INTEGRATION OF VINE COPULA DEPENDENCE STRUCTURES INTO SUBSET SIMULATION FOR ACCIDENT PROBABILITY QUANTIFICATIONS

Lukas Höhndorf*, Chong Wang*, Phillip Koppitz*, Sebastian Weiß*, Sergio Olmos**, Claudia Czado***, Florian Holzapfel*
*Institute of Flight System Dynamics, Technical University of Munich,
**Evelop Airlines, ***Chair of Mathematical Statistics, Technical University of Munich

Keywords: *quantification of aviation accident probabilities, safety assessment, subset simulation, high dimensional dependence structures, vine copula models*

Abstract

To estimate accident probabilities for commercial airlines, a sampling method called subset simulation is used at the Institute of Flight System Dynamics. Thereby, a physical model is utilized that considers a set of contributing factors and returns a so-called incident metric which describes the criticality of the particular flight with respect to a specific accident category. For accident probability estimations, multiple high-dimensional samples, i.e. virtual flights are generated and the physical model is applied for them. The goal of this paper is to integrate vine copula dependence models into the subset simulation implementation. These mathematical dependence models very flexibly describe high-dimensional and non-linear dependencies between the contributing factors. Eventually, the prevailing dependence structures of the contributing factors are recognized during the generation of the samples leading to more realistic results.

1 Introduction

In civil aviation, safety and methods to quantify and increase safety are central. Various target safety levels exist, one of them, set by the European Commission, is to have less than one accident[1] per ten million commercial aircraft flights by the year 2050, see page 17 of [2]. This corresponds to an accident probability of 10^{-7} per flight. Another example is the target safety level of Lufthansa formulated as an accident probability of less than 10^{-8} per flight [3]. For the Lufthansa operation, this corresponds to flying more than 100 years without an accident based on the flight numbers of 2009 [3].

Once a target safety level is defined, the crucial question is how to estimate the current safety level of an airline. Since the accident rates for a particular airline and also the entire industry are quite small, the straight forward statistical approach of dividing the number of accidents by the number of total flights is not feasible.

The estimation of the current safety level of an airline is one of the central motives of the Flight Safety working group at the Institute of Flight System Dynamics of the Technical University of Munich, see [4] and Section 9 of [5]. To achieve this, the methodology of subset simulation is used and applied to physical models. Certain parameters of these models are used as contributing factors whose distributions are drawn from available flight data. For the work described in this paper, an existing physical model for the aviation accident category *Runway Overrun* is

[1]The definition of “accident” of the International Civil Aviation Organization (ICAO) can be found in [1]

used. It has been mainly developed in [6] and further developed in [4].

The overall objective of this project is to increase the quality and confidence of the results from the subset simulation outcomes. With subset simulations, accident or failure probabilities of very small magnitude can be estimated efficiently. Within this paper, the dependence structures of the contributing factors are characterized by vine copula dependence models, see e.g. [7, 8]. For the generation of samples, these dependence structures are taken into account. First implementations of this idea have been conducted in the Bachelor's theses [9] and [10] that have been supervised by the lead author of this paper.

In the following chapter 2, the characteristics of the available data are described. Chapter 3 gives an overview of the existing *Runway Overrun* model that is one pillar of the research conducted in the given paper. A further basis is the subset simulation, which is summarized in chapter 4. The central part of this paper is the integration of vine copula structures into the subset simulation. A summary of the main theoretical aspects of vine copulas is given in chapter 5 and the proposed integration is described in chapter 6. Chapter 7 reflects the results of the subset simulation runs with integrated vine copula structures and compares them with the results before the integration. Finally, chapter 8 concludes the paper and gives an outlook of future tasks.

2 Recorded Flight Data Characteristics

A civil aircraft records data consecutively during the entire flight. The recorded time series describe the current state of the aircraft and multiple properties of various components. In Fig. 1, a recording of the barometric altitude is illustrated. The data is handled, stored, and analyzed as part of the so-called Flight Data Monitoring (FDM) activities of an airline, which is mandatory by law for commercial operators.

Based on the recorded time series, specific time points of the flight can be determined. The most important time point for the scope of this

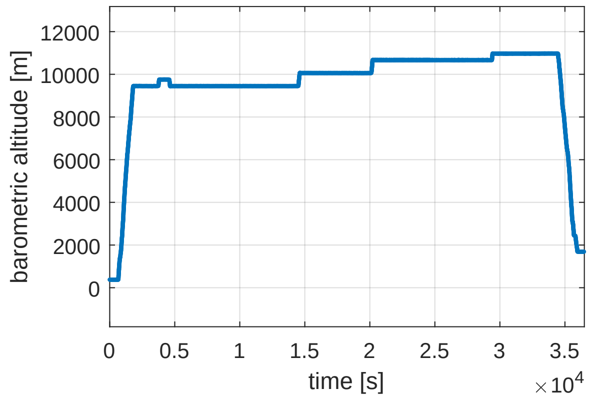

Fig. 1 Time series recording of barometric altitude

paper is the touchdown of the aircraft [11]. Taking the time series and the time points together, so-called measurements or snapshots can be calculated for all available flights. One example is the ground speed of the aircraft at touchdown. Measurements and especially their dependencies are a central aspect of this paper. In particular, the contributing factors of the *Runway Overrun* model that are described in the following chapter 3 are measurements.

3 Runway Overrun Model and Accident Probability Estimation

The research carried out in this paper is based on an implemented physical *Runway Overrun* model and a subset simulation framework.

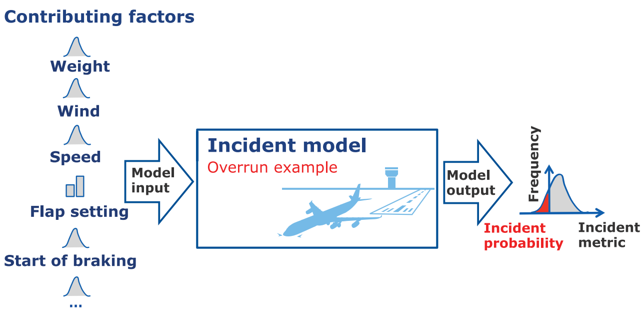

Fig. 2 Overview *Runway Overrun* model, source: Figure 3, Section 9 of [5]

On the left side of Fig. 2, the contributing factors are illustrated. The twelve contributing factors of the utilized *Runway Overrun* model are atmospheric temperature and pressure at touchdown, headwind at touchdown, aircraft landing mass at touchdown, duration from touchdown to spoiler deployment, duration from touchdown to start braking, duration from touchdown to reverser deployment, duration from touchdown to end of braking, distance from threshold to touchdown point, approach speed deviation in a specified time window, average reverse N1 during active reverse, and the commanded aircraft deceleration, see also [12]. Since the entire operation and not only single flights shall be investigated, the contributing factors are described by statistical distributions.

The utilized physical model describes the deceleration performance of the aircraft on the runway. The sum of all forces acting on the aircraft are split up into gravitation, aerodynamics, propulsion, and landing gear forces. Each of these components are described by further physical models. It is not within the scope of this paper to describe any detail of the *Runway Overrun* model. As an example, the equation for the braking force F_B is given.

$$F_B = \mu \cdot F_N = \mu \cdot (m \cdot g - L) = \mu \cdot \left[m \cdot g - \frac{1}{2} \cdot \rho \cdot V^2 \cdot S \cdot (C_{L,0} + C_{L,sp}(t_{sp}))\right] \quad (1)$$

In Equation 1, the following nomenclature is used. The friction coefficient is denoted by μ, the normal, i.e. vertical force acting on the landing gear F_N, aircraft mass m, gravity g, lift force L, atmospheric density ρ, aircraft true airspeed V, reference wing area S, zero lift coefficient $C_{L,0}$, spoiler lift coefficient $C_{L,sp}$, and the time of spoiler deployment t_{sp}. Any further details regarding the physical model can be found in [6, 4].

The outcome of the physical model is the stop margin, which equals the runway length minus the landing distance, see Fig. 3. In other words, it is the remaining runway after a (virtual) stop of the aircraft on the runway. The stop margin can be considered as an indicator for the *Runway Overrun* risk of a particular flight. The smaller the stop margin, the closer the flight is to a *Runway Overrun* accident. In case the stop margin is negative, a *Runway Overrun* accident occurred.

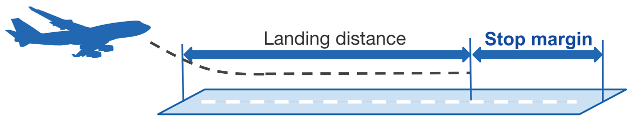

Fig. 3 Stop margin of the *Runway Overrun* model, source: Figure 2, Section 9 of [5]

The estimation of the *Runway Overrun* probability is conducted for different system and environment settings, so-called causal chains, see [6, 4]. One example for a causal chain is landing with full flaps and full slats extension together with a dry runway condition. Once the *Runway Overrun* probability is estimated for any causal chain, the probability for the *Runway Overrun* in general can be aggregated by also taking the probability of the causal chain itself into account. For any details the reader is referred to [6, 4].

4 Subset Simulation

Subset simulation is an advanced Monte Carlo method that is used for the generation of rare failure samples to enable the estimation of low occurrence probabilities. The main reference for subset simulation used within this paper is [13], further information can also be found in [14, 15, 16]. The generation of samples is controlled in such a way, that they are advancing towards the critical region CR[2]. This leads to a smaller number of required samples compared to the Monte Carlo algorithm. Since the physical model describing the aircraft motion is applied to the samples, the less samples lead to a significant run time reduction.

"Subset Simulation is based on the idea that a small failure probability can be expressed as the product of larger conditional probabilities of intermediate failure events, thereby potentially converting a rare event simulation problem into a sequence of more frequent ones. A general failure event is represented as $CR_b = \{Y > b\}$, where

[2]Observe that in [13] the symbol F is used for the critical region. To avoid synonyms, the symbol CR is used within this paper.

Y is a suitably defined 'driving response' characterizing failure.", see page 68 of [13]. This concept is also related to the incident metrics of [5].

The following problem setting is taken from [13] and the central idea of this paper can be formulated subsequently. Let $\boldsymbol{X} = (X_1, \ldots, X_{12})$ be a set of random variables characterizing the problem. Without loss of generality, $\boldsymbol{X}$ is assumed to be standard Gaussian (i.e. the means are zero and the covariance matrix is the identity matrix) and i.i.d. (independent and identically distributed). The standard Gaussian Probability Density Function (PDF) that is induced by $\boldsymbol{X}$ is denoted by ϕ. Dependent non-Gaussian random variables can be constructed from Gaussian ones by proper transformation. In this setting, the accident probability can be calculated as

$$\mathbb{P}(CR) = \int I(\boldsymbol{x} \in CR) \cdot \phi(\boldsymbol{x})\, d\boldsymbol{x}, \qquad (2)$$

see Equation (1) of [13].

For the utilized data set of the given *Runway Overrun* model, see chapter 3, the assumption of standard Gaussian variables are not fulfilled and, therefore, a proper transformation needs to be found, see chapter 5. More information regarding the i.i.d. property of recorded flight data can be found in [17].

An overview of the subset simulation is given in Algorithm 1. The first step of the subset simulation is a basic Monte Carlo step generating N samples. Subsequently, a value b is chosen such that the samples fulfilling the extended critical region $CR_b = \{Y > b\}$ can be selected as seeds for conditional sampling steps of the subsequent subsets [13]. The value b is chosen such that $p_0 \cdot N$ seeds are selected. A typical value for the level probability p_0 is 0.1, see e.g. page 70 of [13], i.e. 10 % of the samples are used as seeds for the next subset.

As a post processing step, [18] suggests a Bayesian consideration of the accident probability as a distribution. The results of Theorem 2 of [18] are used within this paper to calculate the accident probability, which is the mean of the fitted beta distribution. For any details, the reader is referred to page 293 of [18].

Step 1: Monte Carlo simulation to obtain
$\boldsymbol{X}^{(1),1}, \ldots, \boldsymbol{X}^{(1),N}$
Step 2: Calculate $Y\left(\boldsymbol{X}^{(1),1}\right), \ldots, Y\left(\boldsymbol{X}^{(1),N}\right)$
and define intermediate failure event CR_b
Step 3: Select $p_0 \cdot N$ seeds
$\boldsymbol{X}^{(1),seed_1}, \ldots, \boldsymbol{X}^{(1),seed_{p_0 \cdot N}}$ associated to CR_b
Step 4: Iterative conditional sampling
while *not sufficient flights in CR available* **do**
  Step 4.1: Conditional sampling of
  $\boldsymbol{X}^{(i+1),1}, \ldots, \boldsymbol{X}^{(i+1),N}$ according to
  Algorithm 2 or Algorithm 3
  Step 4.2: Calculate
  $Y\left(\boldsymbol{X}^{(i+1),1}\right), \ldots, Y\left(\boldsymbol{X}^{(i+1),N}\right)$ and define
  intermediate failure event CR_b
  Step 4.3: Select $p_0 \cdot N$ seeds
  $\boldsymbol{X}^{(i+1),seed_1}, \ldots, \boldsymbol{X}^{(i+1),seed_{p_0 \cdot N}}$ associated
  to CR_b
end
Step 5: Calculate accident probability based on
all subset samples

Algorithm 1: Subset simulation

In the remaining of this chapter, the two algorithms presented in [13] used for the generation of N conditional samples are summarized.

Both algorithms for the conditional sampling assume that a seed sample $\boldsymbol{X}^{(i)} = (X_1^{(i)}, \ldots, X_{12}^{(i)})$ is given which is distributed according to the target conditional distribution, i.e.

$$\phi(\boldsymbol{x}|CR_b) = \mathbb{P}(CR_b)^{-1} \cdot I(\boldsymbol{x} \in CR_b) \cdot \phi(\boldsymbol{x}) \qquad (3)$$

see Equation (3) of [13]. Both algorithms are designed to generate the next sample $\boldsymbol{X}^{(i+1)} = (X_1^{(i+1)}, \ldots, X_{12}^{(i+1)})$ which is again distributed as $\phi(\boldsymbol{x}|CR_b)$.

4.1 Metropolis Algorithm

The first algorithm for the conditional sampling step is the independent-component Markov Chain Monte Carlo (MCMC) algorithm [19] with symmetric proposal distribution given by its density $p_i^*(\cdot;\cdot)$. The algorithm is commonly referred to as Metropolis algorithm and is summarized in Algorithm 2, see [13].

Step 1: Generate $\boldsymbol{X}' = (X_1', \ldots, X_{12}')$
for $j = 1, \ldots, 12$ **do**
·) Generate ξ_j from the proposal distribution given by its density $p_i^*(\cdot; X_j^{(i)})$ and U_j uniformly on $[0,1]$
·) Calculate $r_j = \frac{\phi(\xi_j)}{\phi(X_j^{(i)})}$
·) **if** $U_j \leq r_j$ **then**
$X_j' = \xi_j$
else
$X_j' = X_j^{(i)}$
end
end
Step 2: Check failure
if $\boldsymbol{X}' \in CR_b$ **then**
$\boldsymbol{X}^{(i+1)} = \boldsymbol{X}'$ (Accept)
else
$\boldsymbol{X}^{(i+1)} = \boldsymbol{X}^{(i)}$ (Reject)
end

Algorithm 2: Independent-component MCMC

4.2 Limiting Algorithm

For the Metropolis Algorithm 2, a proposal distribution $p_i^*(\cdot;\cdot)$ needs to be chosen. It reveals that the efficiency of the subset simulation is insensitive of the type of the proposal distribution [15, 14]. This was the starting point at page 68 of [13] to develop a new algorithm for which no choice of proposal distribution is necessary anymore.

It can be proven that the so-called limiting algorithm, here summarized in Algorithm 3, delivers equivalent results more efficiently. For any details, the reader is referred to [13].

Furthermore, information and suggestions for the choice of a_k and s_k^2 are given in [13] and followed by the given paper.

5 Vine Copula Dependence Models

Generating correlated samples of arbitrary high-dimensional distributions is relevant for various scientific disciplines and industry applications and has been an important focus of research for many years, e.g. [20].

The method considered in this paper are vine

Step 1: Generate $\boldsymbol{X}' = (X_1', \ldots, X_{12}')$ as a Gaussian vector with independent components, with mean vector $(a_1 \cdot X_1^{(i)}, \ldots, a_{12} \cdot X_{12}^{(i)})$ and variances $(s_1^2, \ldots, s_{12}^2)$
Step 2: Check failure
if $\boldsymbol{X}' \in CR_b$ **then**
$\boldsymbol{X}^{(i+1)} = \boldsymbol{X}'$ (Accept)
else
$\boldsymbol{X}^{(i+1)} = \boldsymbol{X}^{(i)}$ (Reject)
end

Algorithm 3: Limiting algorithm

copula dependence models. The utilized algorithms were developed as an R package in [21]. In the following, a brief introduction to copulas and vine copula models is given. The theory of copula is a vivid area of current research. There are various relevant publications and the reader is referred to [7, 8] and references therein.

According to the theorem of Sklar [22], a d-dimensional distribution function F can be decomposed into its marginal Cumulative Distribution Functions (CDF) $F_1, \ldots, F_d$ and a distribution C called the copula in the following way

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \quad (4)$$

The idea of vine copula models is to decompose a high dimensional copula into several lower dimensional ones. The concept of conditional distribution provides the mathematical framework for this decomposition where a d-dimensional copula is split up into $\frac{d \cdot (d-1)}{2}$ two-dimensional copulas. Since this decomposition is not unique, a suitable construction needs to be chosen [23]. In addition, the copula families and all the associated copula parameters need to be fitted and estimated.

In this paper, in the mathematical theory of copula, and in the theory of subset simulation, it is very important to differentiate three variable domains. The on-board recording of the FDM data and the calculation of measurements are performed in the so-called X space. Here, the variables have their specific units, e.g. knots for speeds. The CDF of a distribution in the X space is denoted by F. Using the Probability Integral Transformation (PIT), see page 20 of [7], it follows that applying F to the data in the X space leads to uniformly distributed data on $[0,1]$. This motivates to call this space the U space and it plays an essential role in the copula theory. Furthermore, the one-dimensional CDF of the standard normal distribution is denoted by Φ. Applying Φ^{-1} to a distribution in the U space transforms this distribution into the so-called Z space. This was the space considered for the subset simulation in chapter 4. A summary of this chain of domains and the associated transformations for $k = 1, \ldots, d$ is given by

$$X_k \underset{F_k^{-1}}{\overset{F_k}{\rightleftarrows}} U_k \underset{\Phi}{\overset{\Phi^{-1}}{\rightleftarrows}} Z_k. \tag{5}$$

Observe that in the setting of this paper, the inverse of F_k is assumed to exist.

6 Integration of Vine Copula Structures into Subset Simulation

The flight data considered in this paper is stored, handled, and analyzed within the IT environment of the Flight Safety working group at the Institute of Flight System Dynamics. The calculation of measurements, the fitting of marginal distributions, the application of the physical model, see chapter 3, and the subset simulation, see chapter 4, are conducted within that environment.

In addition, the R software package *VineCopula* is utilized [21]. For the fitting of the vine copula model, the function *RVineStructureSelect* is used. Thereby, the vine copula model is fitted to the calculated measurements and, therefore, representing the true dependence structures of them. According to chapter 4, two transformations are required to use the subset simulation suggested in [13]. The first required transformation is to remove the dependence structure of a given seed. Subsequently, the proposed sample generation of either Algorithm 2 or Algorithm 3 is conducted. After that, the dependence structure is again integrated using the second transformation. Finally, the physical model can be run on the generated samples taking the dependence structures into account and the seeds for the next subset chosen. An overview of the suggested procedure is given in Fig. 4.

The first transformation to remove the dependence information from the sample is given by the Rosenblatt transformation [24]. According to page 3 of [25], the Rosenblatt transformation, that is hardly known for practical applications, is a generalization of the well known Nataf transformation [26]. Precisely, when the considered dependence structure is Gaussian, the Rosenblatt transformation is equivalent to the Nataf transform, see [27] and page 3 of [25]. For the vine copula models, the Rosenblatt transformation has been introduced in [28] and the algorithm is provided in [21] by the function *RVinePIT*.

The second transformation is also given in [21] by the function *RVineSim*. The main purpose of this function is to generate samples in the U space that are taking the dependence structure into account which is given by the vine copula model. However, the function was designed with an optional argument. This can be an existing high-dimensional sample with independent components. In this case, *RVineSim* integrates the prevailing dependence structure and exactly this is required from the second transformation [21].

7 Investigation of Runway Overrun Probabilities with the Vine Copula Integration

In the following, probabilities for the accident category *Runway Overrun* are calculated. The critical region *CR* of the subset simulation, see chapter 4, for the *Runway Overrun* accident category is given by a stop margin less than 0, see chapter 3. For any of the following results, the numbers of samples per subset N was chosen to be $N = 10.000$ and the level probability p_0 was set to be $p_0 = 0.1$.

In Fig. 5 and Fig. 6 the evolution of the measurement *commanded deceleration of the aircraft* is given as an example. In both cases, the Limiting Algorithm 3 was used and the causal chain *full flaps* and *full slats* configuration with a *wet runway* condition is illustrated. Fig. 5 illustrates the situation without the integration of the vine copula dependence structures and Fig. 6 describes the subsets that were generated with the utilized vine copula dependence models. In both figures, the commanded deceleration measurement is indicated on the horizontal axis. On the vertical axis, the stop margin is illustrated. Due to this selection, the different subsets can be well identified. In both cases, a colour coding is used to highlight the different subsets. The values of b in the second to last subset (which can be roughly identified in Fig. 5 and Fig. 6) is, in the case before the vine copula integration, 123 meters and, in the case after the integration of vine copula structures, 88 meters.

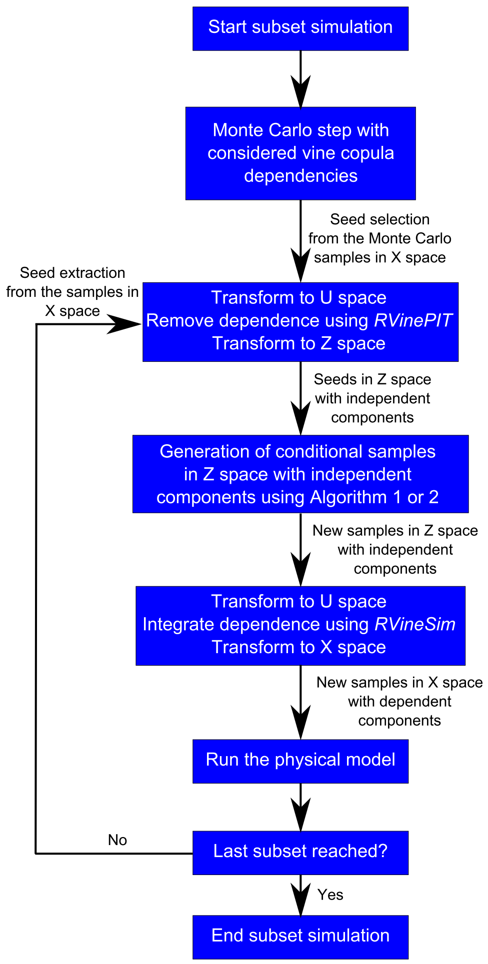

Fig. 4 Overview of the proposed vine copula integration into subset simulation

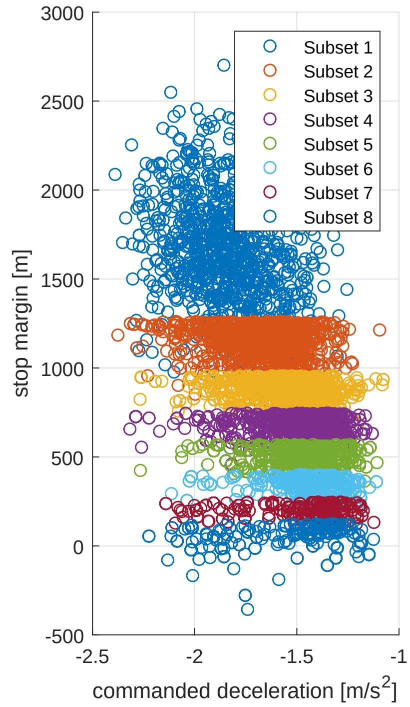

Fig. 5 Evolution of the measurement commanded deceleration - **excluding** vine copula models

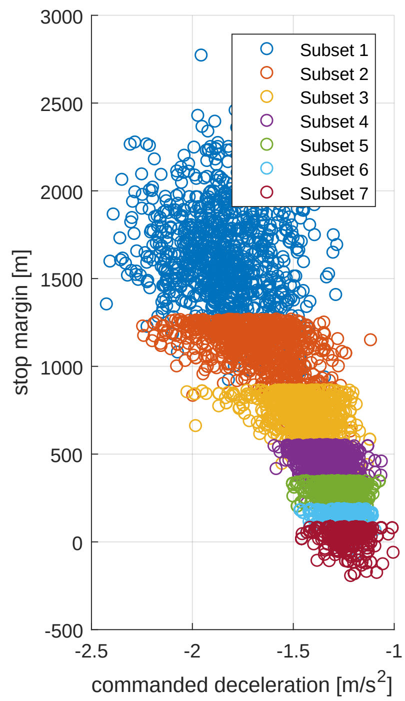

Fig. 6 Evolution of the measurement commanded deceleration - **including** vine copula models

Considering these two types of plots for other contributing factor partially shows different behaviors. Sometimes, the differences before and after the vine copula integration are smaller. These conditions are strongly influenced by the underlying dependence structures captured in the vine copula model.

Fig. 6 shows one subset less compared to Fig. 5 to obtain samples in the critical region *CR*, i.e. samples with negative stop margin. This fact also leads to higher accident probabilities that will be discussed in the following. In addition, a different dependence behavior can be observed particularly in the lower subsets (since the marginal distributions are the same for both figures, the only difference is the dependence structure, compare theorem of Sklar in equation (4)).

Fig. 7 gives an overview of the *Runway Overrun* probabilities for the causal chain flaps setting 25 (second highest setting), full slats setting and a dry runway condition. As sample generation is (pseudo-) random, the results of two individual runs for the same data and the same causal chain are in general different. The estimated *Runway Overrun* probabilities for this causal chain are in the range of 10^{-6}. Fig. 7 shows the results of eight simulations of the same settings. For the cases with Metropolis sampling, a normal distribution with mean 0 and standard deviation 0.7 is used. Some abbreviations are necessary, "Met" stands for the Metropolis Algorithm 2, "Lim" for the Limiting Algorithm 3, "S" for single, i.e. with the contributing factors considered as independent from each other, and finally "C" for the results after the integration of vine copula structures.

In Fig. 7, it can be observed that the integration of the vine copula dependence structures influences the accident probability estimations and both, the accident probabilities and their standard deviation is higher with the vine copula integration. Since the true value of the accident probability is unknown, it can not be clarified whether the results were improved or impaired.

One reason for this probability increase could be the fact that in aviation, accidents can typically be considered as chain of events, see. e.g. [29]. The underlying dependencies between the considered measurements are described by the vine copula models more sophisticatedly. This means that if one measurement is already extraordinary, there is a good chance that also another measurement is particularly high or low. This situation is represented in the vine copula and might contribute to these higher accident probabilities and therefore, the estimation after the vine copula integration can be considered as more realistic.

Furthermore, slightly higher accident probabilities that are caused by the vine copula integration into subset simulation lead to an overestimation of the accident probabilities compared to the results without the integration and therefore towards the safe side.

For now, no explanation for the higher standard deviation could be identified, however, it might be related to a similar situation described on page 53 of [9].

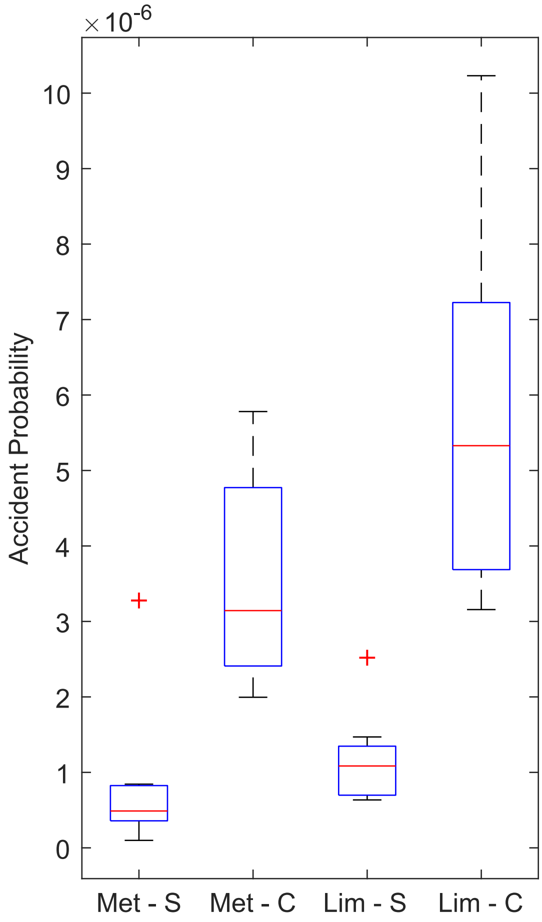

Fig. 7 Subset simulation results

As stated in [13], the efficiency of the Limiting Algorithm 3 is higher than for the Metropolis Algorithm 2 and should therefore be preferred. For the calculations conducted within the scope of this paper, this property resulted in a slightly shorter runtime of Algorithm 3 compared to Algorithm 2.

8 Conclusions and Outlook

Within this project, the proposed technique to integrate vine copula dependence models was applied to a subset simulation algorithm for the estimation of *Runway Overrun* accident probabilities of an airline. Since the underlying flight data measurements show non-trivial dependencies, the estimations for the accident probabilities are affected by the proposed method.

Also for future problems with significant dependence structures between the contributing factors, the integration of vine copula structures into subset simulation can be very beneficial.

References

[1] International Civil Aviation Organization (ICAO). *Aircraft accident and incident investigation: Annex 13 to the Convention on International Civil Aviation*. 11th edition, International Civil Aviation Organization, 2016.

[2] European Commission. *Flightpath 2050: Europe's vision for aviation*. Office for Official Publications of the European Communities, 2011.

[3] Raps J. *Sicherheit muss jeden Tag neu produziert werden: Risk Management bei Lufthansa*. 24.09.2009.

[4] Drees L. *Predictive Analysis: Quantifying Operational Airline Risks*. Dr. Hut, 2017.

[5] International Air Transport Association (IATA). *Safety report 2013*. 50th edition, International Air Transport Association (IATA), 2014.

[6] Wang C. *Quantifizierung von Unfallwahrscheinlichkeiten für Runway Overrun über Subset-Simulation*. Master's Thesis, Technical University of Munich, 2013.

[7] Kurowicka D and Joe H. *Dependence modeling: Vine copula handbook*. World Scientific Pub. Co, 2011.

[8] Joe H. *Dependence modelling with copulas*. Monographs on statistics and applied probability 134, CRC Press, 2015.

[9] Weiß S. *Integration von Abhängigkeitsstrukturen in ein Runway Overrun Incident Model mit Hilfe von Copulas in Matlab und R*. Bachelor's Thesis, Technical University of Munich, 2015.

[10] Olmos S. *Incorporate Dependence Structures in the Runway Overrun Model using MATLAB Copula Models*. Bachelor's Thesis, Technical University of Munich, 2015.

[11] Koppitz P, Siegel J, Romanow N, Höhndorf L and Holzapfel F. Touchdown Point Detection for Operational Flight Data Using Quality Measures and a Model Based Approach. *2018 AIAA Atmospheric Flight Mechanics Conference*, AIAA SciTech Forum, Kissimmee, Florida, USA, 2018.

[12] Wang C, Drees L and Holzapfel F. Incident Prediction using Subset Simulation. *29th Congress of the International Council of the Aeronautical Sciences (ICAS)*, St. Petersburg, Russia, 0758, 2014.

[13] Au S and Patelli E. *Rare event simulation in finite-infinite dimensional space*, Vol. 148, pp 67-77, 2016.

[14] Au S and Wang Y. *Engineering Risk Assessment with Subset Simulation*. 1st edition, Wiley, 2014.

[15] Au S and Beck J L. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, Vol. 16, No. 4, pp 263-277, 2001.

[16] Papaioannou I, Betz W, Zwirglmaier K and Straub D. MCMC algorithms for Subset Simulation. *Probabilistic Engineering Mechanics*, Vol. 41, pp 89-103, 2015.

[17] Höhndorf L, Czado C, Bian H, Kneer J and Holzapfel F. Statistical Modeling of Dependence Structures of Operational Flight Data Measurements not Fulfilling the I.I.D. Condition. *2017 AIAA Atmospheric Flight Mechanics Conference*, AIAA AVIATION Forum, Denver, Colorado, USA, 2017.

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

[19] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, Vol. 21, No. 6, pp 1087-1092, 1953.

[20] Vale C and Maurelli V A. Simulating multivariate nonnormal distributions. *Psychometrika*, Vol. 48, No. 3, pp 465-471, 1983.

[21] Schepsmeier U, Stoeber J, Brechmann E C, Graeler B, Nagler T and Erhardt T. *VineCopula: Statistical Inference of Vine Copulas*. R Package, https://github.com/tnagler/VineCopula, 2018.

[22] Sklar A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, Vol. 8, pp 229-231, 1959.

[23] Dißmann J, Brechmann E C, Czado C, Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, Vol. 59, Supplement C, pp 52-69, 2013.

[24] Rosenblatt M. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, Vol. 23, No. 3, pp 470-472, 1952.

[25] Torre E, Marelli S, Embrechts P, Sudret B. A general framework for uncertainty quantification under non-Gaussian input dependencies. *arXiv.org*, http://arxiv.org/pdf/1709.08626v1.

[26] Nataf A. Détermination des distributions de probabilités dont les marges sont données. *Comptes Rendus de l'Académie des Sciences*, Vol. 225, pp 42-43, 1962.

[27] Lebrun R, Dutfoy, A. Do Rosenblatt and Nataf isoprobabilistic transformations really differ?. *Probabilistic Engineering Mechanics*, Vol. 24, No. 4, pp 577-584, 2009.

[28] Schepsmeier U. Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review. *Journal of Multivariate Analysis*, Vol. 138, pp 34-52, 2015.

[29] Weiss K A, Leveson N, Lundqvist K, Farid N and Stringfellow M. An analysis of causation in aerospace accidents. *20th DASC. 20th Digital Avionics Systems Conference*, Daytona Beach, FL, USA, 7184164, pp 1-11, 2001.

9 Contact Author Email Address

The contact author email address is lukas.hoehndorf@tum.de

Acknowledgements

The research carried out in this paper was supported by the Deutsche Forschungsgemeinschaft (DFG) within the project "Copula based dependence analysis of functional data for validation and calibration of dynamic aircraft models" with the project identifiers CZ 86/5-1 and HO 4190/10-1.

Copyright Statement