Measuring Experts' performance in assessing dependence: experiences and open questions.

- Motivation
- Conditional probabilities of exceedance
- Ratios of rank correlations
- Determinant correlation matrix
- Dependence Calibration Score

Oswaldo Morales Nápoles
CiTG, TU Delft
o.moralesnapoles@tudelft.nl

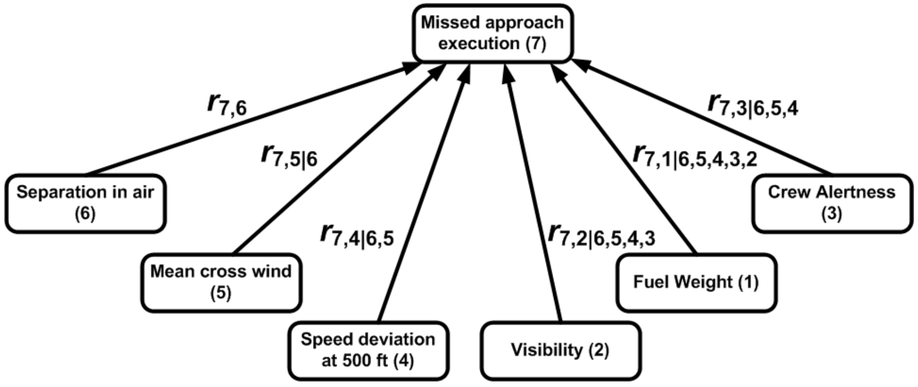
DEPENDENCE ELICITATION IS ESSENTIALLY DIFFERENT THAN UNCERTAINTY ELICITATION.

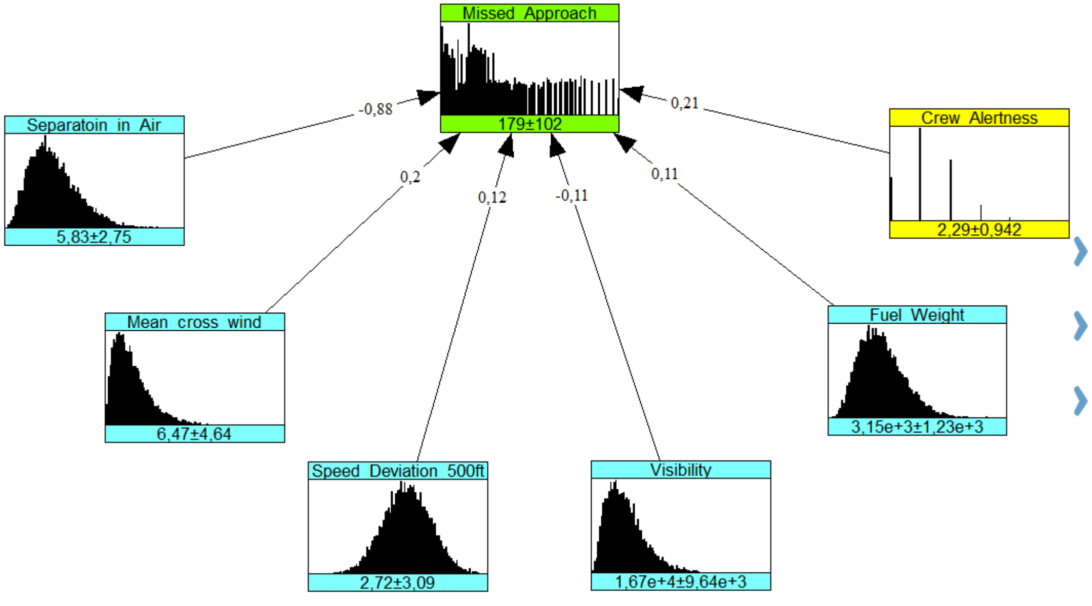
USUAL MEASURES FOR UNCERTAINTY ELICITATION ARE NOT SUFFICIENT.

MOTIVATION

- Not always all JOINT data available.

CONTROLLED FLIGHT INTO TERRAIN





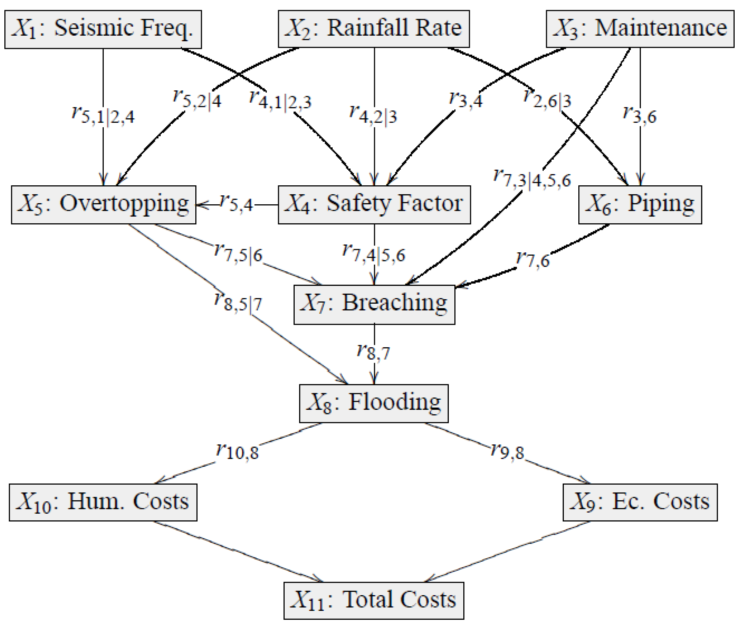
- P(7>med| 6>med)
- P(7>med| 6>med, 5>med) …
- P(7>med| 1>med, 2>med, 3>med, 4>med, 5>med, 6>med)

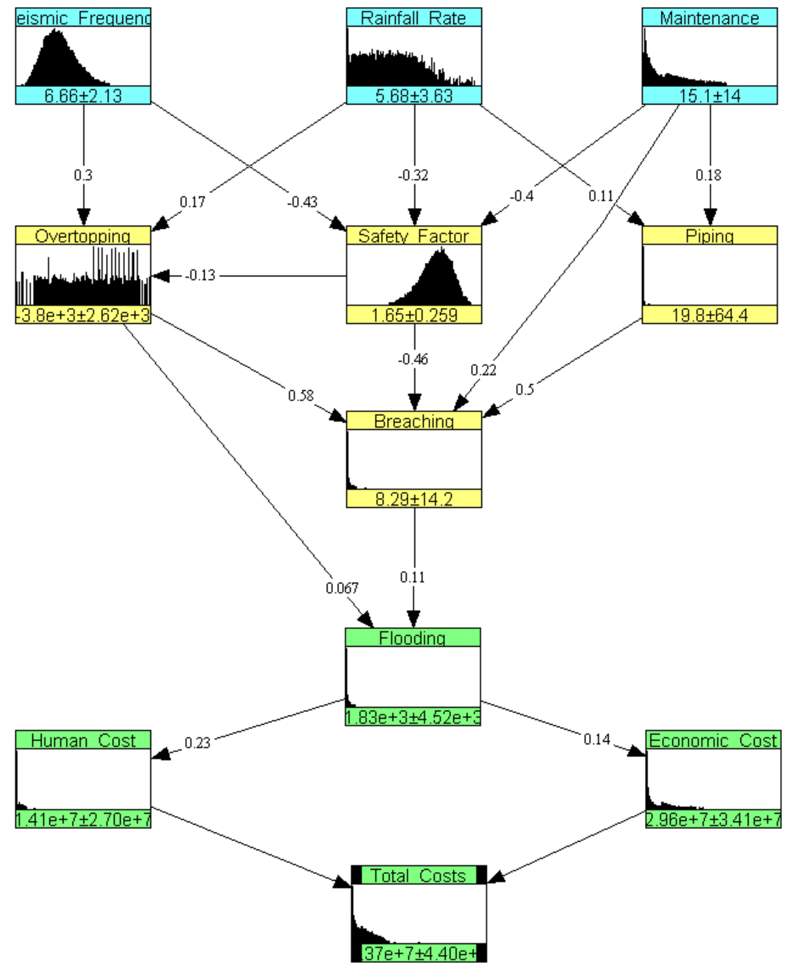
- R(7,6)
- R(7,5)/R(7,6) ….
- R(7,1)/R(7,6)

TU Delft

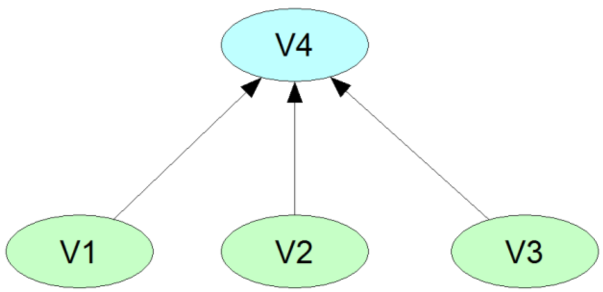
EARTH DAMS IN MEXICO

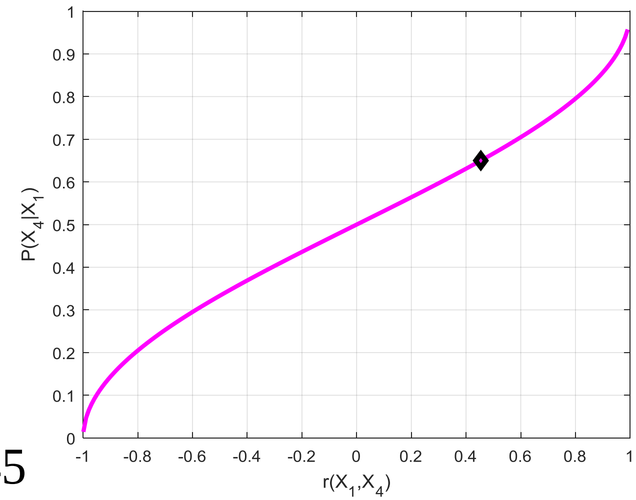
- 4 experts
- 16 (conditional) rank correlations
- Ratios of rank correlations





BN DEPENDENCE QUANTIFICATION

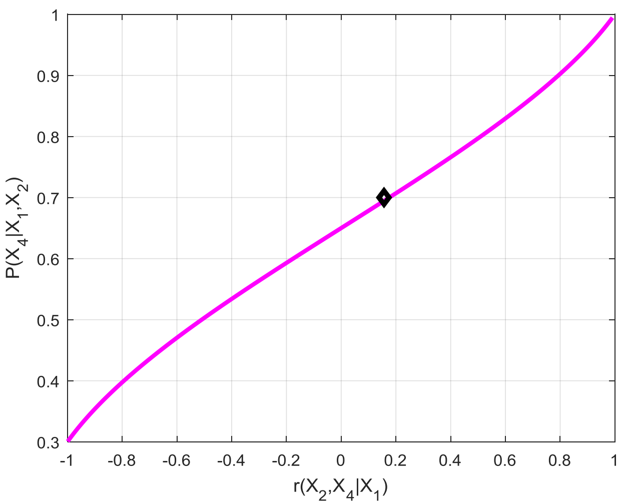


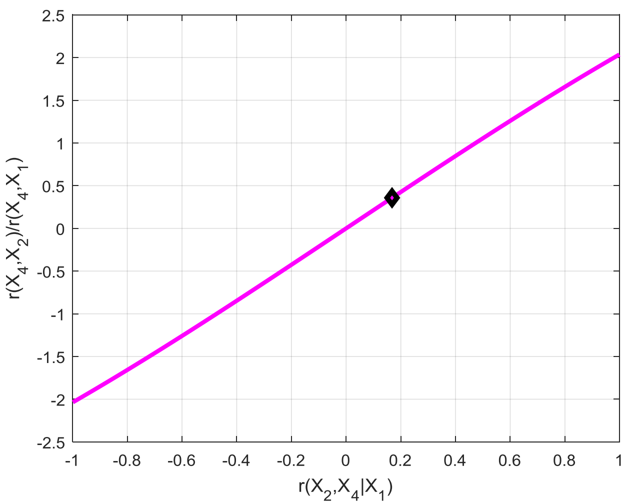


$$P_1^{e_1} = 0.65 \rightarrow r_{4,1}^{e_1} = 0.45 \qquad P_1^{e_2} = 0.65 \rightarrow r_{4,1}^{e_2} = 0.45$$

$$P_2^{e_1} = 0.70 \rightarrow r_{4,2|1}^{e_1} = 0.18 \qquad R_2^{e_2} = 0.35 \rightarrow r_{4,2|1}^{e_1} = 0.18$$

	V1	V2	V3	V4
V1	1	0	0	0,45
V2	0	1	0	0,16
V3	0	0	1	0
V4	0	0	0	1

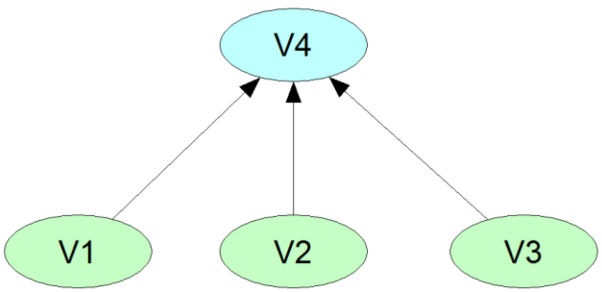


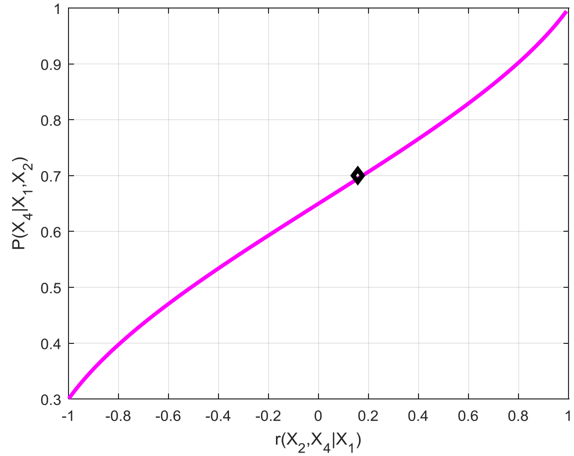


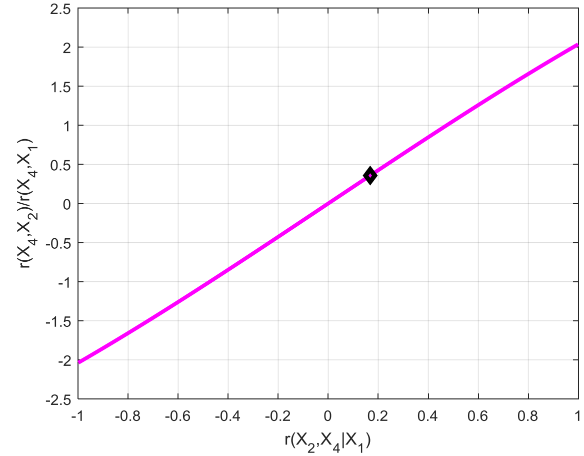
- Cond. Prob. or
- Rat. of Rank corr.
- Are Bounded
- Depends on the corr. Matrix elements
- Positive definite

TU Delft

BN DEPENDENCE QUANTIFICATION





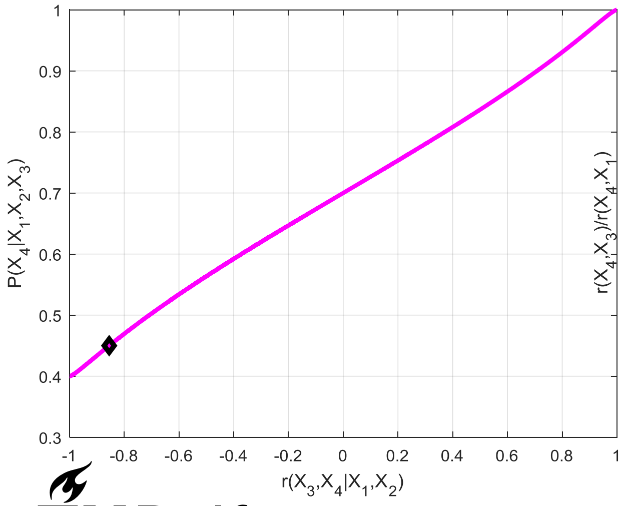


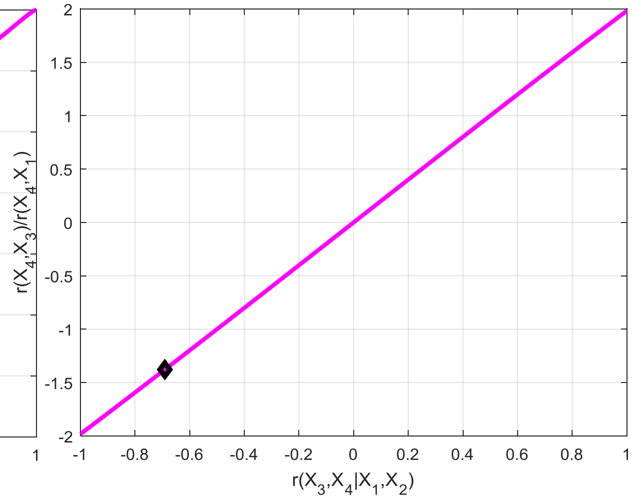
$$P_1^{e_1} = 0.65 \quad \rightarrow \quad r_{4,1}^{e_1} \approx 0.44 \qquad P_1^{e_2} = 0.65 \quad \rightarrow \quad r_{4,1}^{e_2} \approx 0.44$$

$$P_2^{e_1} = 0.70 \quad \rightarrow \quad r_{4,2|1}^{e_1} \approx 0.17 \qquad R_2^{e_2} = 0.36 \quad \rightarrow \quad r_{4,2|1}^{e_1} \approx 0.17$$

$$P_3^{e_1} = 0.50 \quad \rightarrow \quad r_{4,3|2,1}^{e_1} \approx -0.69 \qquad R_2^{e_2} = \text{-}1.38 \quad \rightarrow \quad r_{4,2|1}^{e_1} \approx -0.69$$

	V1	V2	V3	V4
V1	1	0	0	0,45
V2	0	1	0	0,16
V3	0	0	1	-0,62
V4	0	0	-1	1



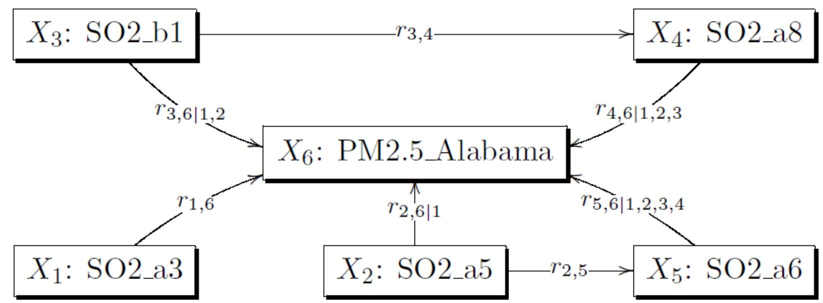


- Bounds for CPE get shorter
- Not so much for RRC
- Which Method gives “better” estimates?
- Can experts estimate dependence?

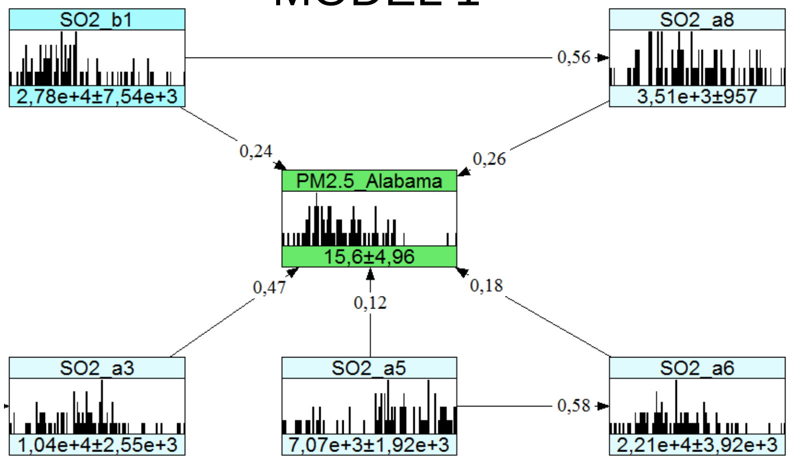
TUDelft

TO ANSWER THE QUESTIONS

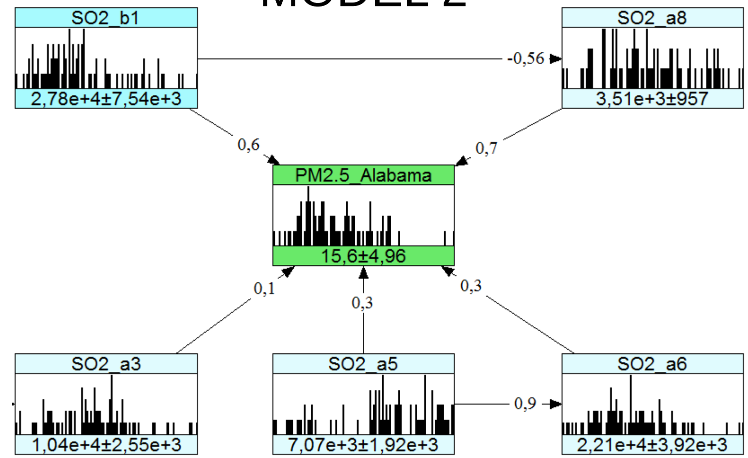
- Let experts quantify models that we know
- Underlying assumptions hold
- SO_2 emissions and $PM_{2.5}$ concentrations
- See how well they can approximate the models
- M1 data M2 fictitious dependence



MODEL 1



MODEL 2



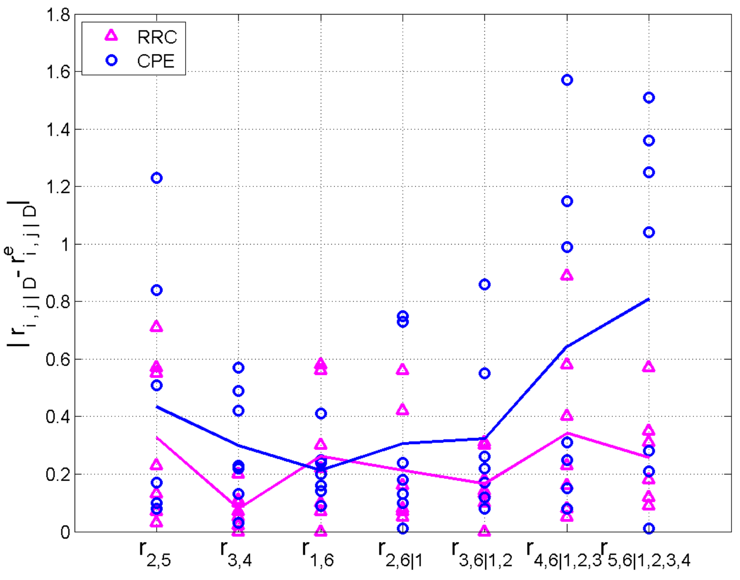
TUDelft

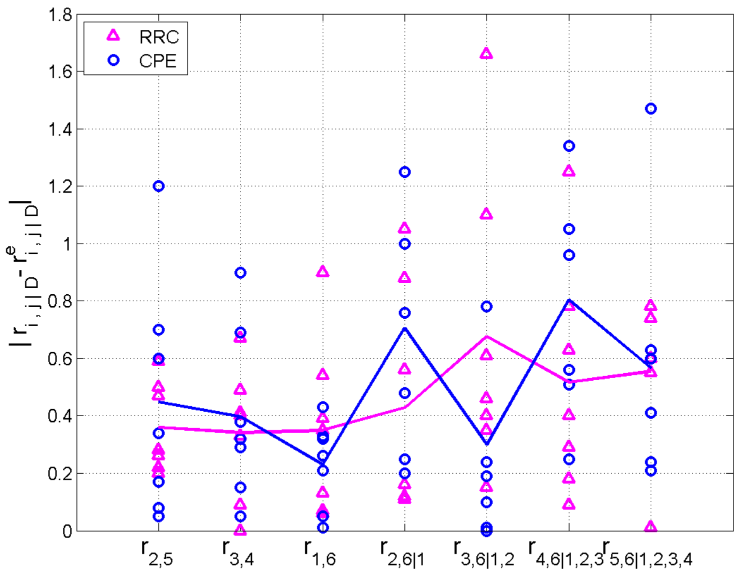
RESULTS (INDIVIDUAL ESTIMATES)

$$\left|r_{i,j|D} - r^e_{i,j|D}\right|$$

$$\bar{\delta}_{M1RRC} = 0.23$$
$$\bar{\delta}_{M1CPE} = 0.43$$
$$\bar{\delta}_{M2RRC} = 0.46$$
$$\bar{\delta}_{M2CPE} = 0.49$$

- Mean difference half of other groups for
- Ratios Rank Correlation with Data
- Diference statistically significant
- Doesn't say if individual experts can approximate the model of interest
- H_0: $BN_e = BN_{true}$
- Initial idea use det of corr. matrices





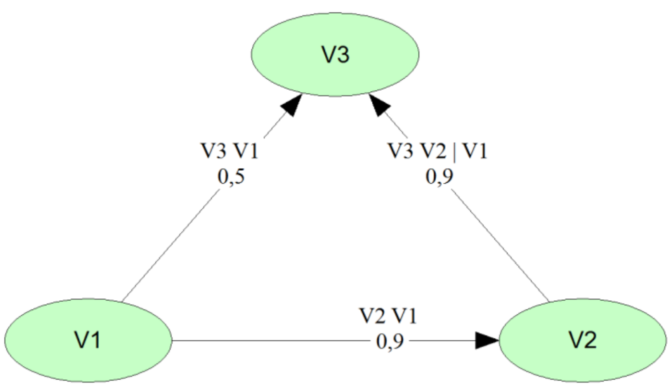
TUDelft

EXAMPLE CORRELATION MATRICES SAME DETERMINANT

- H_0: $BN_e = BN_{true}$
- Sign and magnitude for correlations matter
- Infinitely many CM with same det.

TRUE				
r21 =	0,9			
	0,907981			
r31 =	0,5			
	0,517638			
r32|1 =	0,9			
	0,907981			
	1	0,907981	0,517638	
RT =	0,907981	1	0,795522	
	0,517638	0,795522	1	
DT =	0,022565			

Expert 1			
r21 =	0,9		
	0,907981		
r31 =	0,5		
	0,517638		
r32|1 =	-0,9		
	-0,90798		
	1	0,907980999	0,517638
R1 =	0,907981	1	0,144489
	0,517638	0,144488943	1
D1 =	0,022565		



Expert 2				
r21 =	-0,9			
	-0,90798			
r31 =	-0,9			
	-0,90798			
r32|1 =	-0,5			
	-0,51764			
	1	-0,90798	-0,90798	
R2 =	-0,90798	1	0,733548	
	-0,90798	0,733548	1	
D2 =	0,022565			

Expert 3			
r21 =	0		
	0		
r31 =	0		
	0		
r32|1 =	-0,98751		
	-0,98865		
	1	0	0
R3 =	0	1	-0,98865
	0	-0,98865	1
D3 =	0,022565		

DEPENDENCE CALIBRATION

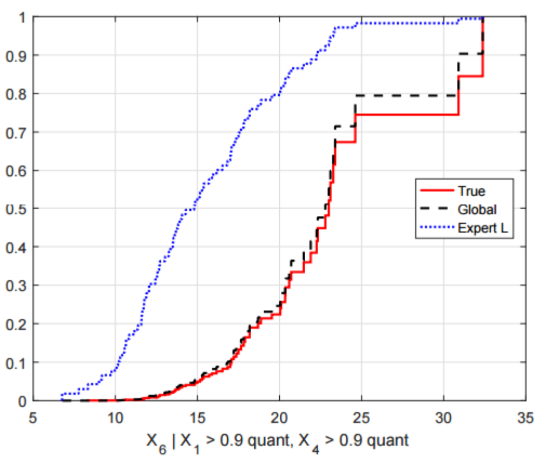
- We have shown:
- dCal = 1 iff $\Sigma_C = \Sigma_E$
- High dCal if calibration CM well approximated element wise
- Low dCal if high (Abs) correlations not well approximated

$$H(f_C, f_E) = \iint_{[0,1]^2} \sqrt{\frac{1}{\sqrt{2}}\left(\sqrt{f_C(u,v)} - \sqrt{f_E(u,v)}\right)^2 \mathrm{dudv}}$$

$$H_G(\Sigma_C, \Sigma_E) = \sqrt{1 - \frac{\det(\Sigma_C)^{1/4}\det(\Sigma_E)^{1/4}}{(1/2\det(\Sigma_C) + 1/2\det(\Sigma_E))^{1/2}}}$$

$$D = 1 - H$$

- Performance in assessing uncertainty and dependence don't correlate perfectly
- We must use other measures of performance for dependence
- Combinations based on dCal outperform individual opinions



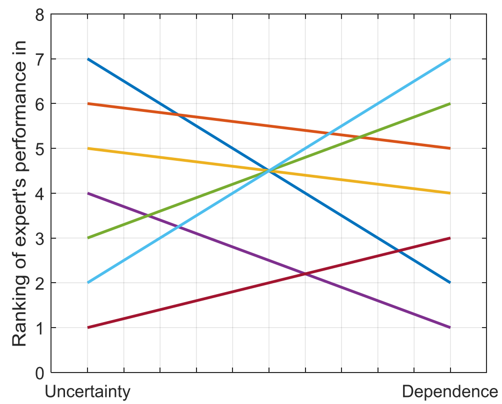


Fig. 1 Ranking of experts performance as uncertainty assessors and dependence assessors.

Group 2				
Id.	Calibr.	Inform.	*dCal* M1RRC	*dCal* M2CPE
D	0.0357	2.745	0.71	0.60
E	0.0063	1.497	0.51	0.32
F	0.7069	0.7571	0.12	0.09
G	l.o.	1.86	0.87	0.49
J	l.o.	2.49	0.32	0.17
L	0.0028	1.169	0.09	0.16
M	0.00131	3.84	0.75	0.32
Eq.	0.5503	0.3009	0.66	0.37
Gl.	0.7069	0.7571	0.95	0.60

Table 3: Calibration, Information and d-Calibration scores for air pollution NPBN experts.

WHAT IF CORRELATION IS NOT SUFFICIENT?

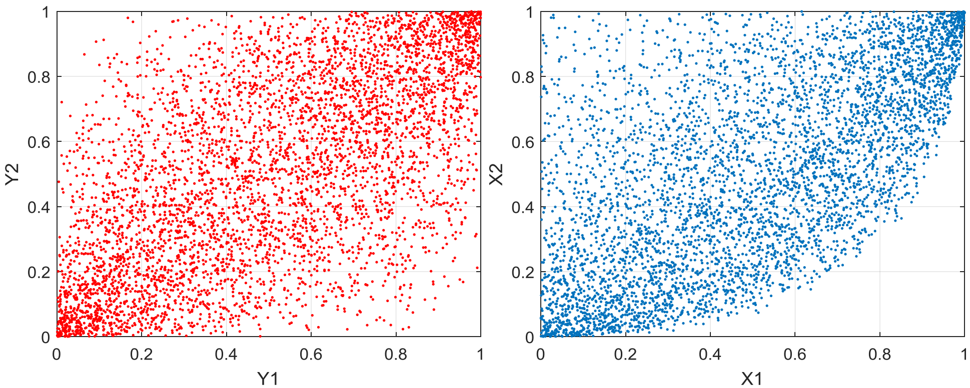
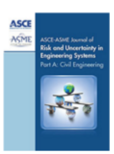


Fig. 3 Samples from a bivariate Gaussian copula (Y1, Y2) and a copula with asymmetries (X1, X2).

$$H(f_C, f_E) = \iint_{[0,1]^2} \sqrt{\frac{1}{\sqrt{2}}\left(\sqrt{f_C(u,v)} - \sqrt{f_E(u,v)}\right)^2}\, dudv$$

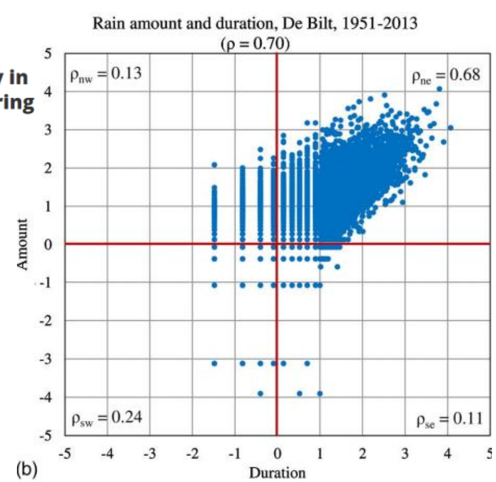
Characterization of Precipitation through Copulas and Expert Judgement for Risk Assessment of Infrastructure



ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering

ISSN (online): 2376-7642

Frequency: Quarterly | Sponsored by the Committee on Technical Advancement



- Started using technique for problems where asymmetries exist.
- Can exp. identify asymmetries?
- Evaluate expert's ability to recognize asymmetries?
- Combine exp. (disagree) w.r.t. possible asymmetries?
- (Dis)agreement grows with increased dimensionality?
- Complexity in combination grows with increased complexity?
- Similar conclusions

TU Delft

CONCLUSIONS

- EXPERTS CAN ESTIMATE DEPENDENCE
- DEPENDENCE ELICITATION IS ESSENTIALLY DIFFERENT THAN UNCERTAINTY ELICITATION
- USUAL MEASURES FOR UNCERTAINTY ELICITATION ARE NOT SUFFICIENT.
- EVALUATE EXPERT PERFORMANCE WITH APPROPRIATE MEASURES
- COMBINATIONS BASED ON DCAL OUTPERFORM INDIVIDUAL OPINIONS
- STILL MUCH TO BE DONE