Comparison of probabilistic and alternative evidence theoretical methods for the handling of parameter uncertainties resulting from variability and/or partial ignorance in safety cases

Rainer Barthel
Brenk Systemplanung GmbH, Aachen, Germany

1. Aleatory and epistemic parameter uncertainties
2. Bayesian Probability Theory (PT)
3. Evidence Theory (ET) – a generalization of PT
4. Propagation of parameter uncertainties in models
5. Description of statistical dependencies by Copulas
6. Conclusions and recommendations

1. Aleatory and epistemic parameter uncertainties

- confidence in the safety of a Deep Geological Repository requires identification & assessment of **uncertainties** concerning evolution **scenarios**, **models** for predicting repository performance and the values/distributions of model **parameters** (100s of uncertain inputs considered in performance assessments for WIPP and YMR)
- aleatory uncertainty = variability (inherent randomness caused by stochastic processes or heterogeneity within a basic population), objective and irreducible
- epistemic uncertainty = incertitude (owing to data imprecision, imperfect knowledge or ignorance, sampling number limitation), subjective and reducible by additional data/information
- both uncertainty types are usually described in terms of probability

2. Bayesian Probability Theory (PT)

- PT is adequate for the modeling of random variability
- Bayesian PT is appropriate for presentation & reduction of incertitude concerning statistical distribution parameters by means of new data:

$p_X(x|\boldsymbol{\theta})$ = PDF of a random variable X; $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ iid sample

likelihood function: $$L(\boldsymbol{\theta} \mid \mathbf{x}) = \prod_{i=1}^{n} p_X(x_i \mid \boldsymbol{\theta})$$

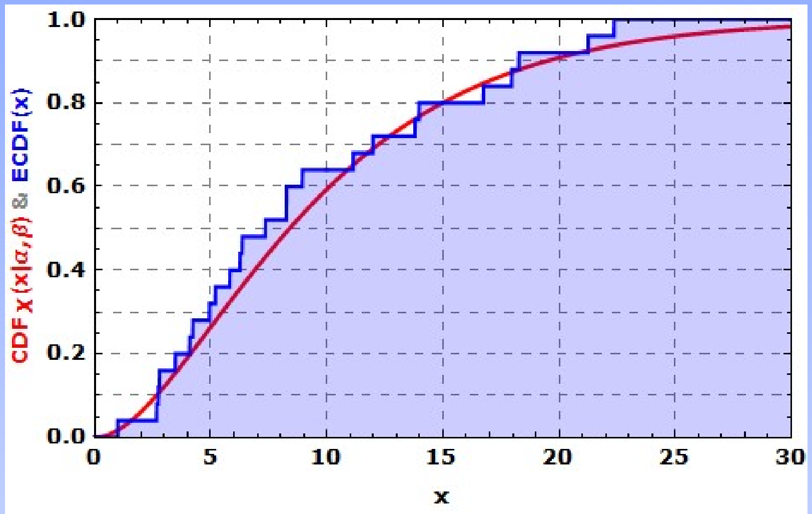
$p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ = prior density of distribution parameter(s) $\boldsymbol{\theta}$

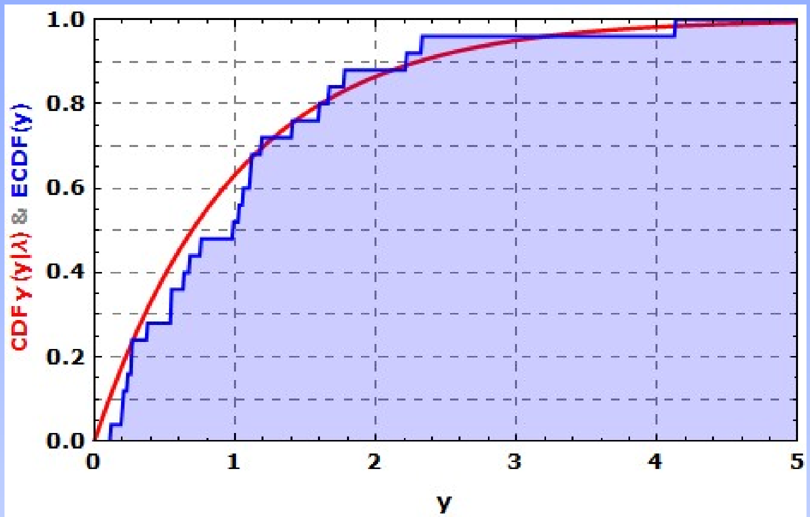
=> posterior density for $\boldsymbol{\theta}$ (incertitude): $$p_{\boldsymbol{\Theta}}(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{L(\boldsymbol{\theta} \mid \mathbf{x}) \cdot p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta} \mid \mathbf{x}) \cdot p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

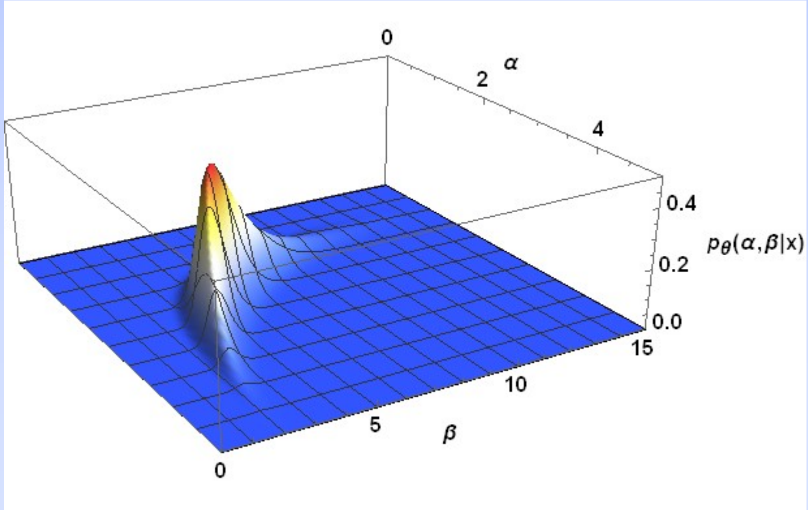
Bayesian PT enables 2nd order probability (2D) Monte Carlo simulation with outer loop for epistemic uncertain distribution parameters Θ and inner loop for random variables X with randomly sampled θ-values.

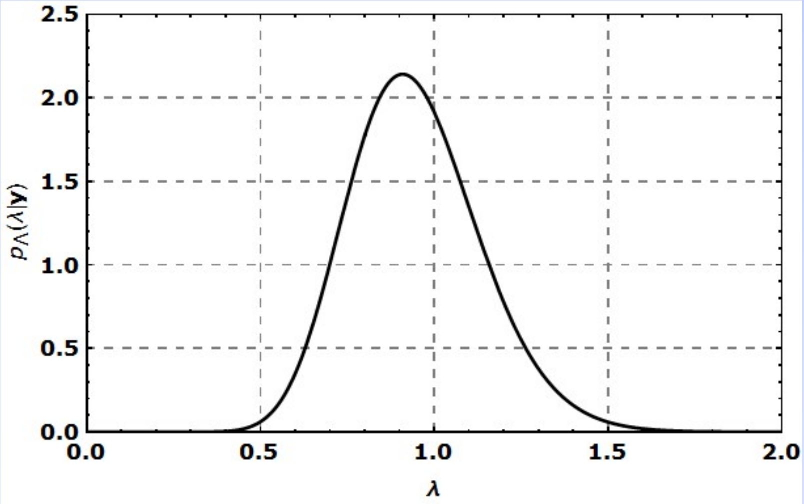
2. Bayesian Probability Theory (PT)
Example 1

$X \sim Gam(\alpha, \beta)$; $Y \sim Exp(\lambda)$; $(\alpha, \beta, \lambda) = (2, 5, 1)$; $n_x = n_y = 25$; $Z = X \cdot Y$





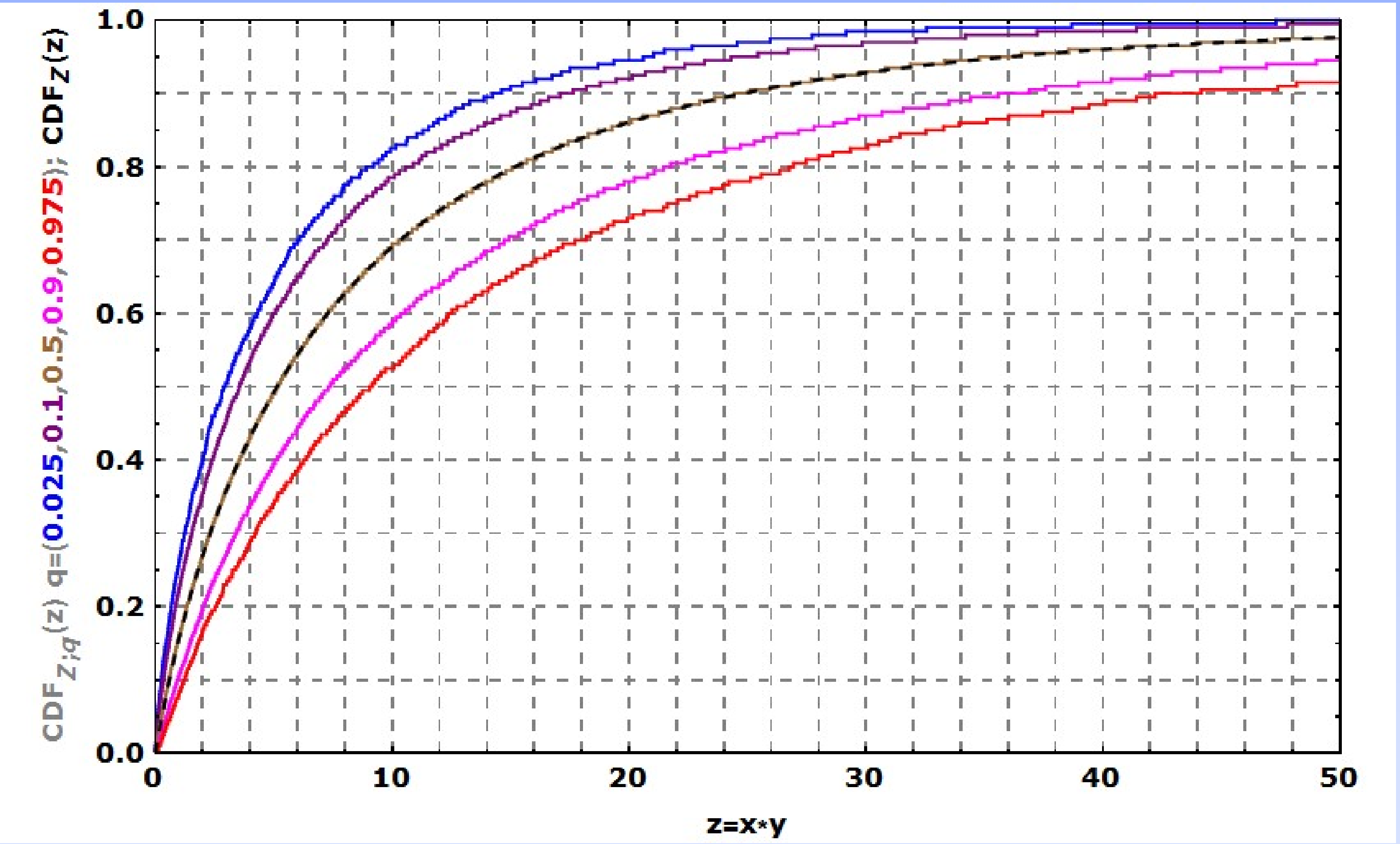




2. Bayesian Probability Theory (PT)

Example 1

Cumulative distribution functions: $CDF_Z(z)$ describes variability; $CDF_{Z;q}(z)$ (q = confidence level) characterize incertitude owing to parameters α, β, λ



3. Evidence Theory (ET) – a generalization of PT

Ω = set of events x_i (i = 1 to k); including Ω and the empty set $\emptyset$, the set of all subsets A (so-called power set P_Ω) has 2^k elements;

PT defines probabilities p_i (chances or subjective degrees of belief) for all singletons x_i with:

$$\sum_i p_i = 1, \quad \Pr(A) = \sum_{i:x_i \in A} p_i \ , \Rightarrow \Pr(A) + \Pr(A^c) = 1$$

ET defines probabilities m(A) (basic probability assignments) for all subsets A by a mapping

$$m : P_\Omega \to [0,1] \quad \text{with} \quad \sum_{A \subseteq \Omega} m(A) = 1$$

Focal sets (A with m(A) > 0) define two non-additive belief measures,

Belief: $\mathrm{Bel}(A) = \sum_{\emptyset \neq B \subseteq A} m(B)$, **Plausibility**: $\mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B)$

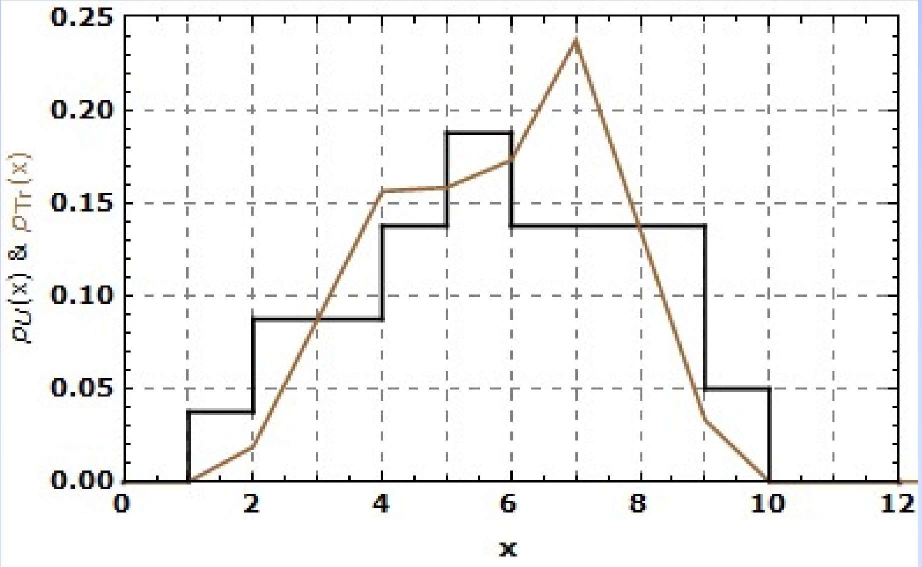
=> $\mathrm{Bel}(A) \leq \mathrm{Pl}(A)$, and (if $m(\emptyset) = 0$) $\mathrm{Bel}(A) + \mathrm{Pl}(A^c) = 1$

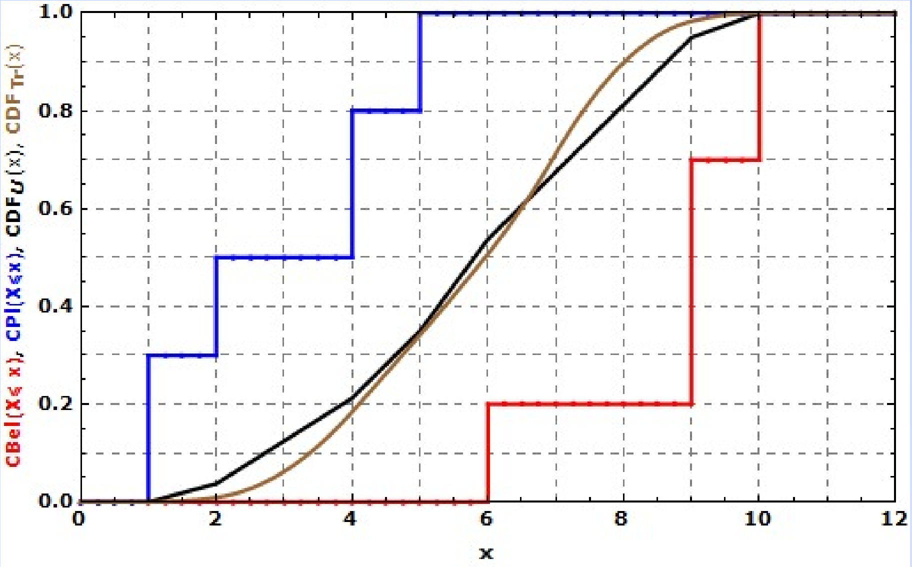
ET generalizes PT; represents degree of **Ignorance** = Pl(A) - Bel(A)

3. Evidence Theory (ET) – a generalization of PT
Example 2

Expert judgement for model parameter X (e. g. evidence transfer from former investigations to another site or to future conditions): $x \in I_i$ with m_i; i = 1 - 4; **{I, m}** = ({[1, 9], 0.3}; {[2, 6], 0.2}; {[4, 10], 0.3};{[5, 9], 0.2});

ET: calculation of cumulative functions CBel(X ≤ x) and CPl(X ≤ x);
Bayesian **PT** could assume uniform / symmetric triangular distributions in the intervals I_i and calculate $p_U(x)$ / p_{Tr} (x) by averaging with m_i

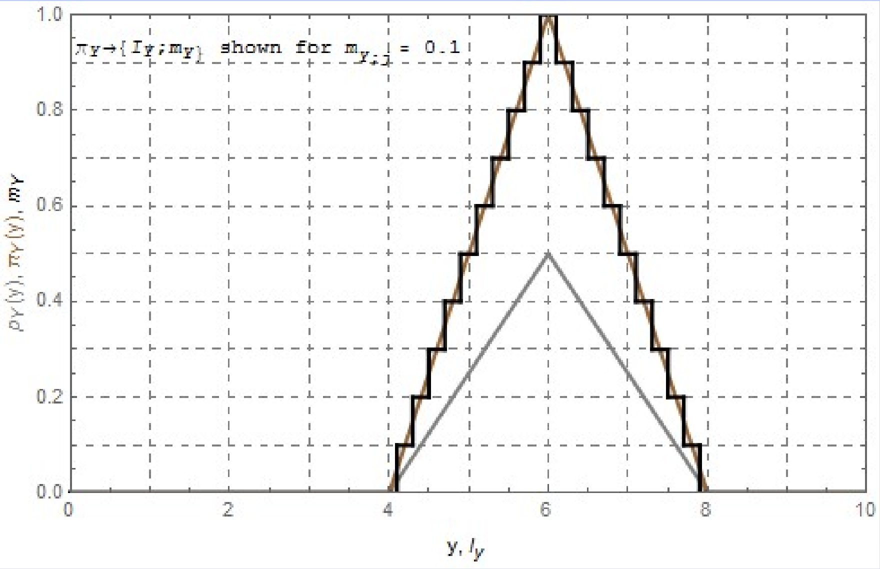


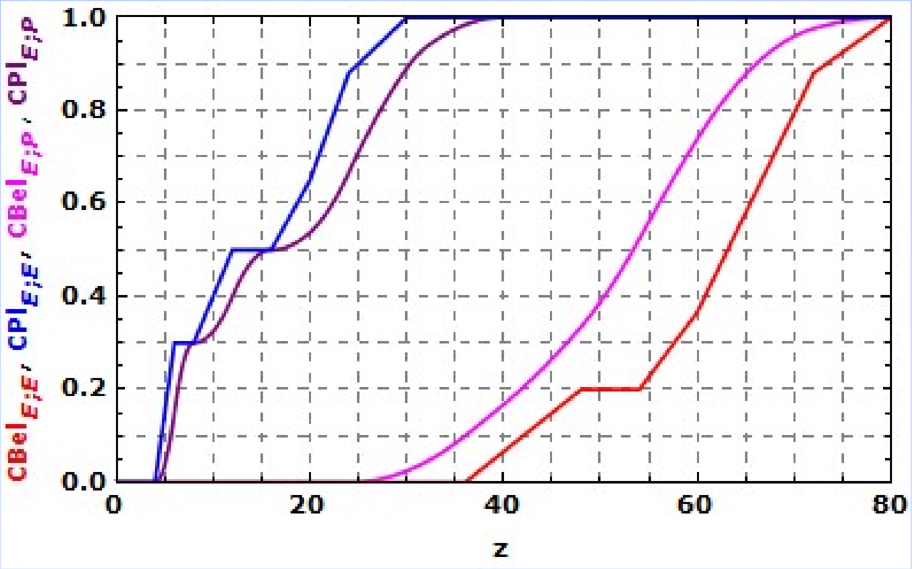


4. Propagation of parameter uncertainties in models

ET includes PT and Possibility Theory (specific case if focal sets are nested), and allows for combined propagation of respective belief measures in models, e. g.:
X and Y independent variables with $\{\mathbf{I_X}, \mathbf{m_X}\}$ and $\{\mathbf{I_Y}, \mathbf{m_Y}\}$;
$Z = X\cdot Y$; => $\{\mathbf{I_Z}, \mathbf{m_Z}\}$ with $I_{Z;i,j} = I_{X;i}\cdot I_{Y;j}$ and $m_{Z;i,j} = m_{X;i}\cdot m_{Y;j}$

Example 3: X with ET-measures as in Example 2; for Y a probabilistic triangular distribution Y~Tr(4, 6, 8) leading to Z-case “E;P”, and a respective possibilistic distribution $\pi_Y(4, 6, 8)$ transformed to $\{\mathbf{I_Y}, \mathbf{m_Y}\}$ with $m_{Y;j} = \Delta\pi \rightarrow 0$, leading to “E;E”

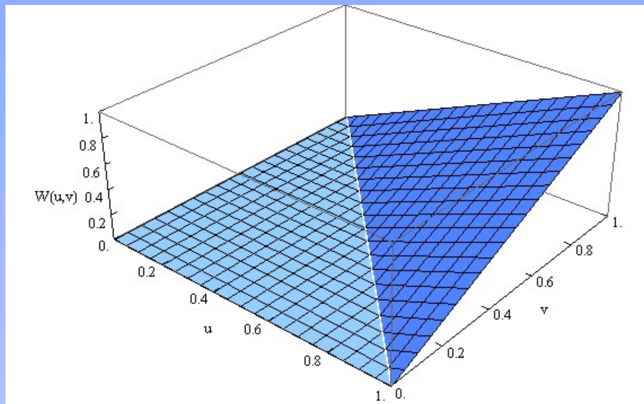




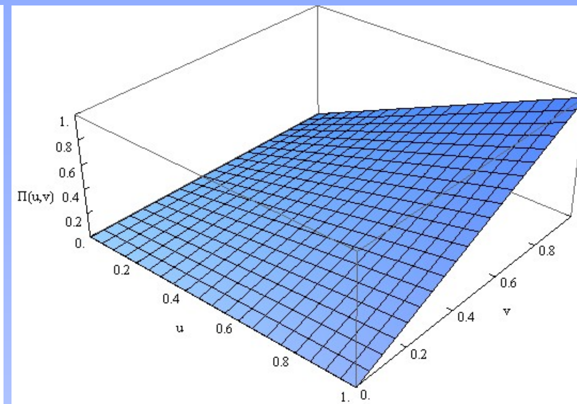
5. Description of statistical dependencies by Copulas

- In Monte Carlo simulation programs, statistical dependencies between variables X and Y is usual modelled by SPEARMAN rank correlation $\rho_{X,Y}$.
- Copulas are much more expressive and allow for separated analysis and simulation of joint distribution functions H(x, y) by their margins F(x) and G(y) and a Copula function C(u, v): $[0, 1]^2 \rightarrow [0, 1]$ that not depends on the types of the margins, but only characterizes dependency between X and Y.
- For H(x, y) with margins F(x) and G(y) exists a Copula C(u, v) such that for $\forall$ x, y: H(x, y) = C(F(x), G(y)); C is unique if F and G are continuous.
- FRÉCHET-HOEFFDING bounds W(u, v) and M(u, v) are universal for all Copulas: $W(u, v) \equiv \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) \equiv M(u, v)$.
  If dependence between X and Y is unknown, copulas W and M provide bounds for Monte Carlo simulation and sensitivity analyses.
- For an excellent introduction and overview see: Roger B. Nelsen, "An Introduction to Copulas" (2nd Edition), Springer 2007.

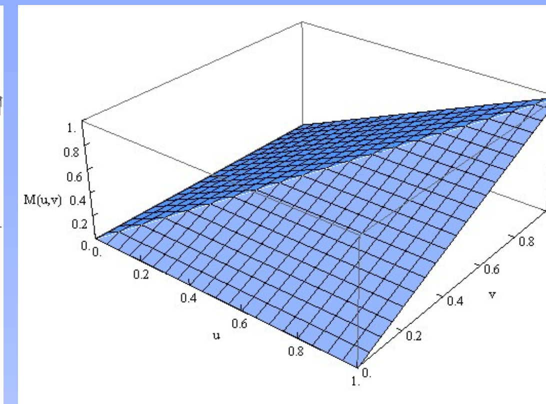
5. Description of statistical dependencies by Copulas - Examples



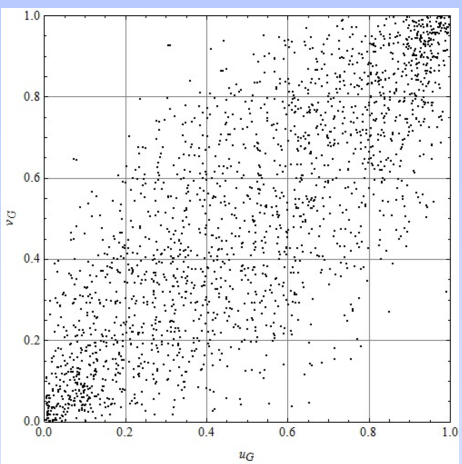
lower FH-bound W



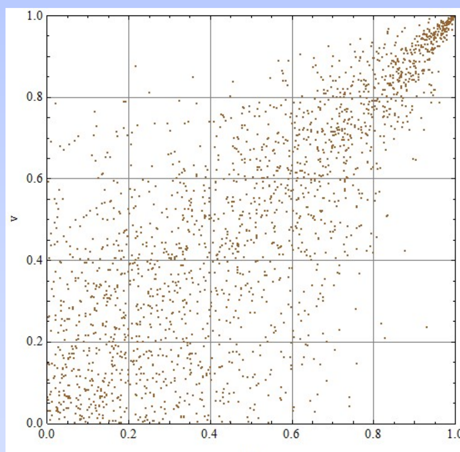
independence Copula $\Pi = u \cdot v$



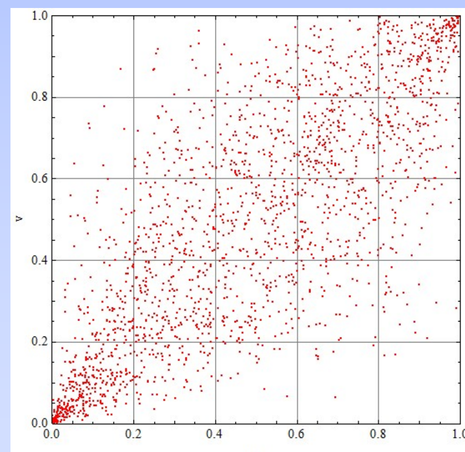
upper FH-bound M



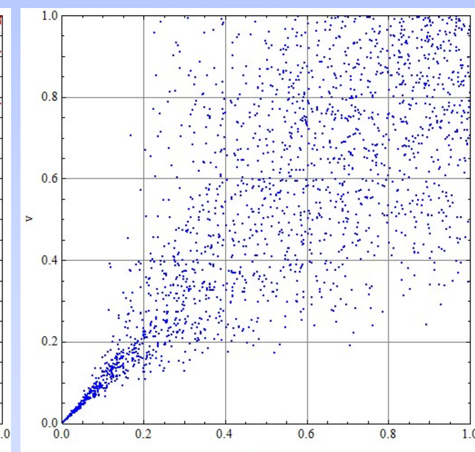
Gauss-Copula C_G



Joe-Copula C_J



Copula C_{12}



Copula C_{19}

(scatterplots of C_G, C_J, C_{12} and C_{19} for KENDALL concordance coefficient $\tau = 0.59$)

6. Conclusions and recommendations

1. Aleatory and epistemic uncertainties (variability and incertitude) of model parameters should be analysed/specified separately. For computer simulations they should be modelled with appropriate mathematical methods that are adequate to the available evidence.
2. Bayesian PT is very useful for assessing incertitude of distribution parameters estimated on the basis of sampling data. However, use of probabilistic terms for subjective degrees of belief concerning pure epistemic uncertain parameters is questionable.
3. Evidence Theory appropriately allows for expressing subjective degrees of belief by two measures, Belief (Bel) and Plausibility (Pl), also quantifying the degree of (partial) ignorance.
4. Modelling of statistical dependencies between random variables by Copulas is recommended. Research seems to be necessary with respect to dependencies between epistemic uncertain parameters.