0.615769
Asymmetric thrust	*Rudder s.d.*	0.590711
Crosswind	*High speed*	0.582155
Crosswind	*Lateral g s.d.*	0.578374
Arrival airport	*Crosswind gust*	0.543608
Heading deviation	*Lateral g s.d.*	0.536829
Arrival airport	*Asymmetric thrust*	0.52684
Lateral g s.d.	*Rudder s.d.*	0.504357
Crosswind	*Rudder s.d.*	0.482741
Crosswind	*Runway wet*	0.298224
Crosswind	*Crosswind gust*	0.274417
High speed	*Lateral g s.d.*	0.171547
Arrival airport	*Runway wet*	nan

Table 5: Strength of influence between nodes, sample size = 500

It is not apparent from GeNIe why the average strength for *Arrival airport* to *Runway wet* has been calculated as 'nan' (not a number), however other distance measures in GeNIe (e.g. Euclidean) suggest a weak influence, similar to that in the previous network. The *Crosswind gust* and *Asymmetric thrust* pairing has been strengthened with the increase in sample size.

Overall the network is slightly simpler with a marginally better log-likelihood, however the changes to the strength of influence of the arcs lead to a conclusion that it is not a significant improvement.

5.1.3 Bayesian Search: sample size = 5000

Further increasing the sample size to 5000 yields the following network.

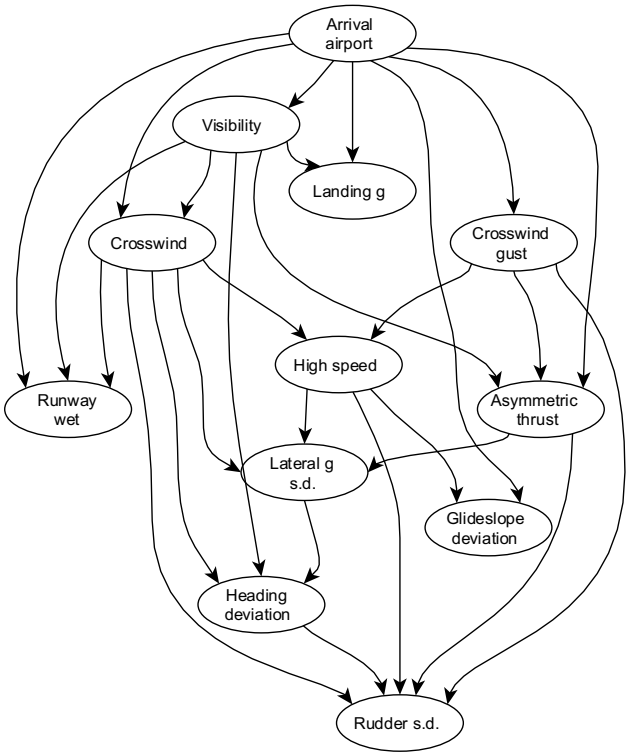

Figure 6: Bayesian Search learnt network, sample size = 5000

The network has increased in complexity, with 30 arcs, and the log-likelihood has reduced to -3.51105e+06, so increasing the sample size has not improved the network. It appears that increasing the sample size possibly leads to overfitting of the network, with some arcs being introduced where no causal influence would be expected, e.g. *Visibility* to *Asymmetric thrust*, *Crosswind gust* to *Visibility*.

5.1.4 Summary of the effect of sample size

Sample sizes of 250, 375, 750 and 1000 were also tried, and the results are summarised in Table 6.

SAMPLE SIZE	ARCS	EM LOG-LIKELIHOOD
50	22	-3.49206e+06
250	21	-3.48748e+06
375	20	-3.48794e+06
500	20	-3.48774e+06
750	21	-3.48729e+06
1000	22	-3.48708e+06
5000	30	-3.51105e+06

Table 6: Summary of effect of Bayesian Search sample size on the learnt network

Mid-table, there is little variation, and the sample size changes and the number of arcs are at a minimum of around 375 to 500, so a sample size of 500 seems like a reasonable compromise.

5.1.5 Bayesian Search summary

Bayesian Search is a popular algorithm for learning Bayesian networks, and it has produced several candidate networks from the observed flight data. The resulting number of arcs, and thus the model's overall complexity, depends on the number of samples used. Likewise, the EM Log-Likelihood score is also influenced by the number of samples, and it seems that, in this case, somewhere between 300 and 500 samples would be a good compromise, generating less complicated networks, but with good likelihood scores when compared to very low and very high sample sizes.

The Bayesian Search using 500 samples, shown in Figure 5, appears to be the best network generated by this algorithm, and it could be further enhanced with expert input by carefully removing questionable weak arcs and possibly introducing one or two others.

5.2 Greedy Thick Thinning

Greedy Thick Thinning is closely related to the Bayesian Search algorithm. The network learnt by the algorithm, using the same background knowledge as before, is shown in Figure 7.

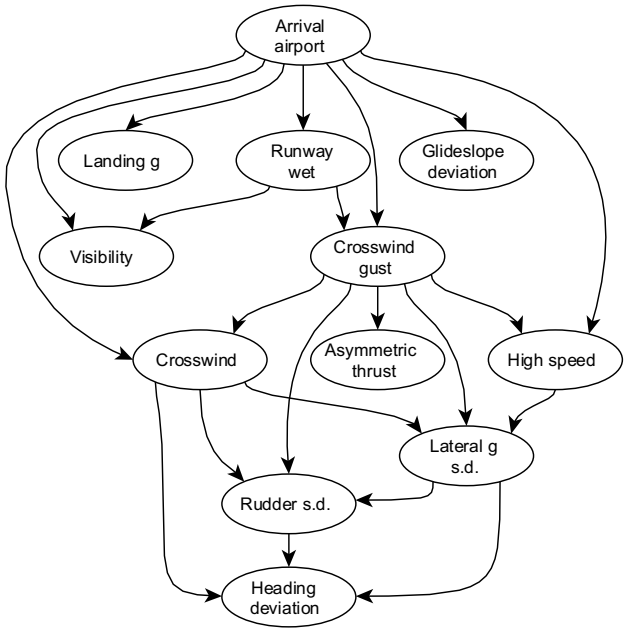

Figure 7: Network learnt using Greedy Thick Thinning algorithm

The resulting network is similar to the networks learnt by Bayesian Search. In total, there are 21 arcs, and they mostly follow an intuitive causal logic. Table 7 shows the strength of influence of the arcs.

Parent	Child	Average distance
Runway wet	*Visibility*	0.812903
Lateral g s.d.	*Rudder s.d.*	0.80508
Arrival airport	*Glideslope deviation*	0.744525
Crosswind gust	*Crosswind*	0.740155
Crosswind gust	*High speed*	0.739413
Crosswind	*Lateral g s.d.*	0.710408
Arrival airport	*Landing g*	0.700233
Crosswind	*Rudder s.d.*	0.691147
Arrival airport	*Visibility*	0.615769
Arrival airport	*Crosswind*	0.586804
Lateral g s.d.	*Heading deviation*	0.580175
Arrival airport	*High speed*	0.578471
Crosswind gust	*Lateral g s.d.*	0.575704
Runway wet	*Crosswind gust*	0.571688
Arrival airport	*Crosswind gust*	0.481021
Crosswind	*Heading deviation*	0.465797
Crosswind gust	*Rudder s.d.*	0.436036
Arrival airport	*Runway wet*	0.397705
High speed	*Lateral g s.d.*	0.286439
Rudder s.d.	*Heading deviation*	0.129406
Crosswind gust	*Asymmetric thrust*	0.0442262

Table 7: Strength of influence of arcs from Greedy Thick Thinning algorithm

This network also has an arc between *Crosswind gust* and *Asymmetric thrust*, however it has a very weak influence compared to the other arcs.

Overall, Greedy Thick Thinning appears to work well with this dataset. The resulting network has a mostly logical causal flow and is not overly complex, however it does suffer from several nodes which are children only, e.g. *Landing g* and *Glideslope deviation*. This means they do not influence the lateral deviation nodes.

5.3 Sensitivity analysis

Sensitivity analysis can be used to tune parameters of a Bayesian network to provide a desired output, so for example, in this network, sensitivity analysis could be used to determine which parameter needs to be changed, and by how much, to result in an X% drop in veer-off probability (Darwiche, 2009). It is also a method often used to evaluate the output of Bayesian networks, especially when there might be some uncertainty over the model, particularly the structure, or when it has been necessary to elicit expert opinion to help build the network.

Sensitivity analysis shows which nodes have the most influence on any particular target node, and if it is shown that a node has an unexpectedly strong influence, there may be a problem with the

definition of that node or perhaps the structure of the network. That is not to say that all nodes should have equal influence; in the case of a causal Bayesian network, it would be expected that the most prevalent causes would have the most influence.

There are several methods for sensitivity analysis (Castillo, Gutierrez and Hadi, 1997; Jensen, Aldenryd and Jensen, 1995), some of them extremely complex, but for this analysis, the built-in function in GeNIe will be used, which follows the work of Kjærulff and van der Gaag (2000). The method highlights nodes for which a small change in a parameter results in a large change in a target node's posterior probability.

5.3.1 Sensitivity analysis of Bayesian Search network

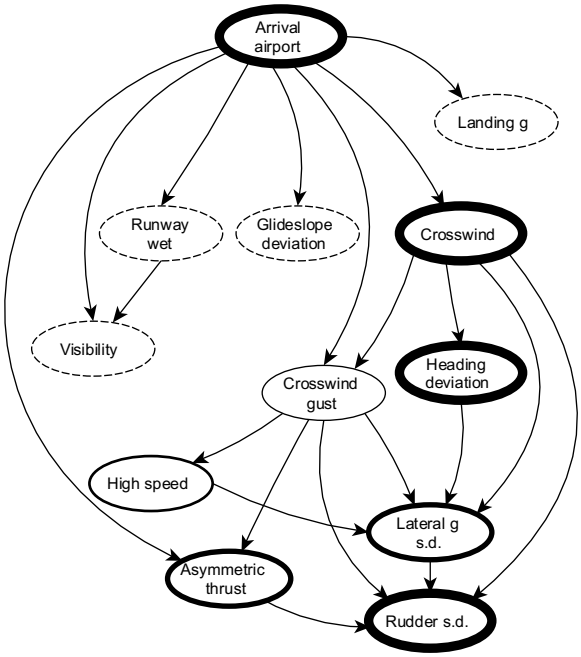

Figure 8: Sensitivity analysis of Bayesian Search learnt network

The nodes' line weight in Figure 8 depicts how sensitive the target nodes are to changes in parameters in the variable nodes. Heavy line weight depicts the most important for the calculation of posterior probabilities in the target nodes, thinner lines less important and dotted nodes have no effect.

The sensitivity analysis of the Bayesian Search learnt network confirms that several nodes are not used in the calculation of the posterior probabilities of the lateral deviation target nodes (i.e. *Lateral g s.d.*, *Rudder s.d.* and *Heading deviation*). This is apparent from the network structure as those nodes have no route to the target nodes. The target nodes are most sensitive to *Arrival airport* and *Crosswind*. A concern in this network is that *Crosswind gust* has little influence over the lateral deviation nodes; it would be expected that a gusting crosswind would influence lateral deviation.

5.3.2 Sensitivity analysis of Greedy Thick Thinning network

The sensitivity analysis of the Greedy Thick Thinning network shows similar results to the Bayesian Search sensitivity analysis, however this time, the deviation nodes are also sensitive to changes in *Crosswind gust*.

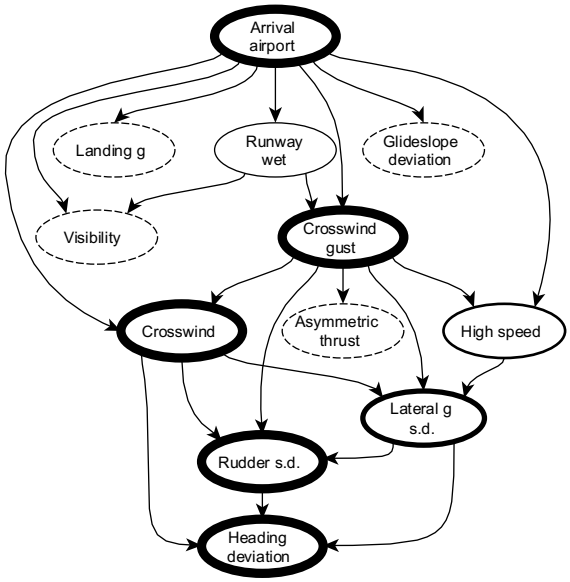

Figure 9: Sensitivity analysis of Greedy Thick Thinning network

Generally, the analysis results are unsurprising, with crosswinds being the leading factor in changes to the lateral deviation probabilities.

5.4 Validation of learnt networks

Cross-validation is often used to validate learnt networks and parameters, and there are various methods available (e.g. leave one out (LOOCV), repeated random sub-sampling). A popular and easy to understand method is k-fold cross-validation. K-fold cross-validation can be less computationally burdensome than LOOCV however it can suffer from more bias, although this can be mitigated to some extent by the choice of *k*.

In k-fold cross-validation, the dataset is split into *k* number of subsets, one of which is used as the validation set, and the remaining *k – 1* subsets used for training. The cross-validation process is repeated k times, using each validation subset once. There are different opinions about the optimal value of *k*; too low a value can result in biased results, whereas high values can result in excessive computing time. A value of *k = 10* is generally regarded as sensible (James et al., 2013; Kuhn and Johnson, 2013).

The following sections show the results of cross-validation carried out on the Bayesian Search network learnt previously.

K-fold cross-validation was carried out using *k = 10*, with *Lateral g s.d.* set as the class node in GeNIe and the *Uniformize* EM option selected. The validation dataset was the same as that used to learn the network, discretised as per 4.2.3.

5.4.1 Bayesian Search network validation

Table 8 shows the validation results for the Bayesian Search learnt network.

Overall	0.478729 (137414/287039)
s1_below_0	0.67232 (48245/71759)
s2_0_0	0.291119 (20891/71761)
s3_0_0	0.255947 (18367/71761)
s4_0_up	0.695546 (49911/71758)

Table 8: Bayesian Search network validation results

The table shows how good the network is at representing observations in each of the four discretisation groupings. The last row labelled *s4_0_up* represents the highest level of standard deviation of lateral acceleration (*Lateral g s.d.*), i.e. the 25% highest observations.

The overall accuracy of the network at predicting *Lateral g s.d.* was just under 48%, however it did much better at predicting the top 25% of occurrences (labelled *s4_0_up*) with an accuracy of nearly 70%.

		Predicted			
		S1_below_0	S2_0_0	S3_0_0	S4_0_up
Actual	S1_below_0	**48245**	14925	6120	2469
	S2_0_0	25791	**20891**	14605	10474
	S3_0_0	11114	16732	**18367**	25548
	S4_0_up	2517	6829	12501	**49911**

Table 9: Validation confusion matrix for Bayesian Search network

Table 9 shows the validation confusion matrix for the Bayesian Search network. The matrix shows correct predictions along the diagonal in bold and errors in the other cells.

Receiver operating characteristic (ROC) curves show a model's ability to classify binary systems, for example, in classifying correct versus false predictions. The ROC curve in Figure 10 shows good performance for the Bayesian Search network in predicting the highest *Lateral g s.d.* values. The grey diagonal represents random performance and, informally, the closer the curve is to the top left corner, the better the performance. The area under the curve (AUC) is around 0.84; perfect performance would result in AUC = 1. A full explanation of ROC curves can be found in Carvalho et al. (2014) and Fawcett (2006).

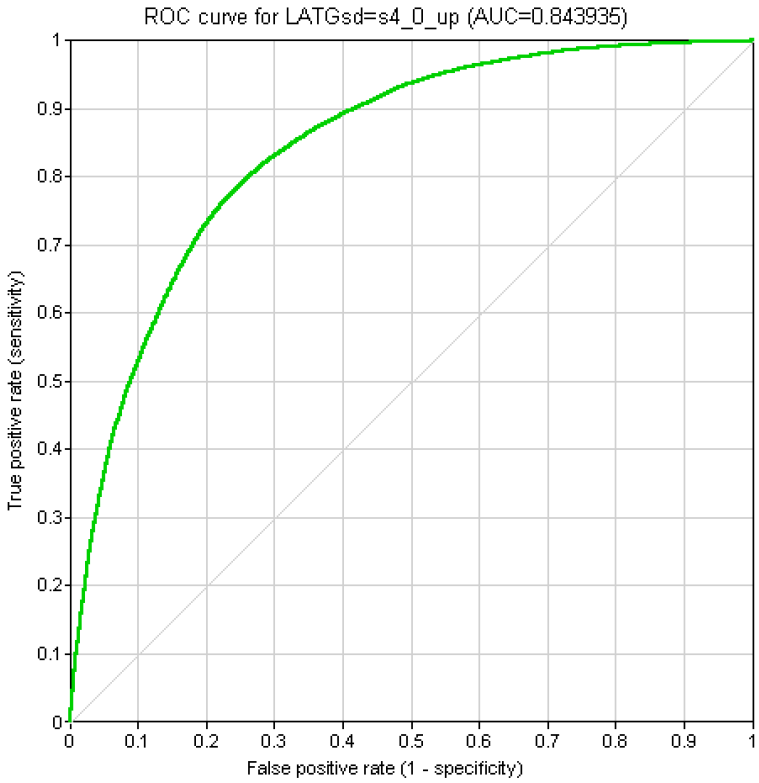

Figure 10: ROC curve for Bayesian Search network validation

5.4.2 Greedy Thick Thinning network validation

Validation of the Greedy Thick Thinning network shows very similar performance to that of the Bayesian Search network, with an overall accuracy of 48% and accuracy for the highest interval of *Lateral g s.d.* of 70%. The confusion matrix and the ROC curve are also very similar.

Overall	0.479297 (137577/287039)
s1_below_0	0.650692 (46693/71759)
s2_0_0	0.311381 (22345/71761)
s3_0_0	0.25741 (18472/71761)
s4_0_up	0.69772 (50067/71758)

Table 10: Greedy Thick Thinning network validation results

		Predicted			
		S1_below_0	S2_0_0	S3_0_0	S4_0_up
Actual	S1_below_0	**46693**	16499	6050	2517
	S2_0_0	24319	**22345**	14531	10566
	S3_0_0	10195	17468	**18472**	25626
	S4_0_up	2248	6978	12465	**50067**

Table 11: Validation confusion matrix for Greedy Thick Thinning network

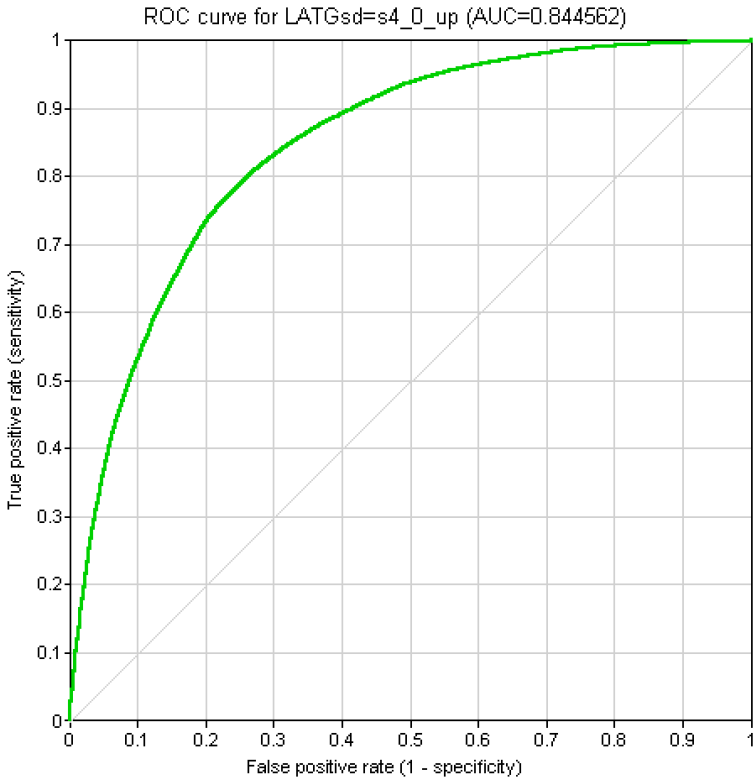

Figure 11: ROC curve for Greedy Thick Thinning network validation

5.4.3 Summary of validation results

The Bayesian Search and Greedy Thick Thinning algorithms performed similarly; however, the latter is marginally better: the Bayesian Search network's sensitivity analysis showed that *Lateral g s.d.* was not very sensitive to gusting crosswinds (*Crosswind gust*), which is counter-intuitive. This problem did not arise in the Greedy Thick Thinning network.

6 Practical application

A specific airport is introduced as evidence at the *Arrival airport* node, and the probability of the most severe cases of *Lateral g s.d.*, *Rudder s.d.* and *Heading deviation* are calculated. The following tables show the results.

Airport - BHD	***Lateral g s.d.***	***Rudder s.d.***	***Heading deviation***
Bayesian Search	0.251	0.253	0.271
GTT	0.254	0.256	0.272

Airport - DLM	***Lateral g s.d.***	***Rudder s.d.***	***Heading deviation***
Bayesian Search	0.143	0.122	0.171
GTT	0.142	0.121	0.165

Airport - DME	*Lateral g s.d.*	*Rudder s.d.*	*Heading deviation*
Bayesian Search	0.219	0.213	0.247
GTT	0.220	0.215	0.246

Airport - EDI	*Lateral g s.d.*	*Rudder s.d.*	*Heading deviation*
Bayesian Search	0.184	0.163	0.200
GTT	0.186	0.166	0.201

Airport - LHR	*Lateral g s.d.*	*Rudder s.d.*	*Heading deviation*
Bayesian Search	0.273	0.277	0.290
GTT	0.270	0.275	0.290

Table 12: Tables showing probabilities of high Lateral g s.d., Rudder s.d. and Heading deviation, for different airports, derived by different learning algorithms

The Bayesian Search and Greedy Thick Thinning learnt networks produce very similar probabilities for the lateral deviation nodes, unsurprisingly given how closely related the algorithms are.

The probabilities shown in the tables above could facilitate a quantitative risk assessment of where a lateral deviation is more likely. Coupled with an assessment of the consequences of a lateral deviation (e.g. cliff edge close to a runway, such as Trabzon or Funchal), these values allow an airline to quantitatively assess the risk, based on objective data rather than opinion or guesses on which current methods depend.

It is also possible to run different scenarios through the network by introducing evidence or known quantities at specific nodes. For example, an airline could quantify the effect of eradicating unstable approaches (e.g. excessive speed) and decide whether the effect warrants the effort. Table 13 shows example scenarios where evidence has been added to the GTT learnt BN. It shows how the presence of veer-off causal factors changes the probability of experiencing high *Lateral g s.d.,* i.e. P(high Lateral g s.d.).

Factors present	*P(high Lateral g s.d.)*	*Increase*
Baseline probability, no evidence added	0.249	
Wet	0.251	x 1.01
Wet + Xwind	0.5	x 2.01
Wet + Xwind + Gust	0.724	x 2.91
High speed	0.299	x 1.20
High speed + Xwind	0.566	x 2.27
High speed + Xwind + Gust + Wet	0.770	x 3.09

Table 13: Comparison of scenarios using GTT learnt BN

For example, a nearly threefold increase in *P(high Lateral g s.d.)* occurs if the runway is wet and a gusting crosswind is present. The author is not aware of any methods in use in airlines today that can provide these relative probabilities, hence the capability to conduct quantitative risk assessments are limited for this kind of event. A BN, such as this, fed with flight data from an airline's FDM programme along with meteorological data, is a viable solution to that problem.

For day-to-day operations, the network could be used dynamically with live feeds of weather forecasts, and the latest result data from the airline's FDM programme, to provide a contemporary

risk assessment. This could be published to pilots during pre-flight briefings so they could be forewarned of potentially higher risk. In mitigation, they may decide, for example, that the more experienced pilot should fly the approach and landing.

7 Discussion

Qualitative forms of risk assessment methods dominate in airline operations, despite the availability of data sources, such as FDM, which could feed quantitative analyses. As described in the Literature Review, regulators expect FDM to be used to quantify risks (European Aviation Safety Agency, 2014), however the three main methods in use (risk matrices, bow ties and ARMS) are mostly qualitative and rely on subjective judgement.

This paper has demonstrated how learning Bayesian networks can be applied to airborne recorded flight data to help quantify risk in airline operations and satisfy regulatory expectations. Selecting and extracting features directly related to known risk causal factors allowed a very large dataset of over 300,000 flights to be used efficiently for network learning. This data pre-exists in most airlines through FDM, however it is often under-utilised and this paper offers a method whereby airlines can exploit their data more thoroughly.

The output from the method is a Bayesian network which will allow airlines to quantify the likelihood of an event and run scenario analyses to assess the likely effect of risk mitigation strategies. The method, demonstrated here using runway veer-off risk, can be applied to other high-level risks where features from flight data can be extracted, including loss of control inflight, controlled flight into terrain and longitudinal runway excursions.

The paper has compared two learning algorithms, Bayesian Search and Greedy Thick Thinning. The algorithms are closely related and produced similar networks from the data, however a subjective analysis of the resulting networks led to the conclusion that GTT is preferred. GTT also has the benefit of being deterministic, generating the same structure each time from a given set of data, and this could be important in a safety-critical domain where the stochastic nature of BS could potentially undermine confidence in results amongst practitioners unfamiliar with BN learning.

This paper's main finding is that through the use of existing data sources, such as FDM and weather databases, airlines can complement their existing qualitative risk assessment methods for operational safety with quantitative methods, such as Bayesian networks. This paper describes how this can be achieved using commercially available software.

A note of flight data parameter availability

This research has highlighted that despite taking flight data from a modern aircraft fleet, there was no direct measure for lateral deviation distance from the runway centreline recorded within the flight data, so a proxy parameter was used. It is acknowledged that accident investigation drives flight data recording requirements, and often for investigation purposes, parameters with high resolution and recording frequencies are not necessary. However, higher quality parameters are valuable and necessary for research into accident prevention, so it is recommended that these requirements are considered in future flight data parameter specifications.

8 References

Air Accidents Investigation Branch (2014) *Boeing 757-2T7, G-MONC, 22 May 2002 - GOV.UK*., *AAIB Bulletin* Available at: https://www.gov.uk/aaib-reports/boeing-757-2t7-g-monc-22-may-2002 (Accessed: 19 December 2018).

Australian Transport Safety Bureau (2008) 'Runway Excursions', *Aviation Research and Analysis Report*, 018(1), pp. 1–103.

Aviation Risk Management Solutions (2010) *ARMS Methodology for Operational Risk Assessment in Aviation Organisations*. Available at: http://www.skybrary.aero/bookshelf/books/1141.pdf (Accessed: 20 December 2015).

BayesFusion LLC (2019) *GeNIe Modeler* 2.3.1.

Beretta, S., Castelli, M., Gonçalves, I., Henriques, R. and Ramazzotti, D. (2018) 'Learning the structure of Bayesian networks: A quantitative assessment of the effect of different algorithmic schemes', *Complexity*

Buntine, W.L. (1991) 'Theory Refinement on Bayesian Networks', D'Ambrosio, B. and Smets, P. (eds.)*Proceedings of the Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-91)*. University of California at Los Angeles, Los Angeles, CA, USA, Vol.UAI1991, pp. 52–60.

Calle-Alonso, F., Pérez, C.J. and Ayra, E.S. (2019) 'A Bayesian-Network-based Approach to Risk Analysis in Runway Excursions', *Journal of Navigation*, pp. 1–19.

Castillo, E., Gutierrez, J.M. and Hadi, A.S. (1997) 'Sensitivity analysis in discrete Bayesian networks', *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(4), pp. 412–423.

Cheng, J., Bell, D.A. and Liu, W. (1997) 'An algorithm for Bayesian belief network construction from data', *Proceedings of AI & STAT'97*., pp. 1–8.

Chidester, T.R. (2007) Voluntary Aviation Safety Process : Preliminary Audit of Distributed FOQA and ASAP Archives Against Industry Statement of Requirements *DOT/FAA/AM-07/7*.

Cho, K., van Merrienboer, B., Bahdanau, D. and Bengio, Y. (2014) 'On the Properties of Neural Machine Translation: Encoder-Decoder Approaches', Wu, D., Carpuat, M., Carreras, X. and Vecchi, E. M. (eds.)*Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, Vol.abs/1409.1, pp. 1–9.

Civil Aviation Authority (2011) *CAA 'Significant Seven' Task Force Reports*. Available at: https://publicapps.caa.co.uk/docs/33/2011_03.pdf (Accessed: 13 March 2017).

Civil Aviation Authority (2013) *CAP739 - Flight Data Monitoring*. Available at: http://publicapps.caa.co.uk/docs/33/CAP739.pdf (Accessed: 18 February 2018).

Civil Aviation Authority (2009) *Air Navigation Order 2009*. 2009 No. 3015.

Civil Aviation Authority (no date) *UK CAA Bowtie guidance*. Available at: http://www.caa.co.uk/Safety-Initiatives-and-Resources/Working-with-industry/Bowtie/ (Accessed: 16 July 2018).

Cooper, G.F. and Herskovits, E. (1992) 'A Bayesian Method for the Induction of Probabilistic Networks from Data', *Machine Learning*, 9(4), pp. 309–347.

Cox, L.A. (2008) 'What's Wrong with Risk Matrices?', *Risk Analysis*, 28(2), pp. 497–512.

Darwiche, A. (2009) *Modeling and Reasoning with Bayesian Networks*. Cambridge: Cambridge

University Press.

Das, S., Li, L., Srivastava, A. and Hansman, R.J. (2012) 'Comparison of Algorithms for Anomaly Detection in Flight Recorder Data of Airline Operations', *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. Reston, Virigina: American Institute of Aeronautics and Astronautics.

European Aviation Safety Agency (2008) *COMMISSION REGULATION (EC) No 859/2008*. *Annex 3*.

European Aviation Safety Agency (2014) *Acceptable Means of Compliance (AMC) and Guidance Material (GM) to part -ORO*.

European Operators Flight Data Monitoring Working Group B (2020) *Guidance for the implementation of FDM precursors Rev 03*.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, 27(8), pp. 861–874.

Federal Aviation Administration (2004) *AC 120-82 - Flight Operational Quality Assurance*. Available at: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-82.pdf (Accessed: 12 December 2017).

Fenton, N. and Neil, M. (2013) *Risk Assessment and Decision Analysis with Bayesian Networks*. 1st edn. CRC Press.

Flight Safety Foundation (2009) 'Reducing the Risk of Runway Excursions', *Report of the Runway Safety Initiative*, , p. 235.

Gorinevsky, D., Matthews, B. and Martin, R. (2012) 'Aircraft anomaly detection using performance models trained on fleet data', Chawla, N. V and Srivastava, A. N. (eds.)*2012 Conference on Intelligent Data Understanding*. IEEE, pp. 17–23.

Haverdings, H. and Chan, P.W. (2009) 'Quick Access Recorder (QAR) Data Analysis Software for Windshear and Turbulence Studies', *1st AIAA Atmospheric and Space Environments Conference*. Reston, Virigina: American Institute of Aeronautics and Astronautics, pp. 22–25.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', *Neural Computation*, 9(8), pp. 1735–1780.

Hubbard, D.W. (2009) *The Failure of Risk Management*. 1st edn. Hubbard, D. W. (ed.) Hoboken, NJ, USA: John Wiley & Sons, Inc.

International Civil Aviation Organisation (2010) *ICAO Annex 6 - Operation of Aircraft*. 9th edn. International Civil Aviation Organisation.

International Civil Aviation Organisation (2018) *Safety Management Manual (SMM)*. 4th edn. International Civil Aviation Organisation.

International Civil Aviation Organisation (2006) *Safety Management Manual (SMM)*. 1st edn. International Civil Aviation Organisation.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An Introduction to Statistical Learning*. New York, NY: Springer New York, Springer Texts in Statistics.

Jensen, F. V., Aldenryd, S.H. and Jensen, K.B. (1995) 'Sensitivity analysis in Bayesian networks', in *Symbolic and Quantitative Approaches*. , pp. 243–250.

Kjærulff, U. and van der Gaag, L. (2000) 'Making Sensitivity Analysis Computationally Efficient', Boutilier, C. and Goldszmidt, M. (eds.)*Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc, pp. 317–325.

Koiter, J.R. (2006) *Visualizing Inference in Bayesian Networks*., *MSc Thesis* Available at: http://www.kbs.twi.tudelft.nl/docs/MSc/2006/JRKoiter/thesis.pdf (Accessed: 6 February 2019).

Kuhn, M. and Johnson, K. (2013) *Applied Predictive Modeling*. New York, NY: Springer New York.

Li, L., Das, S., John Hansman, R., Palacios, R. and Srivastava, A.N. (2015) 'Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations', *Journal of Aerospace Information Systems*, 12(9), pp. 587–598.

Li, L., Gariel, M., Hansman, R.J. and Palacios, R. (2011) 'Anomaly detection in onboard-recorded flight data using cluster analysis', Burkle, M. and Herndon, A. (eds.)*2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. Seattle, Washington: IEEE, pp. 1–31.

Li, L., Hansman, R.J., Palacios, R. and Welsch, R. (2016) 'Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring', *Transportation Research Part C: Emerging Technologies*, 64(Supplement C), pp. 45–57.

Mendes, H. (2012) *Study of Mathematical Algorithms to Identify Abnormal Patterns in Aircraft Flight Data*. Available at: https://fenix.tecnico.ulisboa.pt/downloadFile/395144227066/ResumoAlargadoHM12Jun2012.pdf (Accessed: 13 December 2017).

Moretti, L., Di Mascio, P., Nichele, S. and Cokorilo, O. (2018) 'Runway veer-off accidents: Quantitative risk assessment and risk reduction measures', *Safety Science*, 104, pp. 157–163.

Mugtussidis, I.B. (2000) *Flight Data Processing Techniques to Identify Unusual Events*., *Doctoral Dissertation* Available at: http://hdl.handle.net/10919/28095 (Accessed: 29 April 2015).

Nanduri, A. and Sherry, L. (2016a) 'Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)', *2016 Integrated Communications Navigation and Surveillance (ICNS)*. Herndon, VA: IEEE, pp. 5C2-1-5C2-8.

Nanduri, A. and Sherry, L. (2016b) 'Generating Flight Operations Quality Assurance (foqa) data from the X-Plane Simulation', *2016 Integrated Communications Navigation and Surveillance (ICNS)*. Herndon, VA: IEEE, pp. 5C1-1-5C1-9.

Narasimhan, R. (2017) *weatherData: Get Weather Data from the Web*. Available at: https://cran.r-project.org/package=weatherData (Accessed: 1 August 2018).

Post, J.A. (2015) *Identification and analysis of veer- off risk factors in accidents / incidents*., *Future Sky Safety* Available at: https://www.futuresky-safety.eu/wp-content/uploads/2016/01/FSS_P3_NLR_D3.4_v2.0.pdf (Accessed: 14 November 2017).

Reason, J. (1997) *Managing the Risks of Organizational Accidents*. 1st edn. London: Ashgate.

Scutari, M. and Denis, J. (2014) *Bayesian networks - With examples in R*. 1st edn. Boca Raton, FL: CRC Press.

Smart, E. (2011) *Detecting Abnormalities in Aircraft Flight Data and Ranking their Impact on the Flight*. Available at: https://researchportal.port.ac.uk/portal/files/6061688/thesis.pdf (Accessed: 13 December 2017).

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. and Cowell, R.G. (1993) 'Bayesian Analysis in Expert Systems', *Statistical Science*, 8(3), pp. 219–247.

Spirtes, P. and Meek, C. (1995) 'Learning Bayesian Networks with Discrete Variables from Data', *Proceedings of First International Conference on Knowledge Discovery and Data Mining*. Montreal, Qu.

Streiner, D.L. and Cairney, J. (2007) 'What's under the ROC? An Introduction to Receiver Operating Characteristics Curves', *The Canadian Journal of Psychiatry*, 52(2), pp. 121–128.

Tonda, A., Spritzer, A. and Lutton, E. (2014) 'Balancing User Interaction and Control in BNSL', in Legrand, P., Corsini, M., Hao, J., Monmarche, N., Lutton, E. and Schoenauer, M. (eds.) *Lecture Notes*

in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, pp. 211–223.

Wang, L., Ren, Y. and Wu, C. (2018) 'Effects of flare operation on landing safety: A study based on ANOVA of real flight data', *Safety Science*, 102, pp. 14–25.

2020-05-23

Estimating runway veer-off risk using a Bayesian network with flight data

Barry, David J.

Elsevier

Barry DJ. (2021) Estimating runway veer-off risk using a Bayesian network with flight data. Transportation Research Part C: Emerging Technologies, Voume 128, July 2021, Article number 103180

https://doi.org/10.1016/j.trc.2021.103180