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 bei München, 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. Extreme value theory can be applied to directly estimate the occurrence probabilities from recorded flight data by making statements beyond existing data points. 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 of aircraft parameters that are well known. The hidden relationships and unknown contributing factors, in contrary, can be found and described using the mathematical concept of copulas. 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. The terms *accident* and *serious incident* are defined by the International Civil Aviation Organization [1]. Accident usually refers to events that lead to death of involving persons or a hull loss of involving aircraft. Serious incident refers to circumstances indicating that an accident nearly occurred. In recent years, the safety levels, if expressed as incident rates, have reached magnitudes that are too small 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 flights. This method fails when assessing the safety level of a single airline because such 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.

Our research focuses on calculating occurrence probabilities of serious incidents, as defined by ICAO Annex 13, Attachment C using operational flight data. The methods rely on data that are recorded during daily flight operations using the on-board Quick Access Recorder (QAR). For the assessment of incidents, we use safety critical measures (SCM), which indicate how critical the examined flight performed with respect to the considered incident. The other type of observed measurements are specific factors and we want to investigate their impact onto the incident. In the example of the runway overrun, those factors could be speed at touchdown, height above threshold, flaps setting etc. whereas the SCM could be the actual landing distance. We call those factors, which have a physical causal impact onto the incident, contributing factors. Some of those causal links can be easily described, such as the impact of the aircraft weight on the landing performance. Others, however, can only be expressed qualitatively, even though the relationship seems to be clear. An example could be the impact of pilot duty time on the precision of a manual approach.

Within this paper, we will give an overview of two different approaches to quantify the likelihood of an incident within a single airline based on flight operational data as well as a method to identify hidden relationships between factors that greatly influence the likelihood of an incident.

First, incident probabilities can be obtained directly from operational flight data using extreme value statistics. The method estimates the distribution of the SCM from recorded flight data by estimating this specific measure directly. As incidents rarely occur, it is likely to find no measurement that exceed certain high values to result in serious incidents. However, extreme value methods are designed to extrapolate a distribution function beyond the range of the data.

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. We propagate the probability distributions of contributing factors through this model to obtain the probability distribution of measures that directly describe, whether or not the incident occurred, i.e. the SCM. Incident probabilities are simulated. The method of subset simulation is used to significantly reduce computing time compared to classical Monte Carlo algorithms. With these tools it is not only possible to quantify the occurrence probability, but also to determine the significance of each contributing factor.

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 and are not easily described by physical relations. 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 which are of high relevance for our applications – to quantify incident probabilities.

II. Extreme Value Theory to Estimate Incident Probabilities in Flight Operation

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

We model any SCM, which is either recorded during flight operation or computed from recorded parameters, in a probabilistic way. Extreme value statistic is based on those data X, which are particularly critical; i.e. those exceeding a high threshold. It is thus especially important to find good models for $P(X>x)=1-F(x)=\bar{F}(x)$ for large x. The function $\bar{F}(x)$ is called the tail of the distribution function F.

An example of a safety critical measure for the incident 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, it is especially important to estimate the tail of the distribution for high values as precise as possible. 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 ([2], Chapter 4) and ([3], Chapter 6.5).

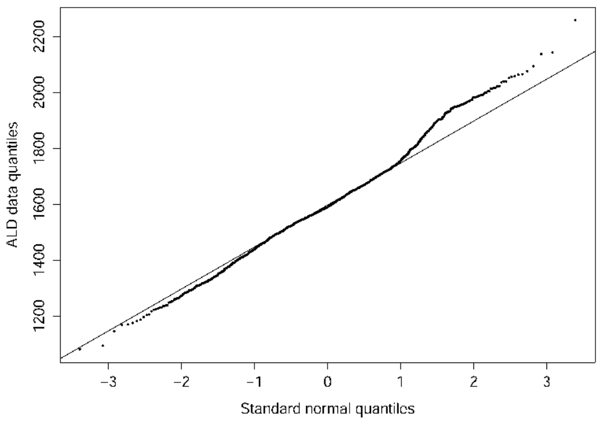

Figure 1. QQ-plot of ALD data versus the standard Gaussian distribution

TABLE I. Estimated probabilities of exceeding different ALD values based on the ALD measurements using the POT method (equation (3))

ALD	1950 m	2000 m	2050 m	2100 m	2150 m
Probability	3.8 %	1.5 %	0.6 %	0.3 %	0.1 %

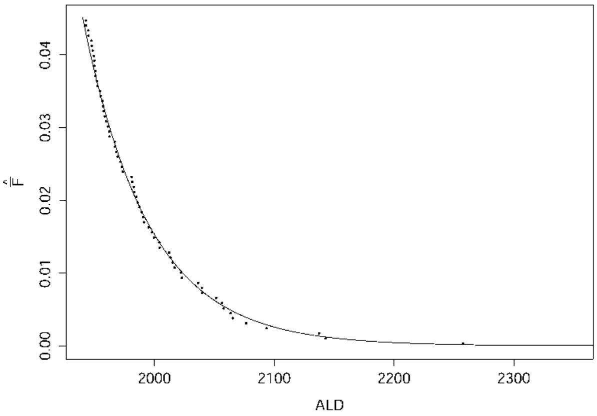

Figure 2. Tail-fit of the ALD dataset using the POT method (equation (3)). The plotted data points are the largest 66 ALDs. The solid line indicates the estimated tail based on these data.

The POT method uses the fact that exceedances over high thresholds for large samples follow a generalized Pareto distribution. In contrast to the Gaussian distribution it has a slowly decreasing tail $\bar{F}$ modelling large data appropriately. The method provides the following tail estimator, for given measurements $X_1, \ldots, X_n$ (independently and identically distributed) and large measurements u:

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

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

$$\hat{\xi}=0.012 \text{ and } \hat{\beta}=55.031. \tag{2}$$

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

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

This equation allows us to estimate the probability of the ALD to exceed 1940 m by y m. Table I shows estimated probabilities of exceeding different ALD values (1940 m + y m, y≥0). The estimated tail depicted in figure 2 fits the ALD measurements perfectly, in particular, in the far end tail. Although only few data are available in the far end tail, such a perfect fit can only be achieved by an extreme value model, which is tailor-made for such problems.

Instead of estimating the probabilities of the ALD exceeding certain critical values, the inverse problem is also of importance. Namely, given a small probability 1-q, what is the critical ALD which corresponds to this probability? This is the question an airline would ask if a certain safety target is set and actions have to be performed to achieve the target. This value is referred to as the q-quantile. By inversion of equation (1) for a given $q \in (0,1)$, we obtain an estimator $\hat{x}_q$ of the q-quantile of the form

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

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), 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 as it follows the data much more closely. This entails the superiority of the quantile estimation of the POT method to the Gaussian quantile. Since risk is associated with the tail of the distribution, extreme value statistics, which is based on extreme values of the SCM, results in a realistic tail and quantile estimates than any other distribution, in particular, the Gaussian distribution. Table II compares more quantile estimates obtained by fitting a Gaussian distribution and by using the extreme value (POT) method, as shown in equation (4).

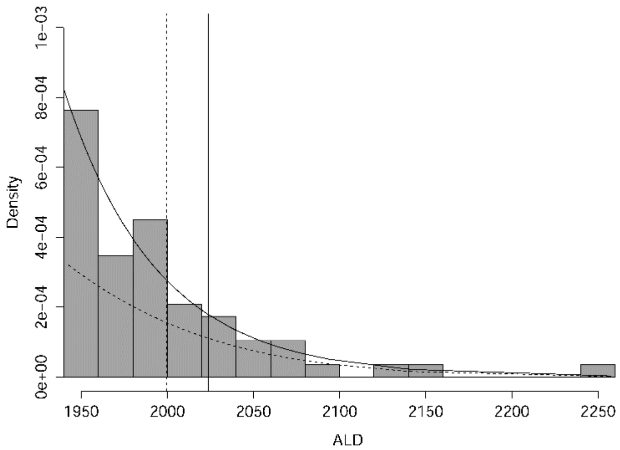

Figure 3. Estimates of the 99 %-quantile based on the ALD measurements obtained by fitting a Gaussian distribution (dotted line) and by using the extreme value (POT) method (solid line). Depicted is the histogram as well as the fitted densities for ALD values greater than 1940 m. The estimated quantiles are 1999 m for the Gaussian fit and 2024 m for the extreme value fit.

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))

Quantile	96 %	97 %	98 %	99 %
Gaussian	1901 m	1923 m	1953 m	1999 m
POT	1947 m	1963 m	1985 m	2024 m

III. MODEL-BASED PREDICTION USING QAR DATA

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 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 [4]. Third, we model the contributing factors using probability distributions. This includes extracting the relevant information (e.g. touchdown points for the runway overrun) from recorded flight data. Fourth, we propagate the probability distributions through the incident models in order to quantify the occurrence probability of the incident that is evaluated.

A. *Development of incident models*

A model of the aircraft dynamics is developed using the existing knowledge of the aircraft behavior. They are represented by a system of usually 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. For example, if only runway overruns are considered with veer-off incidents being neglected, the lateral motion of the aircraft does not have to be taken into account, reducing model complexity and thus computing time.

B. *Estimation of unrecorded parameters*

Currently analysis on QAR data is done through Flight Data Monitoring (FDM) programs that are commercially available.

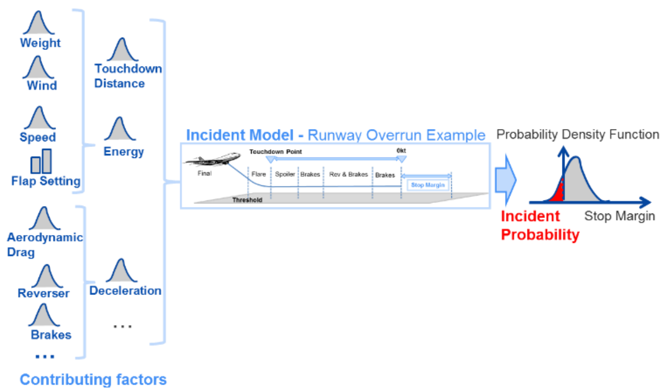

Figure 4. Propagation of probability distributions through a physical model

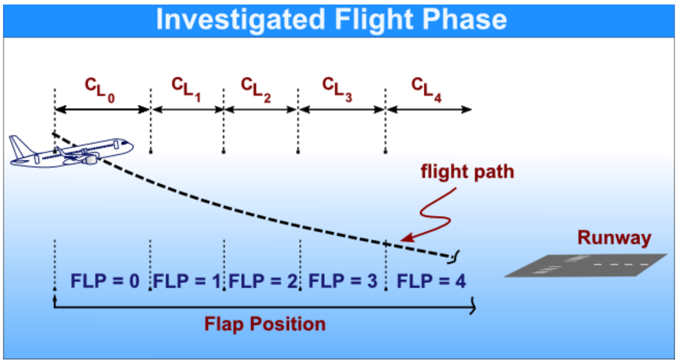

Figure 5. Investigated flight phase

However, there are some events for which the FDM software could not provide enough information about their cause of events such as an altitude drop event occurring during the approach phase. In this event the parameters which might be investigated are elevator position, wind speed, throttle, drag and lift coefficient as well as other related parameters. Another example is the runway overrun incident during landing 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 devices. Yet, all these recorded parameters might not provide enough information to discover the cause of the incident because the aerodynamic coefficients, for example, are not recorded.

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 and wind speed components as the vertical wind speed component is usually not recorded in QAR devices. Some of these parameters are highly influential on the safety of an 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. 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 and not recorded. In addition, this parameter plays an important role in altitude drop event. Figure 5 depicts the flight phase and the related parameters to be estimated. The model used for the selected phase is shown:

$$a_x = 1/m\,(qSC_L \sin\alpha - qSC_D \cos\alpha + \delta_T) + b_x. \qquad (5)$$

Equation (5) shows the acceleration along the longitudinal axis a_x [m/s^2] and the parameters which affect the acceleration on this axis. C_L [-] and C_D [-] are the lift and drag coefficients, respectively, while m [kg] is the mass of the aircraft. S [m^2] represents the wing reference area. An additional parameter b_x [m/s^2] is added into equation in order to deal with bias in the measurement. The parameters such as a_x, true airspeed (TAS) [m/s], which is taken into account by the total pressure q [N/m^2] and the angle of attack α [°] are obtained directly from QAR data. The other parameters in equation (5), such as C_L, C_D, the thrust δ_T [N], and b_x, are parameters to be estimated.

1) *Estimation method*

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) [5] [6]. However, the method used in this paper is Bayesian inference, which estimates the distribution of the parameter Θ. 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 regular 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 being performed in a daily operational condition. Therefore, disturbance, such as turbulences, have to be taken into account as well. 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, which is the probability distribution of the estimated parameters after measured data is taken into account, using the likelihood and the prior of parameters. The prior of the parameters, in contrary, is the probability distribution before obtaining relevant data or measurements. In this paper, we do not consider anything as the prior of parameters which is why it is represented by a uniform distribution of 1, as shown in equation (6) below [7]:

$$p(\Theta|\text{data,model}) = \frac{p(\text{data}|\Theta,\text{model})\cdot p(\Theta|\text{model})}{p(\text{data}|\text{model})} \qquad (6)$$

Observe that p stands for the density of the probability distribution of the parameter Θ. The p(data|Θ,model), p(Θ|model), p(data|model) are likelihood, prior and normalizing constant, consecutively. In this paper, the likelihood model is represented by $p(a_{x\_meas}\,|C_L,C_D,\delta_T,b_x)$. The measurement noise is assumed to be Gaussian distributed as we are referring to measurement noise only and not to other types of noise, e.g. process noise. Measurement devices typically produce noise that is Gaussian distributed. The normalizing constant can be ignored since this parameter is cancelled out during the sampling process. The final form of the posterior formulation is shown as follows:

$$p\left(C_L, C_D, T, b_x | a_{x\_meas}\right) = \frac{1}{\sqrt[N]{2\pi\sigma^2}} e^{\frac{1}{2}\frac{\sum_{i=1}^{N}\left(a_{x\_meas}-a_{x\_model}\right)^2}{\sigma^2}} \times 1. \qquad (7)$$

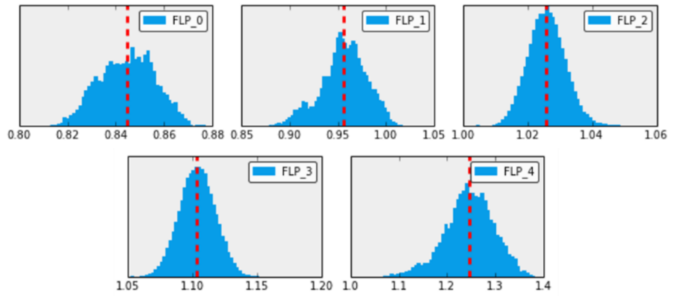

Figure 6. Lift coefficient distribution due to flap deflection

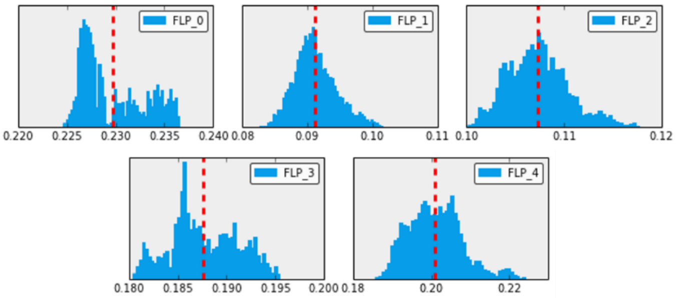

Figure 7. Drag coefficient distribution due to flap deflection

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 [8] [9]. 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.

2) Implementation and results

Implementation is done by sampling the posterior distribution using MCMC technique. 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. Due to the low acceptance ratio, the sampling number was chosen to be 500,000 to achieve a smooth parameter distribution. The results for C_L, C_D at each flap position are shown in figure 6 and 7.

As seen in figure 6, the total lift coefficient increases 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 the lift coefficient constitutes 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. The distribution shape of the drag coefficient forms some modes as shown in figure 7. 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. The estimated thrust varies in the range of 460 to 500 kN.

The Bayesian inference produces parameters in distribution form with all of them being in plausible ranges. The estimated parameters are those which are not recorded or unmeasured in QAR device.

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. 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 according to the measures.

A probability distribution can be seen as a model with unknown parameters $\boldsymbol{\theta}$, i.e. $f(\boldsymbol{\theta})$. The parameters $\boldsymbol{\theta}$ 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, as seen in equations (8), (9) and (10).

$$L(\theta|x) = P(x|\theta) \tag{8}$$

$$BIC = -2\ln(L) + k\cdot\ln(n) \tag{9}$$

$$AIC = 2k - 2\ln(L) + \frac{2k(k+1)}{n-k-1} \tag{10}$$

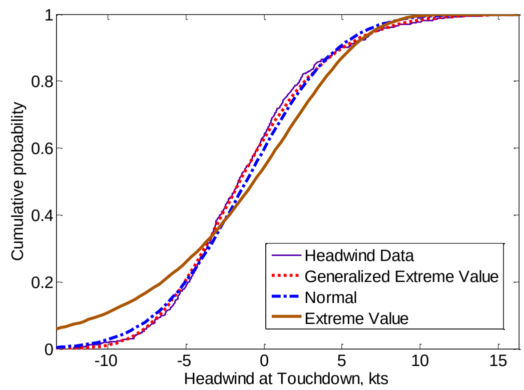

Figure 8. Different fittings for the headwind at touchdown shown as cumulative distribution function (CDF)

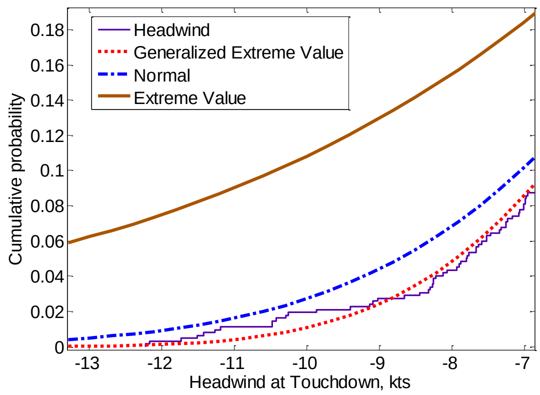

Figure 9. Different fittings for the headwind at touchdown – left tail

Due to the fact that most distribution types only require two or three parameters, the AIC or BIC does not necessarily return proper results, i.e. the selected distribution fits the data well, as seen in figure 9. Therefore, we use different measures [10]. The first measure is the Integrated Quadratic Distance d_{IQ} that is defined as follows:

$$d_{IQ}(F,G)=\int_{-\infty}^{\infty}\left(F(t)-G(t)\right)^2 w(t)\,dt, \qquad (11)$$

where G is the empirical distribution function and F the cumulative distribution function of the distribution candidate. 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 over-fitting due to the weighting function, we also evaluate a second measure that is called the Mean Value Divergence (MVD). 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 Ω 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 figures 8 and 9 show an example of fitting various distribution types to the measured headwind at touchdown. The fitting was performed without any weighting function. For this particular example, the Generalized Extreme Value distribution fits best.

Table III shows the result of the fitting of different distribution types. Generally, the smaller the values of 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 9, 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.

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 SCM 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_F, can generally be obtained by [11]:

$$P_F=P(\boldsymbol{\theta}\in F)=\int I_F(\boldsymbol{\theta})q(\boldsymbol{\theta})d(\boldsymbol{\theta}), \qquad (12)$$

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

TABLE III. DIFFERENT FITTINGS FOR HEADWIND DATA

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

Since P_F 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 [11], the method of subset simulation is presented to express the failure probability as a product of larger, conditional probabilities.

$$P_F=P(F_1)\prod_{i=1}^{m-1}P(F_{i+1}|F_i) \qquad (13)$$

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. The Metropolis-Hastings algorithm is used to obtain samples for the next subset from the samples of the previous subset [12].

In [13], 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.

The approach of propagating the distribution functions of contributing factors to quantify incident probabilities is also used in other domains. The UK railway's Precursor Indicator Model (PIM), for example, is quite similar to our approach. Yet, in our approach we use the aircraft dynamics described by the laws of motion to relate the contributing factors and to obtain their impact the incident probability. In comparison, the PIM uses weighting values that are determined by expert judgment [14]. In addition, the PIM model utilizes incidents that can actually be observed in daily operations whereas our model takes into account events that usually do not exist within the operation of one single airline due to its extremely small occurrence probability.

E. Results

The simulation was performed with 10 000 samples per subset. An 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 standard 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. It is also possible to demonstrate the influence of the contributing factors. Figure 10 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 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 behavior. In comparison, the influence of the air temperature at the airport, however, is only slightly influential. 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 a slightly lower air density, which results from a higher temperature and therefore less effective speed brakes and engine reversers.

The results, as shown in table IV, certainly do meet our expectations. It is obvious that the overrun probability decreases with increasing runway length since a longer distance is available to decelerate the aircraft. The general order of the figures, ranging between 3.2e-04 and 1.8e-06, seems to be consistent with the worldwide occurrence of runway overrun incidents. However, one has to emphasis that the input values are not real data collected from real flight operation, but artificial data that should approximately be similar to what one would obtain from real FDM systems.

IV. IDENTIFICATION OF HIDDEN RELATIONS

The accident of a Boeing 777 on January 17, 2008 at London Heathrow airport revealed high susceptibility of the Fuel Oil Heat Exchanger (FOHE) of the engine when exposed to a certain temperature and the fuel flow reaches a certain value although being compliant with certification regulations. The engine could therefore not provide the power that was required during final approach, the aircraft subsequently landed short of the runway. The accident report [15] indicates that this phenomenon can be found in the recorded data of other flights as well.

Those dependencies, which could be unknown but exist within the flight data, are the focus of this section. Especially if dependencies and unknown factors are relevant in terms of safety, airlines can benefit enormously from statistical methods. In the previous chapter of this paper, we implicitly assumed that we know all the main contributing factors of an incident. To overcome this disadvantage, we want to look deeper into the data and discover unknown phenomena, which could be found in the huge amount of recorded flight data, if existent. Since in many cases extreme realizations are contributing to an incident, our special focus lies on the boundaries of the domain.

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. In statistics, there are various concepts to describe dependencies. Correlation coefficients are often used but they have disadvantages. As a single number, it 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 nonlinear dependencies 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-linear dependencies are evaluated.

To overcome these disadvantages, we use the concept of copulas. They describe the dependencies more thoroughly throughout the entire parameter domain [16]. 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. To explain the statistical method, we have applied the copula concept to a simple situation with data of one flight. We have analyzed the dependence structure between the altitude and the static pressure. In general, algorithms to estimate multivariate distributions are well known and widely used for different purposes. In particular, statisticians develop such fitting algorithms also for the dependence structure, i.e. the copula. To date these algorithms are hardly used in engineering. In the given scenario, a two-dimensional Frank copula, as seen in figure 11, fits to the data best.

TABLE IV. OBTAINED RESULTS FOR RUNWAY OVERRUN

Runway length	Probability	Standard deviation
2500 m	3.2e-04	1.4e-05
3000 m	8.6e-05	4.1e-06
3500 m	4.2e-05	2.1e-06
4000 m	6.7e-06	3.6e-07
4500 m	1.8e-06	1.0e-07

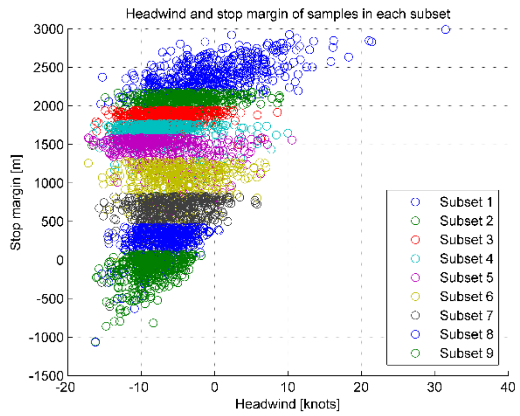

Figure 10. Samples of 9 subsets with the respective headwind components and stop margins

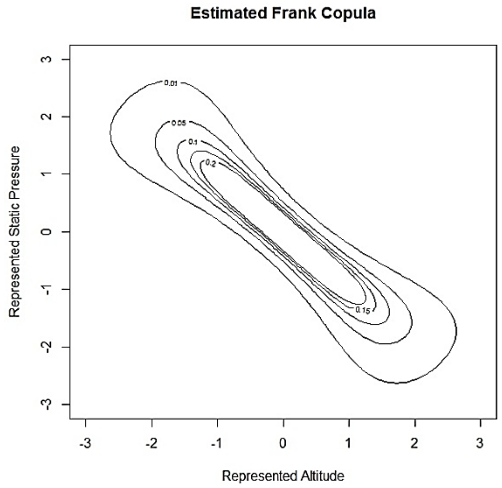

Figure 11. Contour plot of the estimated Frank copula

We do not give a detailed interpretation, but it can be seen from the bivariate density contour lines that there is a dominant negative dependence: if one parameter is small then the contour lines show that the other one is with high probability big, which is consistent with our knowledge about the earth atmosphere. Observe that the values of the two axis are not in physical range since the data points are transformed to a standardized 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:

$$F(x_1,\ldots,x_d)=C\big(F_1(x_1),\ldots,F_d(x_d)\big). \tag{14}$$

In addition, it describes the conditions of existence and uniqueness of the copula. For details, we refer to [17].

Once the copula is estimated, an efficient sampling from the joint distribution is possible and a topic of current research [18]. 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 highly interesting issues and challenging questions for any aircraft operator. We presented several methods that are able to obtain the relevant figures based on recorded QAR flight data. It is possible to receive probabilities of critical incidents 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 a copula is presented to describe those relationships, allowing the detection of unknown contributing factors. 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.

Equally promising is the concept of multivariate extreme value statistics which, as in the one-dimensional setting, focuses on large serious incidents. Extreme dependence measures capture dependence in extreme events, which may occur in observed data or simulated parameters.

REFERENCES

[1] International Civil Aviation Organization. *Annex 13 to the Convention of International Civil Aviation: Aircraft Accident and Incident Investigation*. Nineth Edition, July 2001.

[2] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.

[3] P. Embrechts, C. Klüppelberg, T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 2003.

[4] J. Sembiring, L. Drees, F. Holzapfel. *Extracting Unmeasured Parameters Based on Quick Access Recorder Data Using Parameter-Estimation Method*. AIAA Atmospheric Flight Mechanics Conference, Boston, MA, Aug. 19-22, 2013 AIAA-2013-4848.

[5] R. V. Jategaonkar. *Flight Vehicle System Identification – A Time Domain Methodology*. AIAA, New York, 2006.

[6] V. Klein and E. Morelli. *Aircraft System Identification: Theory and Practice, AIAA Progress in Astronautics and Aeronautics*. AIAA, New York, 2006.

[7] William M. Bolstad. *Introduction to Bayesian Statistics, Second Edition*. John Wiley & Sons, New Jersey, 2007.

[8] William M. Bolstad. *Understanding Computational Bayesian Statistics*. John Wiley & Sons, New Jersey, 2010.

[9] D. Gamermann and H. F. Lopes. *Markov Chain Monte Carlo*. Chapmann & Hall CRC, New York, 2006

[10] T. L. Thorarinsdottir, T. G. Gneiting, N. Gissibl. Using proper divergence functions to evaluate climate models. SIAM/ASA Journal on Uncertainty Quantification, 1(1): 522-534, 2013

[11] K.-S. Au, J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16: 263-277, 2001.

[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6): 1087-1092, 1953.

[13] K. M. Zuev, J. L. Beck, S.-K. Au, L. S. Katafygiotis. Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. *Computers and Structures*, 92-93: 283-296, 2012

[14] Rail Safety and Standards Board Ltd. *Annual Safety Performance Report 2012/13*. 2013.

[15] Air Accidents Investigation Branch. Report on the accident to Boeing 777-236ER, G-YMMM, at London Heathrow Airport on 17 January 2008, February 2010.

[16] C. Klüppelberg and C. Stelzer. *Dealing with Dependent Risks*. In: C. Klüppelberg, D. Straub and I. Welpe (Eds.). *Risk – A Multidisciplinary Introduction*. Springer, Heidelberg, 2014.

[17] D. Kurowicka and R. M. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons, New Jersey, 2006.

[18] D. Schmidl, C. Czado, S. Hug and F. Theis. A Vine-Copula Based Adaptive MCMC Sampler for Efficient Inference of Dynamical Systems. *Bayesian Analysis*, 8(1): 1-22, 2013.