An Application of D-vine Regression for the Identification of Risky Flights in Runway Overrun

Hassan Alnasser
Technical University of Munich
Munich, Germany

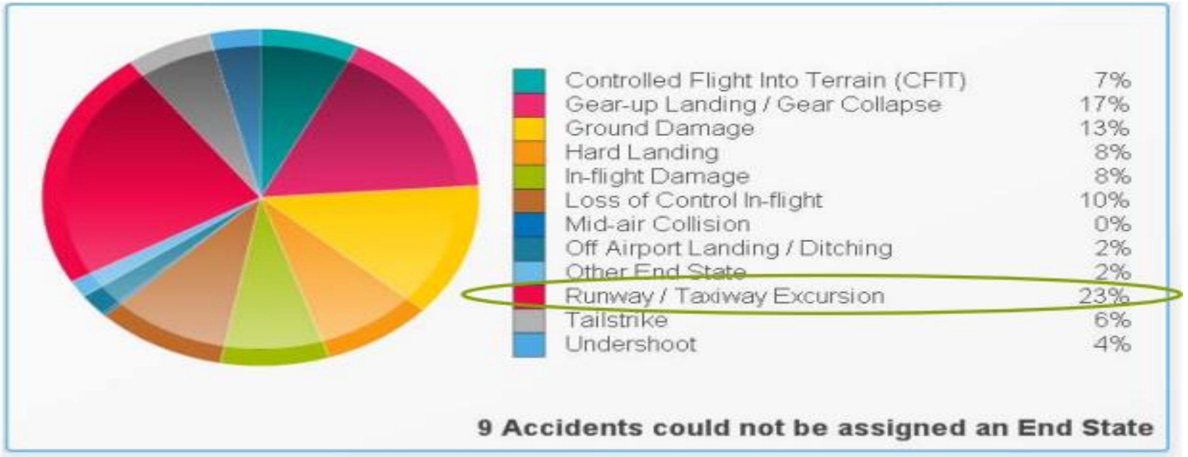
Claudia Czado
Technical University of Munich
Munich, Germany

11.10.2022

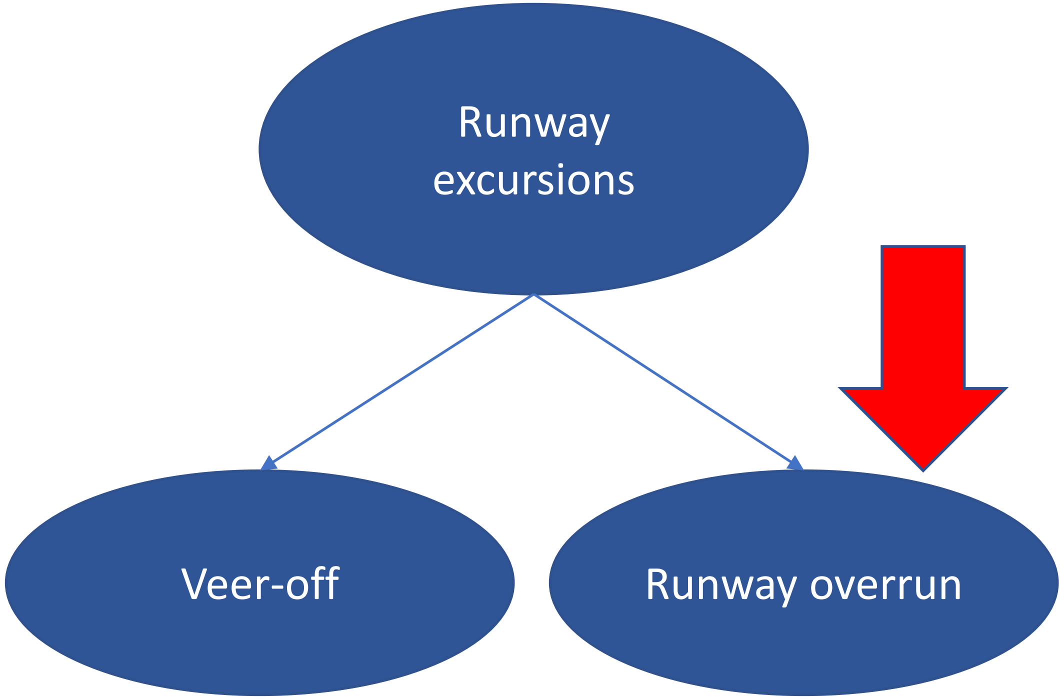
Table of contents

Motivation

Statistics: rate and the number of runway excursions remained the same during the last decade (IATA, 2019).



Percentage of runway excursions between 2005 and the first half of 2019, ICAO.

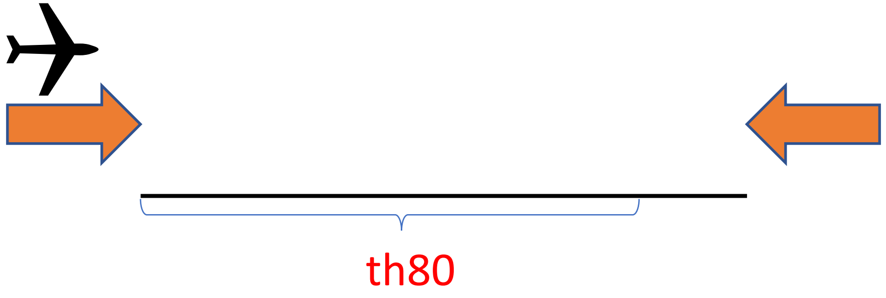


http://aerossurance.com/safety-management/sia-b777-autoland-localiser-re/

https://www.cbc.ca/news/canada/nova-scotia/halifax-airport-skylease-runway-overrun-1.4901699

Data set

- 711 flights of the same aircraft type.
- Landed on the same runway



Response: Smaller distance to controllable speed of 80 knots (th80) lowers the risk of runway overrun.

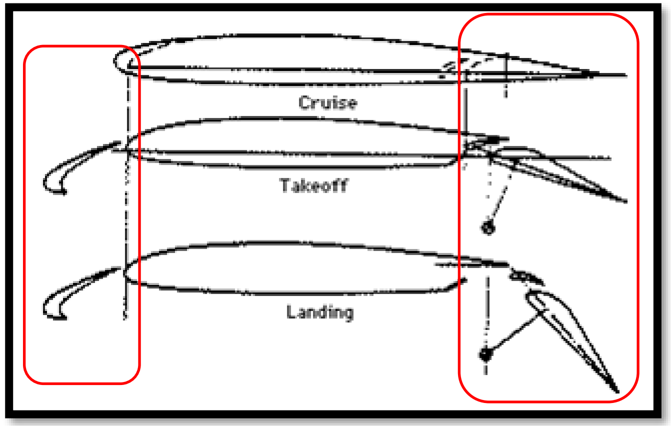
Covariates	App. speed deviation (asd)	Time brake started (tbs)
Headwind speed (hws)	Time reversers deployed (trd)	Break duration (bd)
Temperature (temp)	Time spoilers deployed. (tsd)	Touchdown distance (td)
Ref. air pressure (refAP)	Landing mass (lm)	Equivalent acceleration (ea)

Data set

Further factors

flapConfig	CONF 0^o	CONF 10^o	CONF 20^o	CONF 25^o	CONF 30^o
slatConfig	CONF 0^o	CONF 5^o	CONF 10^o	CONF 20^o	
revThrust	3/ALL OUT	FullRev	2 OUT		
splrSysStat	OP	$\leq$ 2 FAULT	$\leq$ 4 FAULT	5/ALL FAUL	
brkSysStat	OP	DEGRADED	INOP		
rwyCond	DRY	WET	VINCITY		

Flaps and Slats



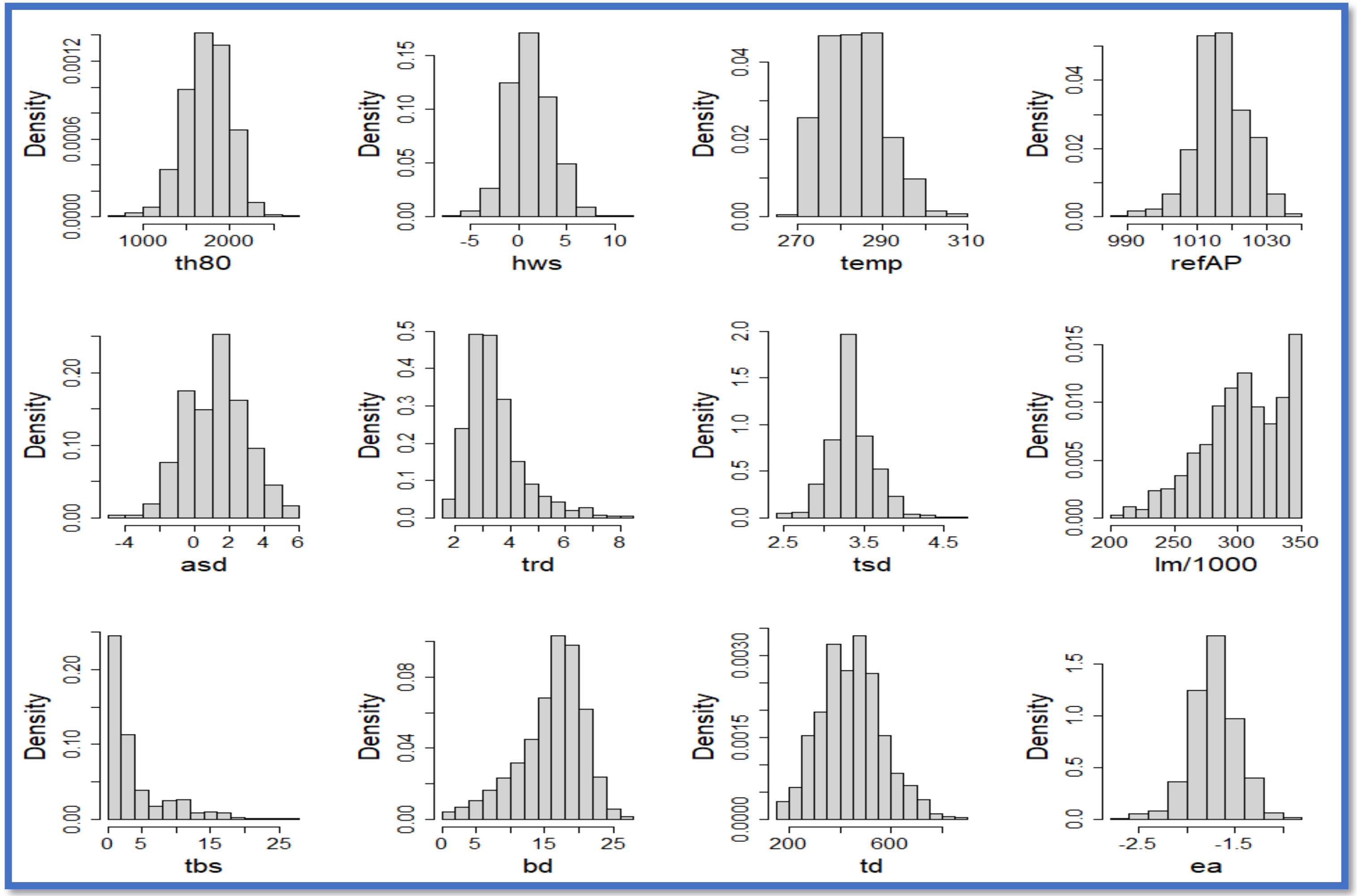
http://www.ae.utexas.edu/courses/ase463q/design_pages/spring03/active_wing/ATAK%20Technologies%20Website/Final%20Report/Web%20Theory.htm

https://www.quora.com/What-is-a-spoiler-in-aviation

https://www.quora.com/When-do-airliners-put-their-reverse-thrust-on-during-a-landing-Is-it-different-for-different-model-airliners

Data set

Histograms



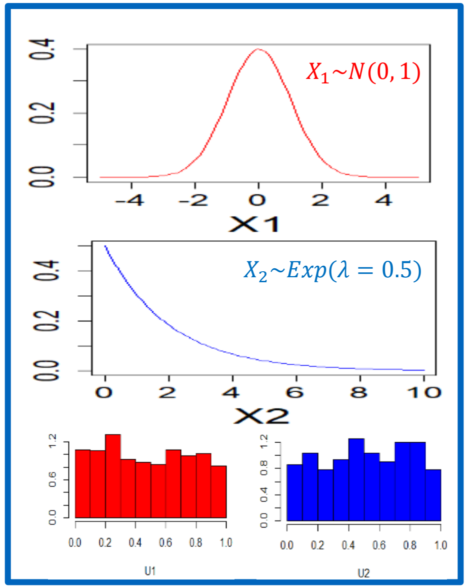
Methodology

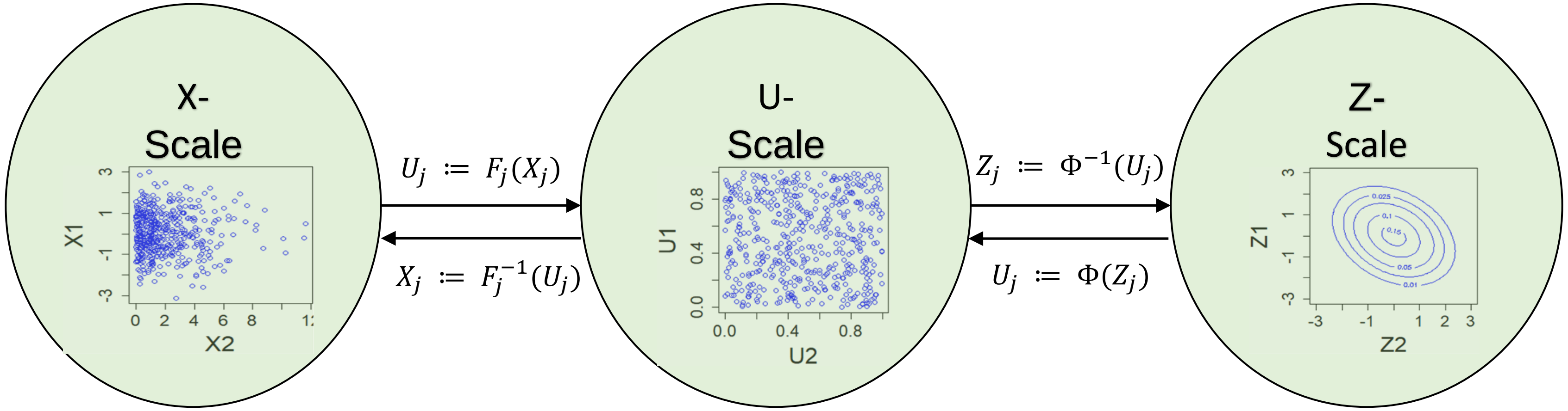
Copulas

How to construct a multivariate distribution with different margins and allow for flexible dependence patterns.

The theory of copulas started in 1959 with Sklar.

A d-dimensional copula $C(u_1, \dots, u_d)$ is a multivariate distribution on $[0,1]^d$ with uniformly distributed margins.



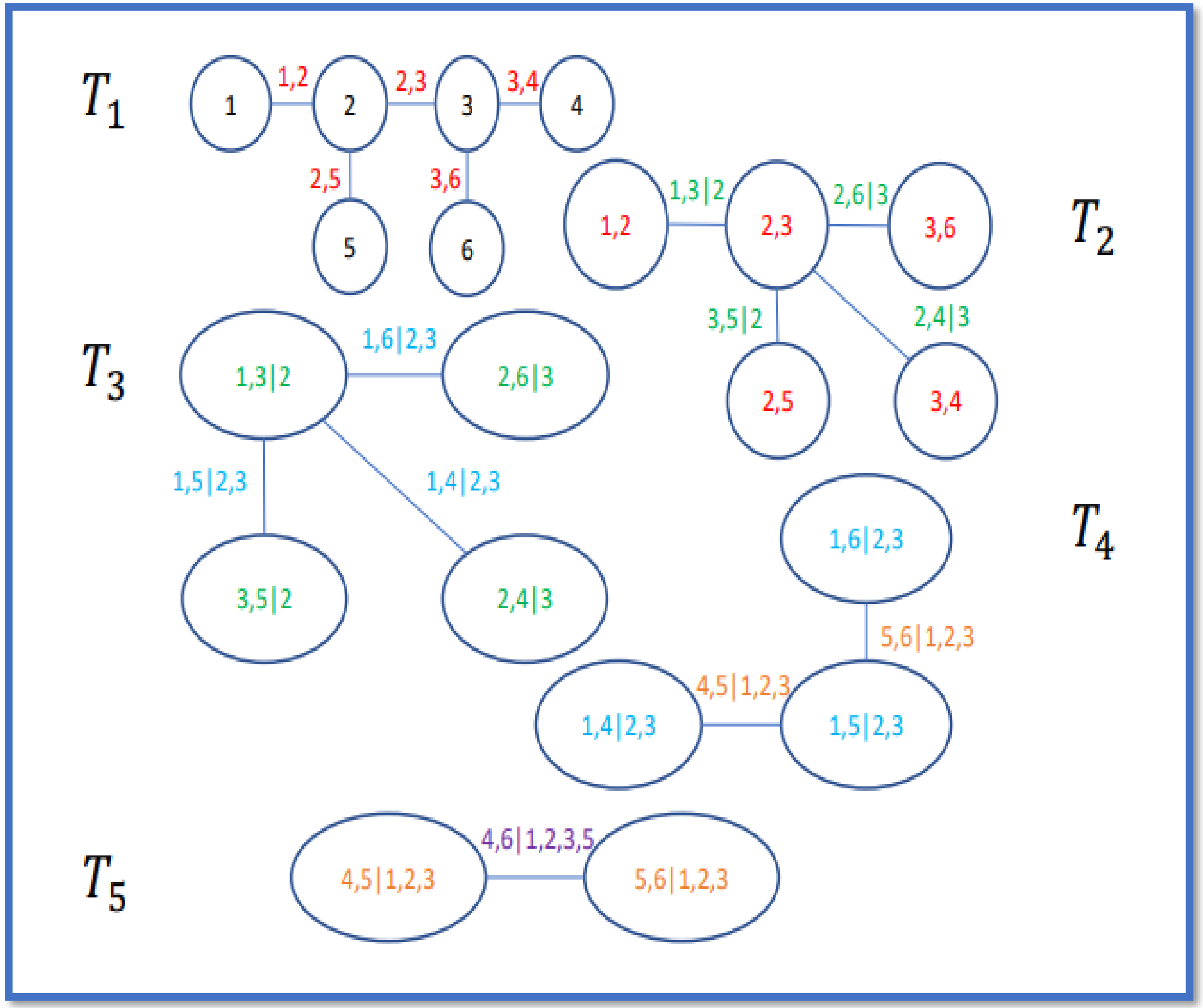


Methodology

Vine-copula

- Joe (1996) used conditioning to construct multivariate copulas using only bivariate copulas as building blocks.
- Bedford and Cooke (2001, 2002) proposed a vine tree structure, which identified regular (R)-vine copulas.
- Two sub classes of R-vine:
  - C-vine
  - D-vine

Six-dimensional R-vine



Methodology

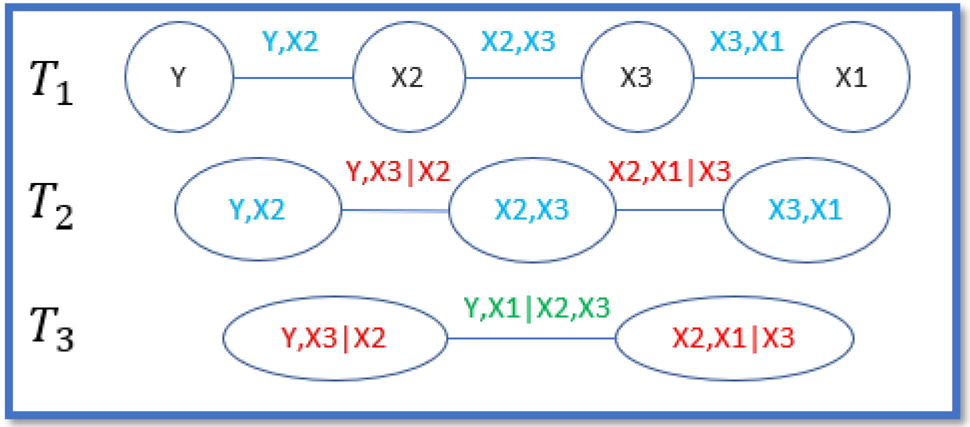
Drawable (D)-vine copula class

D-vine quantile regression:

$$q_\alpha(x_1, \ldots, x_d) := F^{-1}_{Y|X_1,\ldots,X_d}(\alpha | x_1, \ldots, x_d)$$

- Proposed by Kraus and Czado (2017).
- Expresses the conditional distribution function of the response given some variables.
- Allows for flexible modeling.
- Accounts for asymmetric tail dependencies.
- Is suitable to model extreme lower and upper tail dependencies.

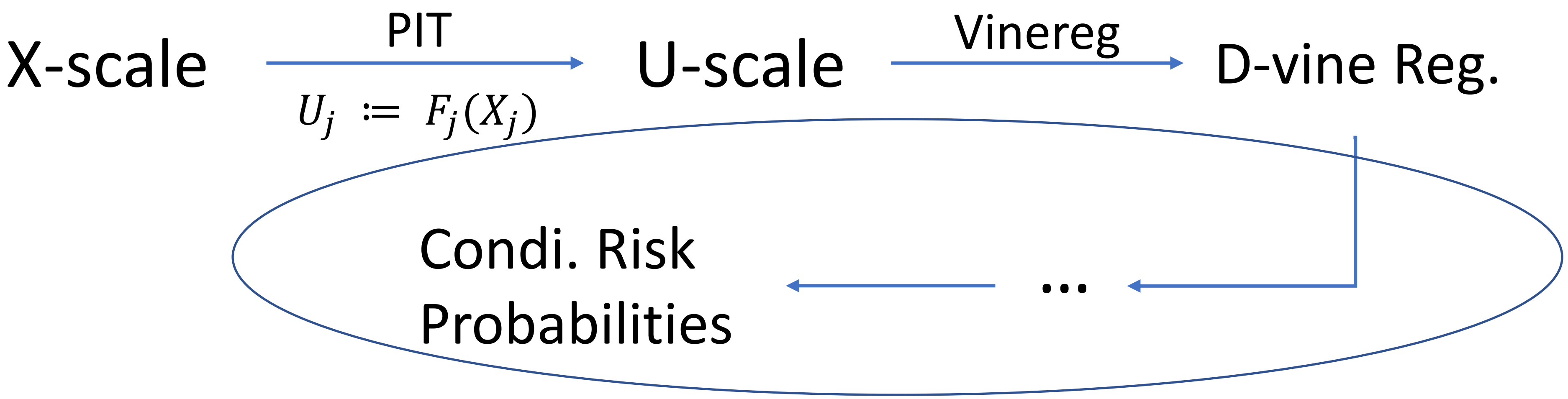
Four-dimensional D-vine



More information: https://www.math.cit.tum.de/en/statistics/statistics/research/vine-copula-models/

Methodology

Framework



- X_j covariate $j, j = 1, \ldots, d$.

Methodology

Estimated D-vine regression model:

$$\hat{q}_\alpha(x_1,\dots,x_d) := \hat{F}_Y^{-1}\left(\hat{C}^{-1}_{V|U_1,\dots,U_d}(\alpha|\hat{u}_1,\dots,\hat{u}_d)\right) \quad (1)$$

where:

- α quantile level
- $\hat{u}_j := \hat{F}_j(x_j)$ the estimated *PIT* of $x_j, j = 1,\dots,d$;
- $\hat{F}_Y^{-1}$ estimate of the inverse of the marginal distribution function;
- $\hat{c}^{-1}_{V|U_1,\dots,U_d}$ estimate of the inverse of the conditional copula quantile function of $V|U_1,\dots,U_d$.

Critical event probabilities using the fitted D-vine regression model:

$$\begin{aligned}\alpha_c(\boldsymbol{x}) &:= P(Y > c|\boldsymbol{X} = \boldsymbol{x}) = 1 - P(Y \le c|\boldsymbol{X} = \boldsymbol{x}) \\ &= 1 - F_{Y|X_1,\dots,X_d}(c|x_1,\dots,x_d)\end{aligned}$$

where:

- $\alpha_c(\boldsymbol{x})$ inverse of the conditional quantile function in (1) *w.r.t.* α;
- $c \in \mathbb{R}^+$, i.e. c = 2500.

Question: How to obtain $\alpha_c(\boldsymbol{x})$?

- Rosenblatt transform (Rosenblatt, 1952)

Methodology

Rosenblatt transform

Purpose: compute the conditional distribution function at c, $F_{Y|X_1,\dots,X_d}(c|x_1,\dots,x_d)$.

Idea:

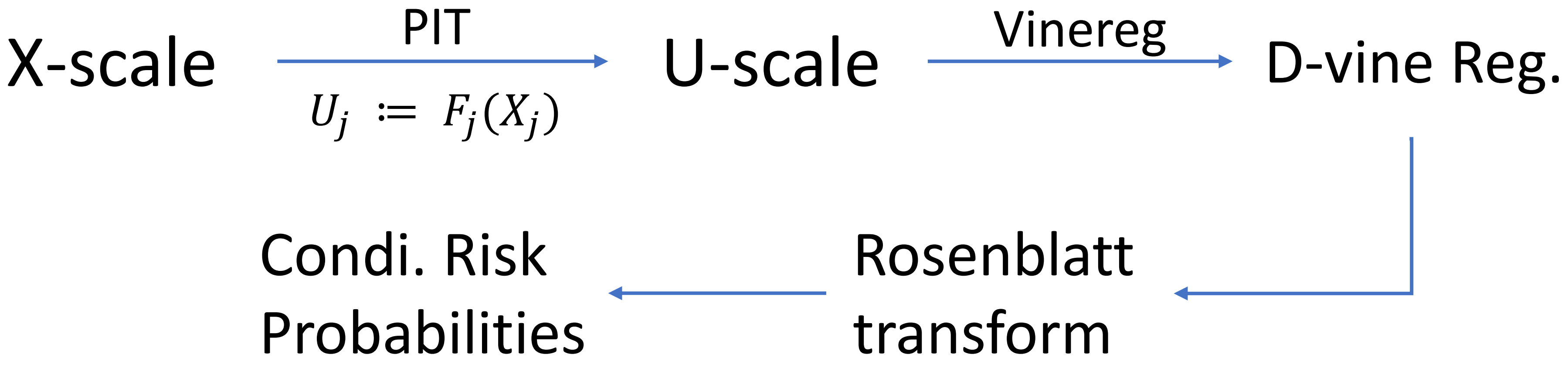
- For a random vector $\boldsymbol{V} = (V_1,\dots,V_p)^T$ with joint distribution function F and conditional distribution function $F_{j|1,\dots,j-1}, j = 2,\dots,p$.
- $\boldsymbol{U}$ is joint uniformly distributed $= \begin{cases} U_1 := F(V_1), \\ \vdots \\ U_j := F_{j|1,\dots,j-1}(V_j|V_1,\dots,V_{j-1}), j = 2,\dots,p. \end{cases}$
- Using the joint density f for a D-vine distribution (Czado, 2010), we can express $F_{Y|X_1,\dots,X_d}(c|x_1,\dots,x_d)$ for aribitrary values of c given the value $\boldsymbol{X} = x$.

$$f(x_1,\dots,x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{i=1}^{d-1} \prod_{j=i+1}^{d} c_{ij;i+1,\dots,j-1}\left(F_{i|i+1,\dots,j-1}(x_i|x_{i+1},\dots,x_{j-1}),\right.$$
$$\left.F_{j|i+1,\dots,j-1}(x_j|x_{i+1},\dots,x_{j-1});x_{i+1},\dots,x_{j-1}\right),$$

for distinct indices i and $j, i < j$.

Methodology

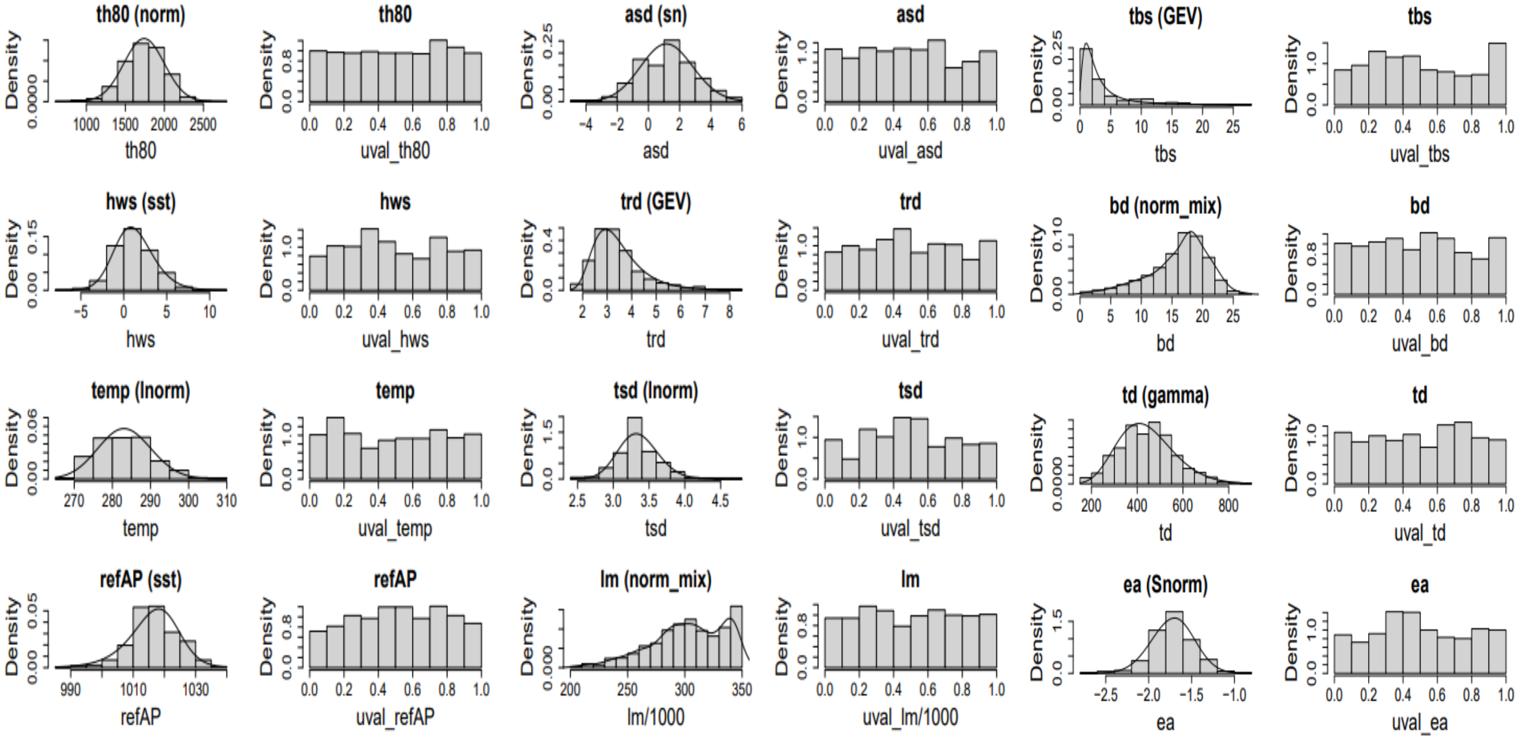
Framework



Analysis and estimated conditional risk Probabilities

Back to the data

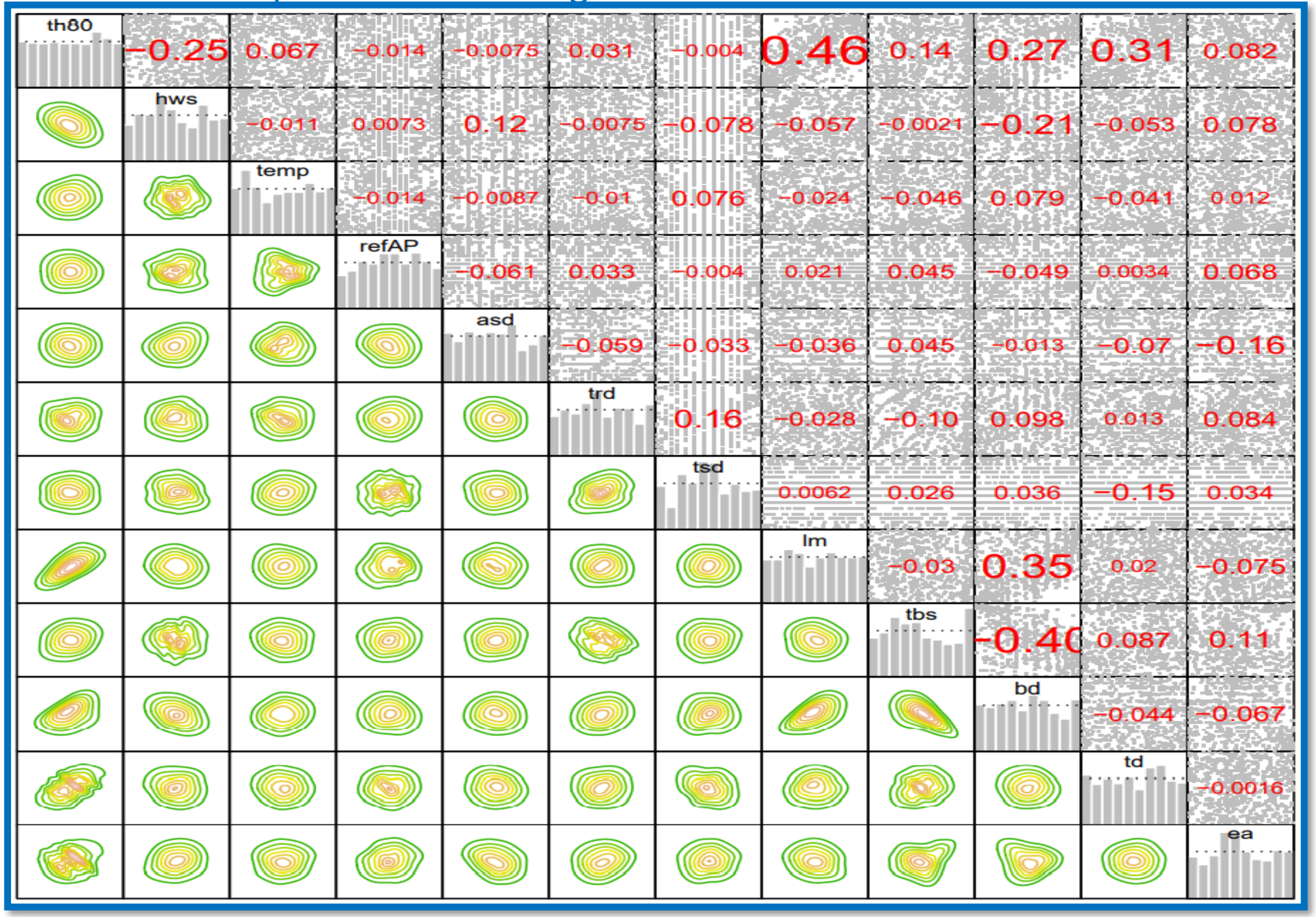
- Parametric distributions



(th80)	Normal
Headwind speed (hws)	Skew Student t.
Temperature (temp)	Log Normal
Ref. air pressure (refAP)	Skew Student t.
App. speed deviation (asd)	Skew Normal
Time reversers deployed (trd)	Generalized Ext. Value
Time spoilers deployed. (tsd)	Log Normal
Landing mass (lm)	Mixture of Normals
Time brake started (tbs)	Generalized Ext. Vaue
Break duration (bd)	Mixture of Normals
Touchdown distance (td)	Gamma
Equivalent acceleration (ea)	Skew Normal

Analysis and estimated conditional risk Probabilities

Dependencies among the variables on U-scale



Analysis and estimated conditional risk Probabilities

Estimated D-vine regression model

$$\hat{q}_\alpha(hws, \ldots, ea) = \hat{F}_{th80}^{-1}\left(\hat{C}_{V_{th80}|U_{hws},\ldots,U_{ea}}^{-1}(\alpha|\hat{u}_{hws}, \ldots, \hat{u}_{ea})\right)$$

Note:

- tsd was not selected to maximize the conditional log likelihood (*cll*)

Objective:

Find flight i with

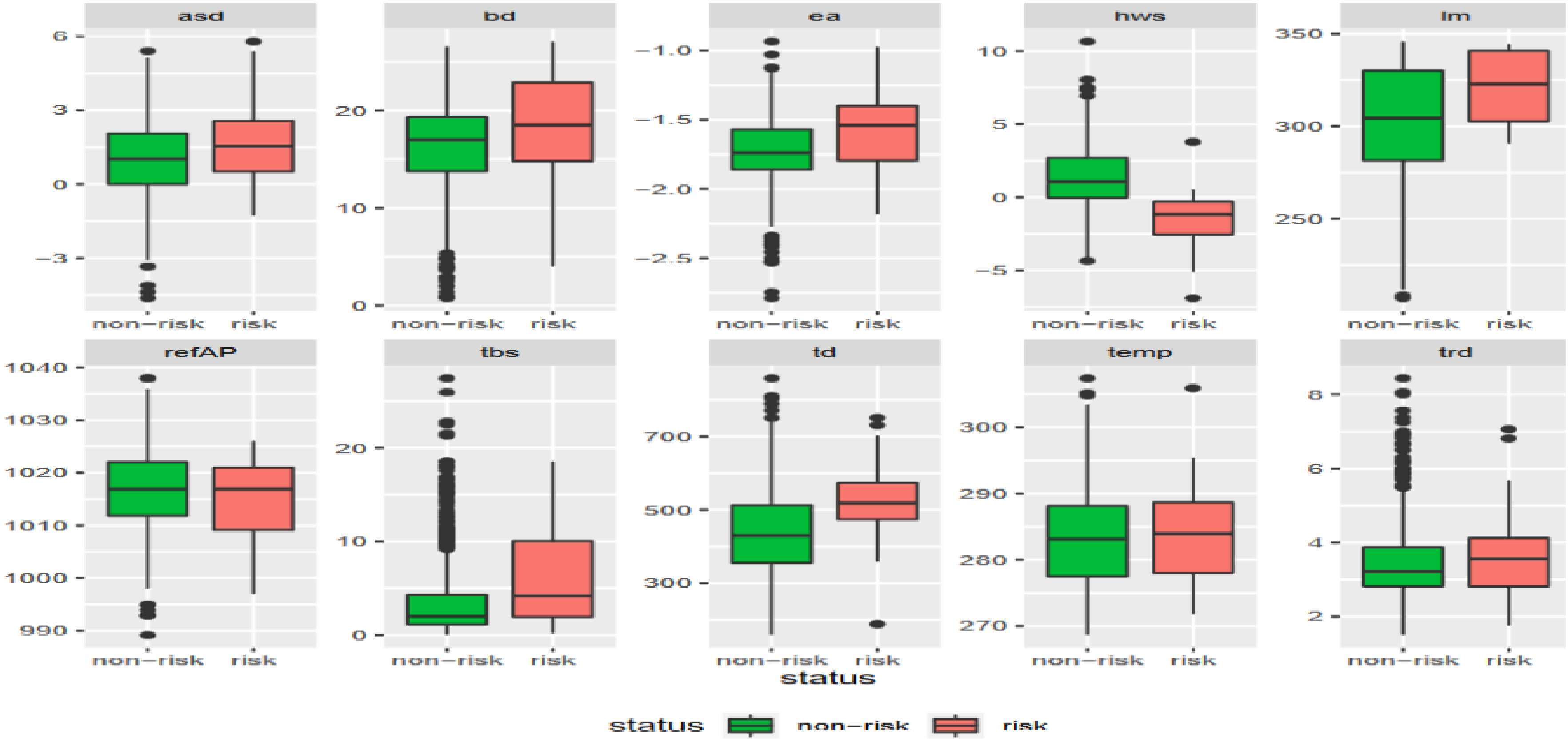
$$P(th80_i > 2500\, m | hws_i, \ldots, ea_i) > 10^{-3}, \quad \text{for } i = 1, \ldots, 711.$$

Paper: HH Alnasser, C Czado - arXiv preprint arXiv:2205.04591, 2022

Analysis and estimated conditional risk Probabilities

Results:

- Identified 41 risky flights. ($c = 2500\ m,\ Prob = > 10^{-3}$)

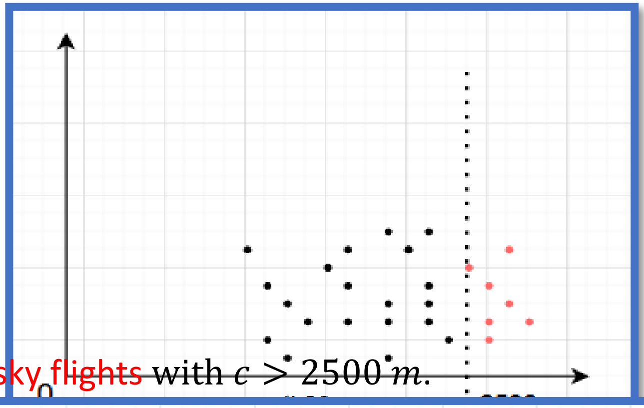


D-vine based subset simulation

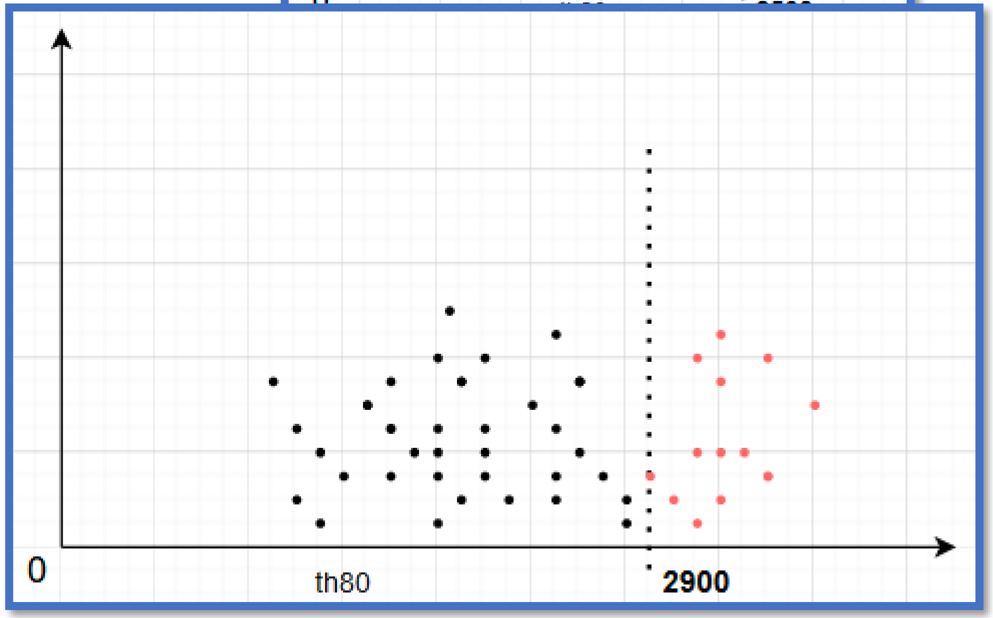
So far:

- Identified 41 risky flights. ($\boldsymbol{c = 2500\,m}$, $\boldsymbol{Prob = > 10^{-3}}$).
- Let's say we want to **generate more flights** with $\boldsymbol{c > 2500\,m}$, $\boldsymbol{Prob\ \ > 10^{-6}}$.

Risky flights with $c = 2500\ m$.



Risky flights with $c > 2500\ m$.



D-vine based subset simulation

Algorithm:

Iteration 0: Crude Monte Carlo simulation (CMS)

- Generate $R\epsilon\mathbb{R}^{+}, i.e. R = 10^{5}$, realizations from the fitted D-vine regression model
  $$y^{r}, x_{1}^{r}, \dots, x_{d}^{r}, r = 1, \dots, R.$$
- How many $y^{r} > c$, $e.g. c = 2900\ m.$
- Sort according to y^{r}.

Iteration 1: Subset simulation (SS)

- Select the highest $P_0\epsilon(0,1), i.e. P_0 = 10\%$. (seeds)
- Generate new realizations $T_1 \in \mathbb{R}^{+}$, i.e. $T_1 = 1, \dots, (1 - P_0) \times R$
  $$y^{t_1}, x_{1}^{t_1}, \dots, x_{d}^{t_1}, t_1 = 1, \dots, T_1.$$
- Capture the dependence via Rosenblatt.
- How many $y^{t_1} > 2900\ m.$
- Sort according to y^{r}.

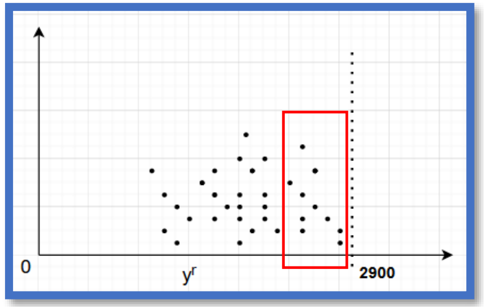
Iteration 2: Subset simulation (SS)

- Select the highest $P_0\epsilon(0,1), i.e. P_0 = 10\%$. (seeds)
- Generate new realizations $T_2 \in \mathbb{R}^{+}$, i.e. $T_2 = 1, \dots, (1 - P_0) \times R$
  $$y^{t_2}, x_{1}^{t_2}, \dots, x_{d}^{t_2}, t_2 = 1, \dots, T_2.$$
- Capture the dependence via Rosenblatt.
- How many $y^{t_2} > 2900\ m.$
- Sort according to y^{r}.

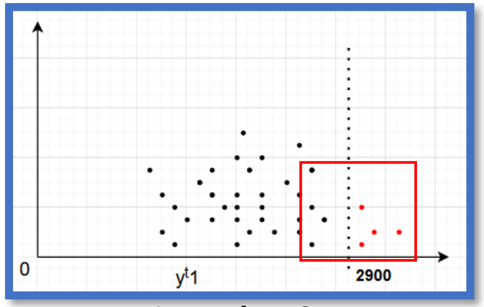
⋮

Continue until a **certain number of observations > 2900** $\boldsymbol{m}$.

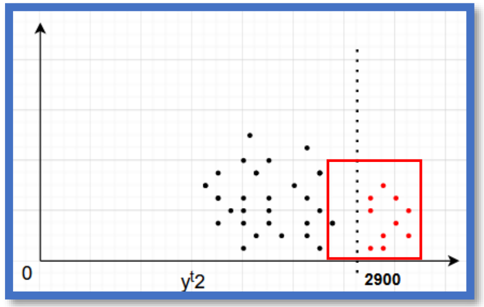
Iteration 0



Iteration 1



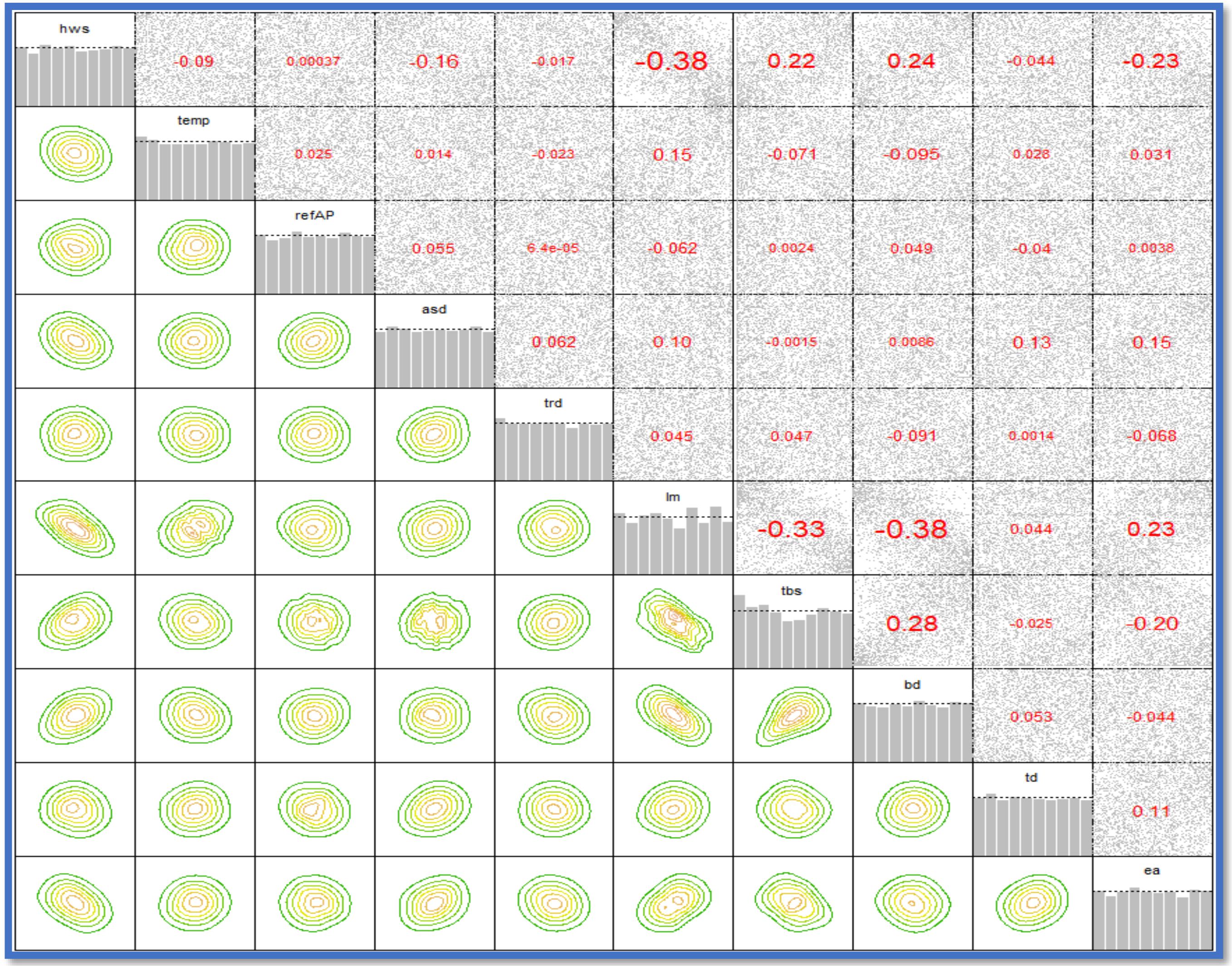
Iteration 2



⋮

D-vine based subset simulation

Dependencies among the covariates on U-scale



Conclusion

I have shown:

- Methodology for the **estimation of conditional risk probabilities** using **D-vine regression.**
- Identified **41 risky flights** (c = 2500 meters, prob. = $> 10^{-3}$)
- **D-vine** based **subset simulation** for the estimation of **rare event probabilities**.

Thank you