Quantification of Incident Probabilities Using Physical and Statistical Approaches

Chong Wang†, Ludwig Drees‡, Nadine Gissibl‡, Lukas Höhndorf†, Javensius Sembiring‡, Florian Holzapfel*  
†PhD student, Institute of Flight System Dynamics  
‡PhD student, Chair of Mathematical Statistics  
*Professor, Institute of Flight System Dynamics  
Technische Universität München  
Garching, Germany  
chong.wang@tum.de

Abstract — This paper provides an overview of methods capable of quantifying the occurrence probabilities of serious incidents for the flight operation of a single airline. Operational flight data is used for the analysis. Both a physical and a statistical approach are used to determine dependencies between contributing factors that can lead to an incident as well as to determine the occurrence probabilities of given incident types. The physical model, represented by the equations of motion of the aircraft, is applied to describe relationships that are well known. The hidden relationships and unknown contributing factors, in contrary, can be found and described using the mathematical concept of copulas. Extreme value theory can be used to directly estimate the occurrence probabilities from recorded flight data by making statements beyond existing data points. The runway overrun serves as an example to demonstrate our methods.

I. INTRODUCTION

Safety has always been of primary concern since the beginning of aviation. In recent years, the safety levels, if expressed as incident rates, have reached standards that are difficult to be determined within a single airline. In a classical approach, one would simply divide the number of incidents that occur during the operation by the total number of operations. Fortunately, this method fails when assessing the safety level of a single airline because serious incidents rarely happen or do not occur at all, some of them are even hard to find for the entire industry. The risks, however, are not zero.

We have applied several methods that are able to quantify the occurrence probabilities of incidents for a given fleet. Each of them takes into account specific aspects. The methods rely on data that are recorded during daily flight operations using the Quick Access Recorder (QAR).

Our research focuses on calculating occurrence probabilities of serious incidents, as defined by ICAO Annex 13, using operational flight data. For the assessment of incidents, we use safety critical measures. These measures indicate how serious the examined flight performed with respect to the considered incident. The other type of observed values are specific factors for which we want to investigate their impact onto the incident. In the example of the runway overrun, those factors could be the speed at touchdown, height above threshold, flaps setting etc. whereas the safety critical measure could be the actual landing distance. We call those factors, which have a relevant impact onto the incident, contributing factors. Within this paper, we will give an overview of three different approaches to quantify the likelihood of an incident within a single airline based on their flight operation data.

First, incident probabilities can be obtained directly from operational flight data using extreme value statistics. The method estimates the distribution of the safety critical measure from recorded flight data by calculating this specific measure directly. As incidents rarely occur, one would not be able to find any measurements that exceed certain values to result in serious incident. However, distributions can be fitted to existing data to obtain the behavior beyond the limits.

Second, we introduce a method to quantify incident probabilities using a physical approach to generate data that are beyond the boundaries of safe flight operation by simulation. A model based on the equation of motion of an aircraft is developed which is tailored to the relevant incident type. In order to obtain a correct model output, several parameters that are not recorded have to be estimated based on flight data. Now we propagate the probability distributions of contributing factors through the model to obtain the probability distribution of measures that directly describe whether or not the incident occurred, i.e. the safety critical measure. A simulation has to be run for obtaining the incident probabilities. The method of subset simulation is used to significantly reduce computing time. Using these tools, one is able not only to quantify the occurrence probability, but also to determine the significance of each contributing factor. When an airline decides to take measures to improve safety, this information is of great importance.

Third, as the physical incident models only take known contributing factors into account, a certain amount of cause and impact relationships are unknowingly neglected in the modeling. Therefore, we introduce the mathematical concept of copulas that is able to identify dependencies within flight data, especially those one would not expect. Copulas are able to describe dependencies more precisely than other measures such as simple correlation coefficients. In addition, they can also offer a description of nonlinear dependencies for a
multidimensional variable space, especially in the tails of a distributions, which are those of high relevance for our applications.

II. EXTREME VALUE THEORY TO ESTIMATE INCIDENT PROBABILITIES IN FLIGHT OPERATION

Serious incidents are obviously extreme events and are usually not predictable from observations. Extreme value theory provides an approach for estimating extreme events from operational flight data.

Let X denote any safety critical measure which is either recorded during flight operation or computed from recorded parameters. The cumulative distribution function of a \( \mathbb{R} \)-values random variable X is given by \( F(x) = \Pr(X \leq x) \) for \( x \in \mathbb{R} \). In words, it indicates the probability with which a value \( x \) will not be exceeded. In many cases, particularly high values of \( X \) correspond to incidents. It is thus especially important to estimate the probability of exceeding high values, i.e. \( \Pr(X > x) = 1 - F(x) = 1 - F(x) \) for large \( x \), precisely. The function \( F(x) \) is referred to as the tail of the cumulative distribution function \( F \).

An example of a safety critical measure for the incident of runway overrun is the actual landing distance (ALD). Figure 1 shows a Quantile-Quantile-plot (QQ-plot) of an ALD dataset (1442 measurements in meters) plotting the empirical quantiles versus the standard Gaussian quantiles. A QQ-plot is a graphical tool to check whether a dataset follows a specific distribution, here the standard Gaussian distribution. The points of the ALD dataset do not lie on a straight line. The Gaussian distribution is therefore not a convincing model for this dataset, especially for the far out right tail. Since high values of the ALD correspond to a runway overrun (values greater than the landing distance available (LDA)), it is especially important to estimate the tail of the distribution for high values precisely. The so-called Peaks Over Threshold (POT) method is used for this purpose. We only give a short overview of the idea of this method, details can be found for instance in ([1], Chapter 4) and ([2], Chapter 6.5).

The POT method uses the fact that exceedances over high thresholds for large samples follow a generalized Pareto distribution. This confirms that the Gaussian distribution is not appropriate for modeling far out tails (cf. figure 1). The method provides the following tail estimator, for given measurements \( X_1, \ldots, X_n \) (independently and identically distributed) and large \( u \):

\[
\hat{F}(u+y) = \frac{N_u}{n} \left( 1 + \frac{y}{\beta} \right)^{-\frac{1}{\xi}}, \quad y \geq 0, \tag{1}
\]

where \( N_u \) denotes the number of all measurements larger than \( u \). The parameters \( \xi \) and \( \beta \) can be estimated from the data which are greater than the threshold \( u \) using maximum likelihood (ML) estimation. For the ALD dataset, we choose the threshold \( u = 1940 \) m, with corresponding \( N_u = 66 \). The ML-estimators are then found to be

\[
\hat{\xi} = 0.012 \quad \text{and} \quad \hat{\beta} = 55.031. \tag{2}
\]

From this, the following estimate of the right tail for the ALD dataset is obtained:

\[
\hat{F}(1940+y) = \frac{66}{1442} \left( 1 + 0.012 \cdot \frac{y}{55.031} \right)^{-1.012}, \quad y \geq 0. \tag{3}
\]

This equation allows us to estimate the probability of the ALD to exceed 1940 m by \( y \). Table 1 shows estimated probabilities of exceeding different ALD values (1940 m + \( y \)), \( y \geq 0 \). The estimated tail depicted in figure 2 fits the ALD measurements perfectly, in particular also in the far end tail.

![Figure 1. QQ-plot of ALD data versus the standard Gaussian distribution](image)

Instead of estimating the probabilities of the ALD exceeding certain critical values, the inverse question is also of importance. Namely, given a small probability \( 1-q \), what is the corresponding critical ALD which corresponds to this probability? This value is referred to as the \( q \)-quantile. By inversion of equation (1), one obtains for a given \( q \in (0,1) \) an estimator \( \hat{x}_q \) of the \( q \)-quantile of the form

\[
\hat{x}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left( \left( \frac{n}{N_u} (1-q) \right)^{\frac{1}{\hat{\xi}}} - 1 \right). \tag{4}
\]

Figure 3 compares two estimates of the 99 %-quantile from the ALD measurements: the empirical ALD measurements exceeding the threshold \( u = 1940 \) m are plotted as a histogram. A first estimate is obtained by fitting a Gaussian distribution (cf. figure 1) to the measurements (estimating mean and variance based on all ALD measurements), and the second by using equation (4) (POT method). The two density estimates are depicted as dotted line corresponding to the Gaussian fit, and the solid line corresponds to the POT-fit. Obviously, the POT-fit is superior to the Gaussian fit. This entails the superiority of the quantile estimation of the POT method to the Gaussian quantile. Since risk is associated with the tails of the distributions, estimation of the 99 %-quantile by means of extreme value theory results in a much more realistic, though larger risk estimate than based on the Gaussian distribution. Table 2 compares more quantile estimates obtained by fitting a Gaussian distribution and by using the extreme value (POT) method, as shown in equation (4).
TABLE I. ESTIMATED PROBABILITIES OF EXCEEDING DIFFERENT ALD VALUES BASED ON THE ALD MEASUREMENTS USING THE POT METHOD (EQUATION (3))

<table>
<thead>
<tr>
<th>ALD</th>
<th>1950 m</th>
<th>2000 m</th>
<th>2050 m</th>
<th>2100 m</th>
<th>2150 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>3.8 %</td>
<td>1.5 %</td>
<td>0.6 %</td>
<td>0.3 %</td>
<td>0.1 %</td>
</tr>
</tbody>
</table>

Figure 2. Tail-fit of the ALD dataset using the POT method (equation (3)). The plotted data points are the largest 66 ALDs. The solid line indicates the estimated tail based on these data.

TABLE II. ESTIMATES OF DIFFERENT Q-QUANTILES BASED ON THE ALD MEASUREMENTS OBTAINED BY FITTING A GAUSSIAN DISTRIBUTION AND BY USING THE EXTREME VALUE (POT) METHOD (EQUATION (4))

<table>
<thead>
<tr>
<th>Quantile</th>
<th>96 %</th>
<th>97 %</th>
<th>98 %</th>
<th>99 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>1901 m</td>
<td>1923 m</td>
<td>1953 m</td>
<td>1999 m</td>
</tr>
<tr>
<td>POT</td>
<td>1947 m</td>
<td>1963 m</td>
<td>1985 m</td>
<td>2024 m</td>
</tr>
</tbody>
</table>

III. MODEL-BASED PREDICTION USING QAR DATA

Due to the high safety standards, certain incidents do not occur at all within the flight operation of a single airline. However, even if the incident itself cannot be observed, the contributing factors leading to incidents can be measured and quantified. We hypothesize, that if the distributions of the contributing factors are known as well as the functional relationships between these factors, we are able to make valid statements of the incident probability by overlapping the distributions of the contributing factors. This approach is shown in figure 4.

Our approach requires basically four steps. First, we develop incident models that describe the functional relationships, including physical, operational and technical relations. Second, we estimate the contributing factors that are neither recorded by the QAR nor can be determined algebraically. Third, we model the contributing factors using probability distributions. This includes extracting the relevant information (e.g. touchdown points for the runway overrun, depending on the incident type) from recorded flight data. Fourth, propagating the probability distributions through the incident models in order to quantify the occurrence probability of the incident that is evaluated. Due to the small probabilities, proper methods in terms of confidence of the result and computational have to be used.

A. Development of incident models

A model of the aircraft dynamics have to be developed using the existing knowledge of the aircraft behavior. It can also be modified according to the incident considered. They are represented by a system of nonlinear differential equations that describe the movement of the aircraft over time, depending on the forces and moments acting on the aircraft. Simplifications can be made to the model depending on the type of incident.

B. Estimation of unrecorded parameter

Currently analysis on QAR data is done through Flight Data Monitoring (FDM) programs that are commercially available for airlines. The analysis is done based on measured parameters and an analytical perspective. The methods implemented on these programs are reliable and used in many today’s airlines, as required by law.

However, there are some events for which the FDM software could not provide enough information about their cause of events. For example, an altitude drop event (figure 5) occurs during the approach phase. In this event the parameters which might be investigated are elevator position, wind speed, throttle and other related parameters. Another example is the runway overrun incident (figure 5) in which the contributing factors to be investigated are thrust, touchdown point, wind speed, touch down speed and other related parameters which are recorded in QAR device. Yet, all these recorded parameters might not provide enough information to the FDM crew hence the cause of the incident might not be revealed.

The method for extending the capability of the FDM program by means of estimating unmeasured or unrecorded parameters is presented. These estimates will provide more information to the analyst for investigating the cause of incidents. Some unrecorded parameters that can be estimated are lift, drag coefficient and their derivatives, runway friction coefficient, wind speed component as the vertical wind speed component is usually not recorded in QAR devices. Some of these parameters highly correlate with the safety of and airline operation, e.g. friction coefficient, and should therefore be monitored. The estimation method employed in this paper is Bayesian inference, which presents the parameters in...
distribution form. Section III.B.2 explains a short description of this method.

1) Problem formulation

As a case study in this paper, the increment of the lift coefficient due to flap deflection (FLP 0 to 4) during approach phase is investigated. Other parameters such as drag coefficient and thrust will also be presented but explanation will be emphasized more on lift coefficient parameters since this parameter is highly influenced by the flap deflection. In addition, this parameter plays an important role in altitude drop event. Figure 6 below depicts the flight phase and the related parameters to be estimated. The model used for the selected phase is shown in equation (5).

\[ a_x = \frac{1}{m} (qSC_D \sin \alpha - qSC_L \cos \alpha + \delta \tau) + b_x \] (5)

Equation (5) shows the acceleration along the longitudinal axis \( a_x \) and parameters which affect the acceleration on this axis. \( C_L \) and \( C_D \) are the lift and drag coefficients, respectively, while \( m \) is the mass of the aircraft. \( S \) represents the reference area. Additional parameter \( b_x \) is added into equation in order to deal with bias in the measurement. The parameters such as \( a_x \), true airspeed (TAS), which is taken into account by the total pressure \( q \) and the angle of attack \( \alpha \) are obtained directly from QAR data. The other parameters in equation (5), such as \( C_L \), \( C_D \), the thrust \( \delta \), and \( b_x \), are parameters to be estimated. The flight data used in this paper are taken from an Airbus A340.

2) Estimation method used

Many estimation methods are available and widely used in the context of aircraft system identification. The most common one is the ML method. This method is based on the maximization of the likelihood function and presents the estimate of a single value (point estimate) [3] [4]. However, the method used in this paper is Bayesian inference, which estimates the distribution of the parameter. There are several reasons for employing the Bayesian inference instead of Maximum Likelihood method on QAR data. In QAR data, there are more uncertainties than in flight testing data. These uncertainties come from several factors such as low and different sampling rates of the recorded parameters, the lack of dedicated flight maneuver to excite certain parameters, and flights are performed in a daily operational condition, thus disturbance, such turbulences, have to be taken into account. Due to these uncertainties, it is better to estimate the parameter in the distribution accounting for the uncertainties. Bayesian approach formulates the problem in the probability form and hence produces the result in probability form as well. The benefit of this formulation is that we can use even a complex model and estimate the probability distribution of all parameters involved in the model. The Bayesian approach works by computing the posterior of parameters using the likelihood and the prior of parameters as shown in equation (6) below [5]:

\[ p(\Theta|\text{data,model}) \propto p(\text{data|model}) p(\Theta|\text{model}) \] (6)

Observe that \( p \) stands for the density of the probability distribution \( P \) and we assume its existence. The \( p(\text{data|model}) \), \( p(\Theta|\text{model}) \), \( p(\text{data|model}) \) are called likelihood, prior and normalizing constant, consecutively. In this paper, the prior model used is an uninformative prior which is represented by the constant 1 and the likelihood model is represented by \( p(\text{data|model} \mid C_L, C_D, \delta, b_x) \). The measurement noise is assumed to be Gaussian distributed. The normalizing constant can be ignored since this parameter is canceled out during the sampling process. The final form of the posterior formulation is shown as follows:

\[ p(C_L, C_D, T, b_x | a_x\text{\_meas}) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{1}{2} \frac{(a_x\text{\_meas} - a_x\text{\_model})^2}{\sigma}} \times 1 \] (7)

The technique used for posterior sampling is Markov Chain Monte Carlo (MCMC) and the component wise Metropolis-Hastings algorithm is employed as a parameter updating method [6] [7]. To avoid a long burn-in time, the initial parameters are estimated using the least square method and the related covariance matrix is computed based on the estimated parameters.

3) Implementation and results

Implementation is done by sampling the posterior distribution using MCMC technique. The sampling number is
set to 500 000. The acceptance ratio using the Metropolis-Hastings algorithm for each parameter varies between 2-14 % and computation time took around 4-5 minutes for each flap position. The results for each parameter and flap position are shown in figure 7 to figure 9.

As shown in figure 7, the total lift coefficient increases during flap deflection from no flap deflection (FLP 0) until full flap (FLP 4). The shape of the parameter distribution is Gaussian and the expected value of parameter for each segment is denoted by a vertical dash line. By comparing the expected value of each segment, it is found that from FLP 0 to 1, the increment of lift coefficient increases about 12 %, FLP 1 to FLP 2 about 7%, FLP 2 to FLP 3 about 8% and FLP 3 to FLP 4 about 14%.

Drag coefficient distribution also changes due to flap deflection. However, the change of this parameter distribution is not the same as that of the lift coefficient. In addition, the distribution shape of the drag coefficient forms some modes as shown in figure 8. The expected value of parameter for each flap position is depicted by a vertical dash line. The increment of drag coefficient occurs from FLP 1 to FLP 4 by 3.4 % in average. Meanwhile, the decrement of this parameter occurs from FLP 0 to FLP 1 by 13%. The thrust is also estimated during the selected phase. However, this parameter is not affected by flap deflection but by throttle input as well as flight condition. As shown in figure 9, the estimated thrust varies in the range of 460 to 500 kN.

The Bayesian inference is successfully implemented on QAR data, which produces parameters in distribution form. The estimated parameters are those which are not recorded or unmeasured in QAR device. These estimates do not only provide the required parameters for the modeling, but also provide more information to the analyst. C. Modeling of contributing factors as probability distributions

The method for the prediction of incidents using flight data requires the estimation of the relevant contributing factors as probability distributions. Therefore, we identify a distribution type that fits the measured data of each contributing factor best. Many distribution types are available, e.g. Gaussian, Weibull or Gumbel. Due to the large amount and variety of data, a manual identification of a proper distribution type is insufficient. The goal of this section is to present available measures to describe the goodness of fit. Within this paper, we focus on identifying the best distribution type based on the computed measure.

\[
\text{AIC} = -2 \ln(L) + k \ln(n)
\]

A probability distribution can be seen as a model with unknown parameters \( \theta \), i.e. \( f(\theta) \). The parameters \( \theta \) of a probability distribution are estimated using the ML principle. Many measures are developed to describe the fitting of data to a model, for example, the negative Log-Likelihood \( L \), the Akaike information criteria (AIC) or the Bayesian information criteria (BIC). The AIC, as well as the BIC, basically use the negative Log-Likelihood \( L \) and add some penalty terms, taking the number of data points \( n \) and the number of free parameters \( k \) into account.

\[
L(\theta|X) = P(X|\theta) \quad (8)
\]

\[
\text{BIC} = -2 \ln(L) + k \ln(n) \quad (9)
\]

\[
\text{AIC} = 2k - 2 \ln(L) + \frac{2k(k+1)}{n-k-1} \quad (10)
\]

Due to the fact that most distribution types only require two or three parameters, the AIC or BIC does not necessarily return proper results, as seen in TABLE III. i.e. the selected distribution fits the data well. Therefore, we use different...
measures \([8]\). The first measure is the Integrated Quadratic Distance \(d_{Q}\) that is defined as follows:

\[
d_{Q}(F,G)=\int_{\Omega}[(F(t)-G(t))^{2}w(t)dt, \tag{11}\]

where \(G\) is the empirical distribution function and \(F\) the cumulative distribution function of the distribution candidate. In principle, a weighting function \(w(t)\) can be included that allows to emphasize certain areas of interest, e.g. the tails of a distribution. In order to counteract an overfitting due to the weighting function, we also evaluate a second measure that is called the Mean Value Divergence. This ensures that also the first moment (i.e. mean value) of the distribution candidates fit closely to the empirical mean value. Depending on the domain \(\Omega\) of each contributing factor, the distribution candidates are selected and the integration is adapted to the valid domain, e.g. only positive values \(\Omega=\{0;\infty\}\).

The following figures 10 and 11 show an example of fitting various distribution types to the measured headwind at touchdown. The fitting was performed without any weighting function. For this example, the Generalized Extreme Value distribution fits best.

Table III shows the result of the fitting of different distribution types. Generally, the smaller the measures are, the better the fit is. However, within this example, it is clear that using the classical BIC or AIC criteria, one would come to different results and more importantly rate the extreme value distribution better than the generalized extreme value distribution. Figure 10, in contrary, indicates that the generalized extreme value distribution is the best of those that were chosen, especially in the far left tail, which is of more interest for us, as seen in figure 11.

**D. Quantification of incident probabilities**

The distributions of the contributing factors shall now be propagated through the incident model. In our specific incident type of runway overrun, the contributing factors are, for example, the wind and the landing weight of the aircraft. The safety critical measure is the landing distance that is needed to achieve full stop. An overrun occurs when this distance is longer than the available runway length. The probability of any incident, which can be seen as a failure probability \(P_{\Omega}\), can generally be obtained by \([9]\):

\[
P_{\Omega}=P(\Theta|F)=\int I_{F}(\Theta)q(\Theta)d(\Theta), \tag{12}\]

**TABLE III. DIFFERENT FITTINGS FOR HEADWIND DATA**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Best Fit</th>
<th>(d_{Q})</th>
<th>MVD</th>
<th>BIC (10E3)</th>
<th>AIC (10E3)</th>
<th>Neglog (10E3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Generalized Extreme Value</td>
<td>0.0028</td>
<td>0.0002</td>
<td>3.6524</td>
<td>3.6436</td>
<td>1.8198</td>
</tr>
<tr>
<td>2</td>
<td>Logistic</td>
<td>0.0074</td>
<td>0.0652</td>
<td>3.6592</td>
<td>3.6460</td>
<td>1.8200</td>
</tr>
<tr>
<td>3</td>
<td>t-Location Scale</td>
<td>0.0077</td>
<td>0.0400</td>
<td>3.8686</td>
<td>3.8597</td>
<td>1.9279</td>
</tr>
<tr>
<td>4</td>
<td>Gaussian Distribution</td>
<td>0.0139</td>
<td>0</td>
<td>3.6605</td>
<td>3.6516</td>
<td>1.8238</td>
</tr>
<tr>
<td>5</td>
<td>Extreme Value Distribution</td>
<td>0.1212</td>
<td>0.3431</td>
<td>3.6372</td>
<td>3.6240</td>
<td>1.8090</td>
</tr>
</tbody>
</table>

where \(\Theta\) is a state of the system with a probability density function \(q\). \(I_{F}\) is an indicator function, it shows whether or not \(\Theta\) lies within the failure domain \(F\), being either one or zero.

Since \(P_{\Omega}\) is an integral with complex geometry and \(q\) is often not explicitly known, equation (12) cannot be solved directly. Instead, Monte Carlo Simulations (MCS) can be used to estimate the integral. However, MCS methods are not efficient when estimating small probabilities since the required number of samples is inverse proportional to the failure probability. In \([9]\), the method of subset simulation is presented to express the failure probability as a product of larger, conditional probabilities.

\[
P_{\Omega}=P(F_{i})\prod_{i=1}^{n}P(F_{i+1}|F_{i}) \tag{13}\]

Each \(P(F_{i})\) represents the probability of being with the failure domain of the respective subset \(i\). Those probabilities can be chosen to be large (usually of order \(10^{-1}\)), so that the number of samples can be small. The first subset is obtained by a classical Monte-Carlo simulation. A Metropolis-Hastings algorithm is used to obtain samples for the next subset from the samples of the previous subset \([10]\).

In \([11]\), a post-processor is presented as an enhancement to the subset simulation that provides not only an estimate for the failure probability, but also a beta distribution of the probability that describes the confidence of the estimation.

**E. Results**

The simulation was run with 10 000 samples per subset. The Airbus A340 is considered for generic runways with different length in this particular example. The results obtained from the subset simulation are shown in Table IV. The standard deviation of each estimated probability is also shown.

The overrun probability increases significantly with decreasing runway length. The estimations can be considered reliable since the standards deviations are all at least one order below their corresponding probabilities. However, those values only take into account methodic uncertainties. Model uncertainties still remain.
When the samples of each subset is examined, one is able to see them not only converging towards smaller stop margins as the simulation output. It is also possible to demonstrate the influence of the contributing factors. Figure 12 shows the headwind at touchdown for the 4000 meter long runway. For aesthetic reasons, only 500 samples of each subset are shown. The headwind is moving towards smaller (negative) values during the simulation. The samples that result in an overrun incident, which are those with stop margins, i.e. the difference between the actual landing distance and the available runway length, smaller than zero, are almost only those with negative headwinds, i.e. tailwind. Thus, the wind has great influence on the landing distance, which is consistent to the physical knowledge about aircraft.

The influence of the air temperature at the airport, however, is only slightly influential, as seen in figure 13. Samples of a large span of temperature values exist over all subsets. However, a minor tendency towards higher temperatures can be observed as the simulation progresses. This can be explained by lower air density, which result from higher temperature and therefore less effective speed brakes and engine reversers.

### TABLE IV. Obtained results for runway overrun

<table>
<thead>
<tr>
<th>Runway length</th>
<th>Probability</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500 m</td>
<td>3.2e-04</td>
<td>1.4e-05</td>
</tr>
<tr>
<td>3000 m</td>
<td>8.6e-05</td>
<td>4.1e-06</td>
</tr>
<tr>
<td>3500 m</td>
<td>4.2e-05</td>
<td>2.1e-06</td>
</tr>
<tr>
<td>4000 m</td>
<td>6.7e-06</td>
<td>3.6e-07</td>
</tr>
<tr>
<td>4500 m</td>
<td>1.8e-06</td>
<td>1.0e-07</td>
</tr>
</tbody>
</table>

In the next step, we are looking at the contributing factors of the incident in an automated way. Considering that an incident has various reasons and that different incident categories influence each other, it is necessary to be able to capture the dependence structure in high dimensions sufficiently. In statistics, there are various concepts to describe dependencies. Correlation coefficients are often used but they have disadvantages. A single value is not capable of describing the dependence between two random variables for the whole domain. Another aspect is that only linear dependence can be captured correctly, thus many relations cannot be represented. Furthermore, correlation coefficients are only defined for two random variables. Using correlation matrices, higher dimensions can be taken into account, but again, they are not capable of capturing the entire dependence structure satisfactorily since just pairwise evaluations are performed.

To overcome these disadvantages, we use the concept of copulas. They describe the dependencies more thoroughly throughout the entire parameter domain [13]. Since in many cases extremerealizations are contributing to an incident, a special focus lies on the boundaries of the domain. Furthermore, the theory of copulas is applicable in higher dimensions, i.e. description of the dependence between more than two random variables simultaneously is possible.
Based on flight data, a copula related to one or several incidents is estimated. Due to the lack of big data, we have applied the concept to a simple situation with data of one flight. We have analyzed the dependence structure between the altitude and the static pressure. In the given scenario, a two-dimensional Frank copula has been fitted to the data, as seen in figure 14.

We do not give a detailed interpretation, but it can be seen that there is a dominant negative dependence (if one parameter is small then the contour lines show that the other one is big, which corresponds to the real situation). Observe that the values of the two axis are not in usual range since the data points are transformed to a certain space such that the theory of copulas is valid.

The foundation of the theory of copula is the mathematical theorem of Sklar. It describes the interaction of the copula $C$, the joint distribution $F$ and the distributions of each component $F_i$.

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$ (14)

In addition, it describes the conditions of existence and uniqueness of the copula. For details, we refer to [14].

Once the copula is estimated, an efficient sampling from the joint distribution is possible and a topic of current research [15]. In the future, we will examine whether an application of this sampling method for the estimation of aircraft incident probabilities described in chapter III is feasible.

V. CONCLUSION

The assessment of risks and the estimation of incident probabilities are the highly interesting issues and challenging questions for any aircraft operator. We presented several methods that are able to answer these questions based on recorded QAR flight data. It is possible to obtain the probabilities by applying extreme value statistics to make valid statements beyond measured data. Alternatively, a physically motivated incident model to propagate the distributions of the contributing factors through the model to estimate the distributions of safety critical measures is developed. Since some contributing factors are not recorded, we apply estimation algorithms. The use of subset simulation to determine the incident probability can greatly reduce the computing time compared to a classical Monte Carlo simulation. Still, there might be unknown relationships and hidden contributing factors. Therefore, the concept of copula is presented to describe those information. The copula concept is of great potential. It can also be applied for use within the physical model to introduce dependencies between the contributing factors that are currently only covered by correlation coefficients.

REFERENCES