

Invitation

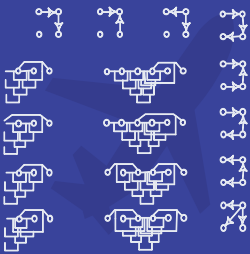
You are welcome to attend
the public defense of my Ph.D.
thesis:

Bayesian Belief Nets and Vines
in Aviation Safety and other
Applications.

It will take place on Monday
February 15, 2010 at 10.00 hrs
in the Senaatszaal of the Aula
of the Delft University of
Technology, Mekelweg 5,
Delft.

Prior to the defence, at 09.30
there will be a 20 minutes
presentation.

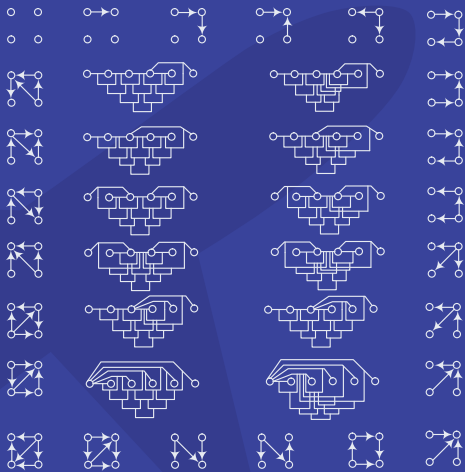
The defence will be followed
by a reception in the Aula.



Bayesian belief nets and vines in aviation safety and other applications

O. Morales Nápoles

Bayesian Belief Nets and Vines
in aviation safety and other applications



Oswaldo Morales Nápoles

PROPOSITIONS
accompanying the thesis

**Bayesian Belief Nets and Vines in Aviation Safety and other
Applications.**

Oswaldo Morales Nápoles.

- (1) The algorithm proposed in Aas et al. [2009, p.189] suggests that operations to assign the ‘best’ *regular vine* to a data set should begin by selecting the first tree and iteratively selecting subsequent trees of the *regular vine*. Algorithm 2.3.2. in this thesis (Morales Nápoles [2009]) is the most advantageous in this case if regular vines are to be generated “on the fly”.
- (2) The notion of *natural order* for regular vines presented in this thesis has helped to enumerate the number of *equivalence classes* of regular vines on n nodes [Joe et al., 2010]. This same notion should shed light on other graphical properties of regular vines such as the number of tree-equivalent classes, the degree sequence, the typical distance and the diameter of each tree in every level of the vine. Moreover, these properties should be further studied.
- (3) The subject of random vines should be developed. One example is the *random vine* $V(n)$ taken at random from the collection of all vines on n nodes. A second example, more important for present applications [Kurowicka and Cooke, 2006], could be the *random regular vine* $RV(n, p)$ on n nodes where each of the $n(n - 1)/2$ edges realizes independently a partial correlation equal to zero with probability p . Moreover, p could be related to the ratio of the number of labeled elements in each equivalent or tree-equivalent class to the total number of labeled regular vines on n nodes.
- (4) The construction of a large data base with all possible *labeled regular vines* on n nodes for a sufficiently large n that includes all known graphical properties is, at this stage, as important for applications as algorithms for generating them.
- (5) The number of stars in the Milky Way as computed from Schneider [2006, p.5] is approximately equal to the number of *labeled vines* on 7 nodes.
- (6) According to the CATS BBN presented in this thesis (Morales Nápoles [2009]), it is possible to reduce the accident rate to a value approximately 6 times smaller than the current value. In order to do this, the flight crew should increase their levels of experience to a value equal to the 97th percentile of their distributions. The aviation sector cannot support this change in the short run. In order to achieve a reduction in the accident rate of similar magnitude, a combination of policies is required. Investigating these possible policies should be at least one of the immediate goals of the CATS BBN.

- (7) *Non-Parametric Continuous-Discrete BBNs* are more flexible with respect to changes in modelling than discrete, Gaussian and discrete-Gaussian BBNs. However, adding or deleting nodes or arcs in the graph could still lead to a re-quantification.
- (8) The assessment of conditional rank correlations through ratios of unconditional rank correlations is easier for experts than its assessment through probabilities of exceedence with a large number of conditions.
- (9) The consequences of an earth dam failure in the State of Mexico will be approximately constant regardless of the size of the failure.
- (10) When the univariate marginal distributions are very different across experts, the joint distribution obtained with the method for combination described in this thesis (Morales Nápoles [2009, Ch.4]) with equal weights tends to suppress the magnitude of the dependence even if individual experts think bivariate rank correlations are of the same sign and magnitude. This is an advantage of combining rank correlations through exceedence probabilities.
- (11) The number of hairs in a cow's tail is a random variable with mean 2,872 (Fauvel and Gerdes [1990]). Efforts to adequately characterize this quantity would lead to a better world.

These propositions are considered opposable and defensible and as such have been approved by the supervisor prof. dr. R.M. Cooke.

REFERENCES

- Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182 – 198, 2009.
- O. Morales Nápoles. *Bayesian belief nets in aviation safety and other applications*. PhD thesis, Delft University of Technology, 2009.
- Harry Joe, Roger M. Cooke, and Dorota Kurowicka. Regular vines: Generation algorithm and number of equivalent classes. In Dorota Kurowicka and Harry Joe, editors, *Dependence Modeling-Handbook on Vine Copulae*, Dependence Modeling, Scheduled Fall 2010.
- Dorota Kurowicka and R.M. Cooke. Completion problem with partial correlation vines. *Linear Algebra and Its Applications*, 418(1):188–200, 2006.
- Peter Schneider. *Extragalactic Astronomy and Cosmology An Introduction*. Springer, 2006. ISBN 10 3-540-33174-3.
- J. Fauvel and P. Gerdes. African slave and calculating prodigy: Bicentenary of the death of thomas fuller. *Historia Mathematica*, (17):141–151, 1990.

STELLINGEN
behorende bij het proefschrift

**Bayesiaanse Netwerken en Vines in veiligheid van de luchtvaart en
andere toepassingen.**

Oswaldo Morales Nápoles.

- (1) Het algoritme voorgesteld in Aas et al. [2009, p.189] suggereert dat activiteiten, om de “beste” (regular vine) aan een reeks gegevens toe te wijzen, moeten beginnen met het selecteren van de eerste boom en daarna iteratief selecteren van latere bomen van de (regular vine). Algoritme 2.3.2. in dit proefschrift (Morales Nápoles [2009]) is de meest voordelige in dit geval, waarin reguliere takken per direct gegenereerd worden.
- (2) Het begrip dat regular vines aan een *natuurlijke orde* voldoen, zoals gepresenteerd in dit proefschrift, heeft bijgedragen aan de mogelijkheid om het aantal equivalentie klassen van reguliere takken op n knopen op te sommen [Joe et al., 2010]. Dit begrip zou tevens licht moeten werpen op andere grafische eigenschappen van reguliere takken, zoals het aantal boom-gelijkwaardige klassen, de graad volgorde, de typische afstand en de diameter van elke boom in elk niveau van de vine. Bovendien moeten deze eigenschappen verder worden onderzocht.
- (3) Het onderwerp van willekeurige vines moet verder worden ontwikkeld. Een voorbeeld hiervan is de *willekeurige vine* $V(n)$ die willekeurig uit de verzameling van alle vines op n knooppunten getrokken wordt. Een tweede voorbeeld, dat belangrijker is voor de huidige toepassingen [Kurowicz and Cooke, 2006], zou de *willekeurige reguliere vine* $RV(n, p)$ op n knopen kunnen zijn, waarbij elk van de $n(n - 1)/2$ randen onafhankelijk een partiële correlatie gelijk aan nul met kans p realiseert. Bovendien zou p gerelateerd kunnen worden aan de verhouding tussen het aantal gelabelde elementen in elke equivalente klasse of boom-equivalente klasse en het totale aantal gelabelde regular vines op n knopen.
- (4) Het maken van een grote databank, met alle mogelijke gelabelde *regular vines* op n knooppunten voor een voldoende grote n , dat alle bekende grafische eigenschappen bevat, is, in dit stadium, van hetzelfde belang voor toepassingen als algoritmen zijn voor het genereren van vines.
- (5) Het aantal sterren in de Melkweg zoals berekend van Schneider [2006, p.5] is ongeveer gelijk aan het aantal *gelabelde vines* op 7 knooppunten.
- (6) Volgens de CATS BBN gepresenteerd in dit proefschrift (Morales Nápoles [2009]), is het mogelijk om het aantal ongevallen te verminderen tot een waarde die ongeveer 6 maal kleiner is dan het huidige aantal. Om dit te doen, moet het cockpitpersoneel hun ervaringsniveau opvoeren naar een niveau gelijk aan het 97ste percentiel van hun verdelingen. De luchtvaartsector kan op korte termijn niet voldoen aan deze wijziging. Om een gelijkwaardig grote reductie in het aantal ongevallen te verkrijgen is een combinatie van maatregelen nodig. Het onderzoeken van deze mogelijke

maatregelen zou ten minste n van de doelen van de CATS BBN moeten zijn.

- (7) *Niet-parametrische continue-discrete BBN* zijn flexibeler met betrekking tot veranderingen in modellering dan discrete, Gaussische en discrete-Gaussische BBNs. Echter, het toevoegen of verwijderen van knooppunten of bogen in de grafiek kan nog leiden tot een nieuwe kwantificering.
- (8) De beoordeling van de voorwaardelijke rang correlaties door middel van verhoudingen van onvoorwaardelijke rang correlaties is eenvoudiger voor experts dan het beoordelen met behulp van kansen op overschrijding met een groot aantal voorwaarden.
- (9) De gevolgen van het falen van een dijk in de Staat van Mexico zal ongeveer constant zijn, onafhankelijk van de omvang van het falen.
- (10) Wanneer de één-dimensionale verdelingen zeer verschillend zijn over verscheidene experts, zal de gezamenlijke verdeling, verkregen met de methode voor de combinatie beschreven in dit proefschrift (Morales Nápoles [2009, Ch.4]), met gelijke gewichten de neiging hebben om de grootte van afhankelijkheid te onderdrukken. Zelfs als individuele experts denken dat rang twee dimensionale correlaties van hetzelfde teken en dezelfde omvang zijn. Dit is een voordeel van het combineren van rang correlaties door overschrijdende kansen.
- (11) Het aantal haren in de staart van een koe is een willekeurige variabele met een gemiddelde van 2.872 (Fauvel and Gerdes [1990]). Inspanningen om deze hoeveelheid adequaat te kwantificeren zou een betere wereld tot gevolg hebben.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof. dr. R.M. Cooke.

REFERENCES

- Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182 – 198, 2009.
- O. Morales Nápoles. *Bayesian belief nets in aviation safety and other applications*. PhD thesis, Delft University of Technology, 2009.
- Harry Joe, Roger M. Cooke, and Dorota Kurowicka. Regular vines: Generation algorithm and number of equivalent classes. In Dorota Kurowicka and Harry Joe, editors, *Dependence Modeling-Handbook on Vine Copulae*, Dependence Modeling, Scheduled Fall 2010.
- Dorota Kurowicka and R.M. Cooke. Completion problem with partial correlation vines. *Linear Algebra and Its Applications*, 418(1):188–200, 2006.
- Peter Schneider. *Extragalactic Astronomy and Cosmology An Introduction*. Springer, 2006. ISBN 10 3-540-33174-3.
- J. Fauvel and P. Gerdes. African slave and calculating prodigy: Bicentenary of the death of thomas fuller. *Historia Mathematica*, (17):141–151, 1990.

BAYESIAN BELIEF NETS AND VINES IN AVIATION SAFETY AND OTHER APPLICATIONS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 15 februari 2010 om
10:00 uur

door

Oswaldo MORALES NÁPOLES

Master of Science in Applied Mathematics
geboren te Toluca, México.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. R.M. Cooke

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. R.M. Cooke	Technische Universiteit Delft, promotor
Dr. D. Kurowicka	Technische Universiteit Delft, copromotor
Prof. dr. B.J.M. Ale	Technische Universiteit Delft
Prof. dr. H. Joe	University of British Columbia, Vancouver
Dr. D. de León Escobedo	Universidad Autónoma del Estado de México, Toluca
Prof. dr. A. Mosleh	University of Maryland, Maryland
Prof. dr. M.J.L. van Tooren	Technische Universiteit Delft
Prof. dr. G. Jongbloed	Technische Universiteit Delft, reservelid

ISBN

Copyright © 2010 by O. Morales Nápoles.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

Cover designed by Sandra Gaytan, Diana Garcia & the author.

Printed in the Netherlands by Wöhrmann Print Service.

Contents

1	Introduction	1
1.1	Probability & uncertainty	1
1.2	Copulae	2
1.2.1	Dependence Measures	2
1.2.2	Two examples of copulae	3
1.3	Overview of the development of vines & BBNs	5
1.3.1	Graph Theory	5
1.3.2	Bayesian belief networks and influence diagrams	6
1.3.3	Undirected Graphs and Vines	8
1.4	Introduction to the Causal Model for Air transport Safety	9
2	About The Number of Vines and Regular Vines on n Nodes.	17
2.1	Introduction	17
2.2	Trees	18
2.2.1	The Number of Labeled Trees on n Nodes and the Prüfer Code	18
2.3	Vines	20
2.3.1	The Number of Vines on n Nodes and the Prüfer Code	20
2.3.2	Regular vines and the line graph	22
2.4	The Number of Regular Vines on n Nodes.	26
2.5	Final Comments	33
3	BBNs in Aviation Safety	35
3.1	Discrete BBNs	35
3.2	Non-Parametric Continuous-Discrete BBNs.	37
3.3	Causal Model for Air transport Safety (CATS)	40
3.3.1	Event Sequence Diagrams & Fault Trees	40
3.3.2	Human Reliability Models	42
3.3.2.1	FCP Model Description	42

3.3.2.2	ATCP Model Description.	44
3.3.2.3	MNTP Model Description.	47
3.3.3	The CATS Model in UNINET	47
3.3.3.1	ESDs & FTs for the CATS model in UNINET	47
3.3.3.2	The error distributions in UNINET	47
3.3.3.3	The complete model	51
3.4	Model Use	56
4	Elicitation and Combination of Dependence.	61
4.1	Introduction	61
4.2	Structured Expert Judgment	62
4.2.1	The Classical Model for Structured Expert Judgment	63
4.2.2	Dependence Elicitation	64
4.2.2.1	Probabilistic Approaches	64
4.2.2.2	Direct Approach	70
4.2.3	Combination of Experts' Dependence Estimates	74
4.3	Final Comments	77
5	Structured Expert Judgment in Aviation Safety	79
5.1	The Missed Approach Model	79
5.1.1	Introduction to the MA model.	79
5.1.2	Description of the MA model.	80
5.1.3	Expert Elicitation Results of the MA Model	82
5.1.4	Updating beliefs in the MA Model	84
5.2	The Flight Crew Performance Model	87
5.2.1	Expert Elicitation Results of the FCP Model	87
5.2.2	Dependence in the FCP Model	89
5.3	The Air Traffic Control Performance Model	90
5.3.1	Expert Elicitation Results of the ATCP Model	90
5.3.2	Dependence in the ATCP Model	91
6	Dams Safety in the State of Mexico	93
6.1	Introduction	93
6.2	Earth Dams in the State of Mexico	95
6.3	Description of the DS model.	96
6.3.1	Model variables & graph	96
6.3.2	Expert Elicitation Results of the DS Model	98
6.3.3	Dependence in the DS Model	100
6.4	Discussion of the DS Model	103
6.5	Final comments of the DS Model	105
7	Conclusions	107
7.1	About Vines	107
7.2	About Bayesian Networks and their Applications	109
7.2.1	Aviation Safety	109
7.2.2	Earth Dams Safety	111
7.2.3	About BBNs.	111

References	113
------------	-----

Appendices

A Regular Vines Catalogue.	121
----------------------------	-----

Summary	155
---------	-----

Samenvatting	157
--------------	-----

Acknowledgments	159
-----------------	-----

Curriculum Vitae	161
------------------	-----

CHAPTER 1

Introduction

1.1 Probability & uncertainty

This thesis explores some properties of graphs and their relation to probability distributions in order to show their use in current applications in risk and uncertainty analysis.

According to David [1955] games of chance were invented some time between 40,000 years ago and the third millennium before Christ. About 960 A.D. the earliest work which mentions the number of ways in which three dice thrown together (or one dice thrown three times) may fall, irrespective of order, is presented by a certain bishop Wibold. In his presentation no attempt to assess relative probabilities may be visualized [Kendall, 1956].

Games of chances were studied by mathematicians such as Cardano, Galileo Galilei, Pascal, Fermat and Huygens until the middle of the 17th century. Concepts such as fair coins, honest dice, equal case of occurrence, equal conditions and others were in the minds of scientists at the time. But it is not until the works of De Moivre and Jacob Bernoulli that a more modern version of the theory is encountered. It appears to be the latter the first who thought of applying the doctrine of chances to the art of conjecture [Kendall, 1956]. Applications of probability theory to the actuarial sciences begin also with the works of Halley and later Montmort and Nicholas Bernoulli [Sheynin, 1968].

From the 1700s on, the development of probability theory and its applications in many fields has progressed rapidly. In particular this thesis is interested in the description of applications of Bayesian belief networks (BBNs) and vines to specific problems in which quantifying uncertainty is of prime importance. BBNs and vines are graphical models used to represent multivariate probability distributions. BBNs will find their application in this thesis in the identification and measurement of risks in the aviation industry and earth dams.

1.2 Copulae

Representing multivariate probability distributions for certain phenomena can be a challenging task. Perhaps the multivariate model which is most widely used is the joint normal distribution. However, many phenomena behave far from normal. This is one of the reasons for researchers to have recourse to alternative models such as copulae.

The use of copulae can be traced back to the 1940s in the work of Hoeffding and the 1950s with the work of Fréchet and Sklar [Nelsen, 1998, p.2]. Copulae are multivariate distributions with uniform margins on $(0, 1)$. This suggests immediately the possibility of inducing a certain dependence structure to given one dimensional margins. Its possibilities for applications in statistics and simulation become evident and today many references can be found for them.

Copulae are part of the building blocks of the graphical models to be used in this thesis and for that reason basic concepts and definitions regarding them are introduced. The book by Nelsen [1998] presents an introduction to the subject. A larger account of the ideas briefly discussed in this section may be found in Kurowicka and Cooke [2006].

Bivariate copulae will be of special interest for us. By copula (or copulae) we mean a bivariate copula (or bivariate copulae) unless otherwise specified. The *bivariate copula* or simply the *copula* of two continuous random variables X and Y is the function C such that their joint distribution can be written as:

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$

Copulae are functions that allow naturally the investigation of association between random variables. Measures of association such as the rank correlation or Kendall's tau may be expressed in terms of copulae [Nelsen, 1998]. The measures of association to be used in this thesis are described next.

1.2.1 Dependence Measures

In this section we briefly present basic concepts and definitions about the measures of association used later on in the thesis. The *product moment correlation* of random variables X and Y with finite expectations $E(X)$, $E(Y)$ and finite variances $var(X)$, $var(Y)$ is:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)var(Y)}}.$$

The *rank correlation* of random variables X , Y with cumulative distribution functions F_X and F_Y is:

$$r_{X,Y} = \rho_{F_X(X), F_Y(Y)} = \frac{E(F_X(X)F_Y(Y)) - E(F_X(X))E(F_Y(Y))}{\sqrt{var(F_X(X))var(F_Y(Y))}}.$$

The rank correlation is the product moment correlation of the ranks of variables X and Y , and measures strength of monotonic relationship between variables. The conditional rank correlation of X and Y given Z is:

$$r_{X,Y|Z} = r_{\tilde{X},\tilde{Y}}$$

where (\tilde{X}, \tilde{Y}) has the distribution of (X, Y) given $Z = z$.

The (conditional) rank correlation is the dependence measure of interest because of its close relationship with conditional copulae used in vines (chapter 2) and non-parametric continuous-discrete BBNs (chapter 3). One disadvantage of this measure however is that it fails to capture non-monotonic dependencies.

Rank correlations may be realized by copulae, hence the importance of these functions in dependence modeling. Partial correlations will also be of interest in this thesis. These can be defined in terms of partial regression coefficients. Consider variables X_i with mean zero and standard deviation σ_i for $i = 1, \dots, n$ and let the numbers $b_{1,2,3,\dots,n}, \dots, b_{1,n;2,\dots,n-1}$ minimize:

$$E[(X_1 - b_{1,2,3,\dots,n}X_2 - \dots - b_{1,n;2,\dots,n-1}X_n)^2]$$

The *partial correlation* of X_1 and X_2 based on X_3, \dots, X_n is:

$$\rho_{1,2;3,\dots,n} = \text{sgn}(b_{1,2;3,\dots,n})(b_{1,2;3,\dots,n}b_{2,1;3,\dots,n})^{1/2}$$

Partial correlations can be computed recursively from correlations [Yule and Kendall, 1965.]:

$$\rho_{1,2;3,\dots,n} = \frac{\rho_{1,2;4,\dots,n} - \rho_{1,3;4,\dots,n} \cdot \rho_{2,3;4,\dots,n}}{((1 - \rho_{1,3;4,\dots,n}^2) \cdot (1 - \rho_{2,3;4,\dots,n}^2))^{1/2}} \quad (1.1)$$

Next two examples of copulae that will appear later in this thesis are presented.

1.2.2 Two examples of copulae

A unique copula that corresponds to any given continuous joint distribution may always be found. In this section two such copulae will be presented. Denote by Φ_ρ the bivariate standard normal cumulative distribution function with correlation ρ and Φ^{-1} the inverse of the univariate standard normal distribution function then

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)); (u, v) \in [0, 1]^2$$

is called the *normal copula*. Notice that ρ is a parameter of the normal copula. The relationship between the correlation of the normal copula r (the rank correlation of the normal variables) and the parameter ρ (the product moment correlation of the normal variables) is known and given by the following formula [Kurowicka and Cooke, 2005, p.55]:

$$\rho = 2 \sin\left(\frac{\pi}{6}r\right). \quad (1.2)$$

In this thesis rank and conditional rank correlations will be of special importance. In general partial correlation is not equal to conditional correlation,

however, for the joint normal distribution the partial and conditional correlations are equal.

A second example of a copula that will be used in this thesis is Frank's copula. Frank's copula [Frank, 1979] is an Archimedean copula that has closed form for the density, conditional and inverse conditional distribution. Additionally, it has the property of reflection symmetry [Kurowicka and Cooke, 2005, p.49]. *Frank's copula* is:

$$C_{\theta}(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]; (u, v) \in [0, 1]^2 \quad (1.3)$$

The parameter θ in equation (1.3) may be expressed in terms of rank correlation. For the Normal copula zero correlation entails independence. For Frank's copula the limit $\theta \rightarrow 0$ yields $C_{\theta}(u, v) = u \cdot v$. The property that zero correlation implies independence is called the *zero independence property* and is of special importance for continuous non-parametric BBNs [Hanea et al., 2006].

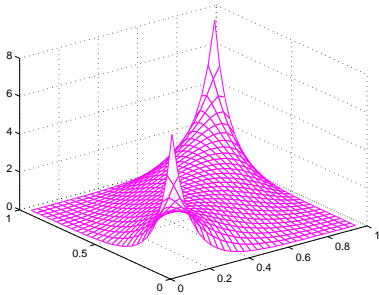


Figure 1.1: *Density of the normal copula with rank correlation 0.7859*

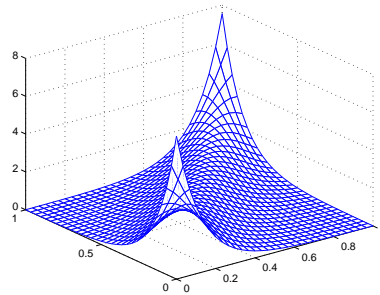


Figure 1.2: *Density of Frank's copula with rank correlation 0.7859*

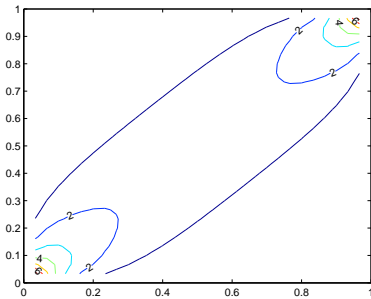


Figure 1.3: *Contour plot of Figure 1.1*

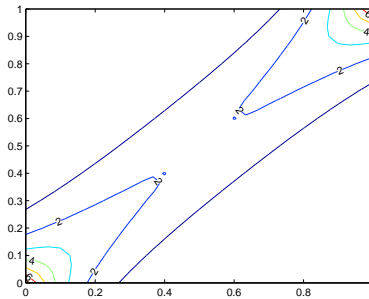


Figure 1.4: *Contour plot of Figure 1.2*

The densities of the normal and Frank's copulae with correlation 0.7859 are presented in Figures 1.1 and 1.2 respectively. At first sight the densities seem

to be almost identical. The differences between the two densities become more evident in Figures 1.3 and 1.4 where contour plots are shown for each of them.

The advantages of using the normal copula for the methods proposed in this thesis will become evident in chapters 3 and 4. The differences that may arise by the choice of a copula in the modeling process will be illustrated comparing these two copulae. Copulae are used in the graphical models discussed in this thesis (vines and BBNs) to construct multidimensional probability distributions. Before discussing vines and BBNs more formally a small overview of the development of such models will be presented.

1.3 Overview of the development of vines & BBNs

1.3.1 Graph Theory

As previously mentioned, vines and BBNs combine graph theory with probability theory. For that reason we begin this thesis with an overview of graph theory. Some concepts and definitions additional to those presented in section 1.2 will be required. They will be presented in this section.

Because of his discussion of a famous problem called the Königsberg bridge problem, Leonhard Euler is acknowledged as the father of graph theory. This problem appears in almost any modern text book on graph theory. Euler's original paper is written in Latin, for a translation to English the reader is referred to [Biggs et al., 1986]. Like many problems in probability theory, some of the early developments of graph theory originated from games. One of these was the hamiltonian game invented by Sir William Hamilton. The hamiltonian game will be used to introduce some definitions and notation that will be used later in the rest of the thesis.

An *undirected graph* $G = (N, E)$ consists of a finite non empty set N of nodes, also called (points or vertices) and a possibly empty set E of edges (lines or arcs) where each element is an unordered pair (α_1, α_2) , where α_1 and $\alpha_2 \neq \alpha_1$ are elements of N . Without loss of generality in this thesis when $N = \{1, 2, \dots, n\}$ we speak of *labeled graphs*. It will be assumed that two distinct edges do not join the same pair of nodes; graphs in which this is allowed are called *multigraphs*. Observe that no self-loops are permitted that is, edges joining nodes with itself. If the pair (α_1, α_2) is ordered then G is a *directed graph* and the pair (α_1, α_2) will be represented as $\alpha_1 \rightarrow \alpha_2$. In this case α_1 will be called a *parent* node of the *child* node α_2 . Examples of undirected and directed graphs are shown in Figures 1.5 and 1.6 respectively.

The cardinality of N is called the *order* of the graph. If the pair $(\alpha_1, \alpha_2) \in E$ then the two nodes α_1 and α_2 are *adjacent* and each one is *incident* with the pair $(\alpha_1, \alpha_2) \in E$. The *degree* of a node is the number of edges incident with it. A *complete graph* CG has every node adjacent to each other. A *path* of length n from α to β is a sequence $\alpha = \alpha_0, \dots, \alpha_n = \beta$ of distinct nodes such that $(\alpha_{i-1}, \alpha_i) \in E$ for all $i = 1, \dots, n$. A *cycle* is a path such that $\alpha = \beta$. If every pair (α_{i-1}, α_i) in a cycle of a directed graph is ordered as in E then it is a *directed cycle* otherwise it is an *undirected cycle*. If a directed graph has no directed cycles, then it is a

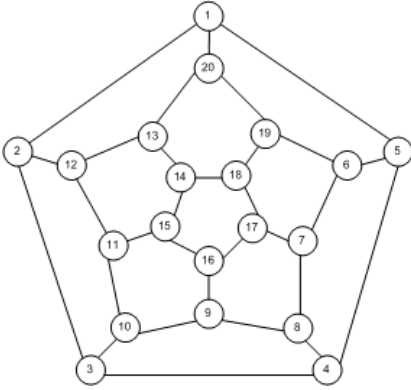


Figure 1.5: Undirected graph of order 20

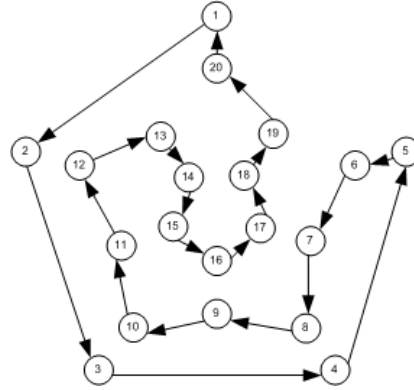


Figure 1.6: Directed graph of order 20 with a cycle.

directed acyclic graph.

Hamilton proposed a graph like the one in Figure 1.5 where each node represented a city of the world and the edges connections between the cities. The object of the game was to travel “Around the World” by finding a route that passes through each node exactly once. In other words, the object of the game was to find a cycle of the graph in Figure 1.5 such that all nodes in N are contained in the cycle. One possible such cycle is represented by the directed graph in Figure 1.6. According to Harary [1967, p.5] “Hamilton sold this idea to a game manufacturer in Dublin for about twenty-five guineas which was wise of him since it was not a commercial success”¹.

Since the introduction of graphs by Euler, its applications to many fields of science has grown. Probability theory has also relied on graphs to advance its methods. Two types of graphs will be of special importance in this thesis: directed acyclic graphs and trees. The presentation continues with a short overview of the development of vines and BBNs.

1.3.2 Bayesian belief networks and influence diagrams

Most of the literature on graphical models reflects the idea of using graphical representations for probabilistic information can be traced to the work of Sewal Wright in the 1920s (See for example Pearl [1988, p.131] and Cowell et al. [1999, p.81]). Figure 1.7 taken from Wright [1921] shows the diagram that Wright used in his guinea pigs example for introducing his method of *path coefficients*. Wright thought of the boxes in Figure 1.7 as variables that are correlated with each other. He thought it was convenient to use a diagram such as the one in Figure 1.7 to represent relations in which the paths of influence between variables are shown by arrows. The sign of the correlations between variables is shown in the arcs of the network. This kind of diagrams have a close relationship to those that will

¹One guinea in Victorian Britain was equivalent with 26.25 pounds.

be used later in this thesis. Wright's method was criticized by Neils [1922] and perhaps that critique contributed to delay the development of graphical models in probability theory [Pearl, 1988, p.131]².

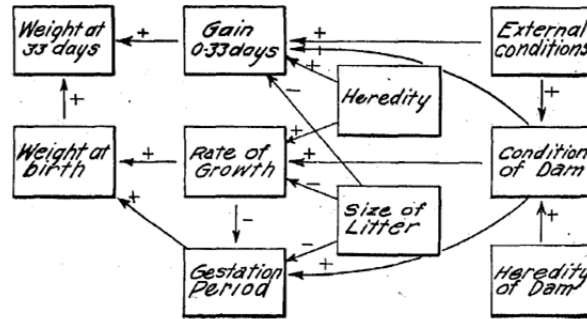


Figure 1.7: Wright's diagram showing the interrelation among the factors determining the weight of guinea pigs at birth and at 33 days.

The use of directed acyclic graphs in combination with probability theory appears to be parallel in decision analysis and artificial intelligence in the late 70s and early 80s [Pearl, 1993]. In Pearl [1982] inference nets where nodes represent discrete variables and arcs conditional probabilities of the child variable given the parents are introduced³. These were extended by Kim and Pearl [1983] and later, in Pearl [1985] a more formal concept of Bayesian Networks is introduced that would lead finally to its formalization in Pearl [1986] and Pearl [1988]. In fact, Pearl [1986, p.246] states that the names "*belief networks, Bayesian networks or influence networks* [will be used] interchangeably, the former two to emphasize the judgmental origin and the probabilistic nature of the quantifiers, the later to reflect the directionality of the links. When the nature of the interaction is perceived to be causal, then the term, *causal network* may also be appropriate". Future developments of discrete Bayesian networks where in the direction of techniques for network updating of which probably the one by Lauritzen and Spiegelhalter [1988] is the most used to date. This technique has been improved however over the years [Cowell et al., 1999, p.123] and [Pearl, 1993, p.55].

Influence diagrams were introduced by Howard and Matheson [1984/2005] as an attempt to form a bridge between qualitative description and quantitative

²According to [Neils, 1922, p.262] there were three fallacies that vitiate the theory: "(1) the assumption that a correct system of the action of the variables upon each other can be set up from *a priori* knowledge; (2) the idea that causation implies an inherently necessary connection between things, or that in some other way it differs from correlation; (3) the necessity of breaking off the chain of causes at some comparatively near finite point."

³Pearl restricts the analysis to "trees" though he recognizes that the model may be generalized to include multiple parents keeping in mind that the states of each variable in the tree may represent the power set of multi-parent groups in the corresponding graph. In the same paper, in a footnote Pearl acknowledges Bayes's essay [Barnard and Bayes, 1958] as the beginning of the science of inductive reasoning. Next year Stigler [1983] makes a suggestion that Bayes may not have been the originator of the theorem named after him.

specification. According to Boutlier [2005], these had their share of influence in artificial intelligence. Pearl [2005] views influence diagrams as informal precursors of belief networks. They might have had a larger impact in representing joint distributions with continuous variables as observed in Pearl [1988] and Schachter and Kenley [1989]. Networks with continuous nodes were restricted to this point to variables with joint normal distributions or discretizing continuous nodes to a finite number of states. Mixed discrete-gaussian models were also made available later Cowell et al. [1999].

Bayesian belief networks⁴ bear the name of the celebre reverend Thomas Bayes⁵ however Bayesian networks as such were not a subject of discussion in his work. Because of his essay [Barnard and Bayes, 1958] he is acknowledged as one of the mayor exponents of the philosophy of induction⁶. Thus the name Bayesian belief networks emphasizes the continuous use of Bayes's rule and inverse probability in the philosophy behind these objects⁷. A recent work by Hanea [2008] compares basic characteristics of these models (discrete BBNs, Gaussian and Discrete-Gaussian BBNs and non-parametric BBNs) and hence that will not be done in the present work. However some concepts and definitions will be repeated for completeness.

1.3.3 Undirected Graphs and Vines

Vines are undirected graphs that specify a multivariate joint distribution. According to Cowell et al. [1999, p.81] undirected models can be traced back to the work of Bartlett [1935] in contingency tables. However the use of undirected graphical models to represent multivariate interactions were formally introduced in Darroch et al. [1980] for discrete variables specified by multidimensional contingency tables. In Speed and Kiiveri [1986] the continuous counterpart is presented for jointly Gaussian random variables. These references make use of undirected graphs to specify conditional independence, however we shall not deal with these kind of models in this thesis⁸.

A more direct ancestor of vines may be found in trees (see section 2.2). Trees were used by Darroch et al. [1980] and Speed and Kiiveri [1986] as special cases of graphical models, however undirected graphs with cycles were also used. Trees were also used in Chow and Liu [1968] to infer discrete distributions⁹ from data. The direct parents of vines are however Markov or Dependence Trees [Meeuwissen, 1993] and [Meeuwissen and Cooke, 1994]. These were used to specify multivariate

⁴In this thesis the name Bayesian belief nets will be used in accordance to previous literature by the group at TU Delft. Pearl [1988] is the first to use this name as far as the author knows.

⁵For a biographical sketch of Bayes see Bellhouse [2004]

⁶Both Hartley and Price recognized the implications that Bayes's theorem would have for methods of reasoning Stigler [1983] and Barnard and Bayes [1958].

⁷For an overview on inverse probability the reader is referred to Dale [1999]. In page 10 Dale uses an example that is recurrent in the early literature on Bayesian networks [Pearl, 1982] and [Kim and Pearl, 1983]

⁸The readers interested in log-linear interaction models and gaussian dependence graphs are referred to Whittaker [1990] and Lauritzen [1996]

⁹Actually the method presented by Chow and Liu [1968] characterized trees as directed graphs and keeps a close relationship with BBNs.

distributions for use in uncertainty analysis. Their suitability for Montecarlo Simulation made them appealing for applications. The concept of a tree was later extended to allow for more complicated dependence structures.

Vines use sequences of conditional distributions to build a multivariate distribution where conditional bivariate constraints are satisfied. The first model with such characteristics was presented in Joe [1996] with no specific relation to graphs. Cooke [1997] introduced independently the formal concept of a vine as a graphical object that uses sequences of trees to build the joint distribution and Bedford and Cooke [2002] developed it further. Vines as graphical models will be discussed in more detail in chapter 2. Relevant information for researchers interested in vines is presented additionally in appendix A.

Vines and continuous BBNs are closely related. This was investigated in Kurowicka and Cooke [2005], Hanea et al. [2006] and Kurowicka and Cooke [2006]. In particular the theory behind Non-parametric Continuous-discrete BBNs (NPCDBBNs) was built around vines. Theorem 1.3.1 shows the main result of the copula vine approach to non-parametric continuous BBNs.

Theorem 1.3.1. *[Hanea et al., 2006] Given:*

- *A directed acyclic graph with n nodes specifying conditional independence relationships in a BBN;*
- *n variables, assigned to the nodes, with invertible distribution functions;*
- *the specification in equation (3.3) of conditional rank correlations on the arcs of the BBN and;*
- *a copula realizing all correlations $[-1, 1]$ for which zero correlation entails independence;*

the joint distribution of the n variables is uniquely determined. This joint distribution satisfies the conditional independence statements implied by the BBN and the conditional rank correlations in 3.3 are algebraically independent.

In the prove of theorem 1.3.1 D-vines (see chapter 2) were used[Hanea, 2008]. NPCDBBNs will be discussed in more detail in chapter 3. Emphasis will be made on the elicitation of rank and conditional rank correlations attached to the arcs of the BBN. The main application driving the ideas discussed in this thesis consists of a large scale NPCDBBN for measuring risks in the aviation industry. For its importance in this thesis the model will be briefly introduced in next section. The model will be explained in more detail in chapters 3 and 5.

1.4 Introduction to the Causal Model for Air transport Safety

As mentioned in section 1.1, BBNs will find their main application in this thesis in modeling risks in the aviation industry. The aviation sector is generally acknowledged for its impressive levels of safety. According to data from the Dutch National Aerospace Laboratory (NLR) [CAANL, 2008], the number of flights

worldwide has roughly doubled from 1980 to 2007. The number of fatal accidents on the other hand has not. Figure 1.8 presents the number of flights and the number of fatal accidents per year for the period between 1993 and 2007. The number of accidents per flight is decreasing, both world wide and for European Air Safety Agency (EASA) countries. Whereas worldwide, the fatal accident rate has been decreasing by 3.5% per year, for EASA countries, the fatal accident rate is decreasing by 5.3% per year (Figure 1.9).

The worldwide and EASA fatal and non-fatal accident frequencies are shown in Figure 1.10. The fatal and non-fatal accident frequency worldwide is decreasing by 1.8% per year. For EASA countries it is decreasing by 1.0% per year. The FAA forecasts growth in civil air transportation volume: “The active general aviation fleet is projected to increase at an average annual rate of 1.0 percent over the 17-year forecast period, growing from an estimated 234,015 in 2008 to 275,230 aircraft by 2025” [FAA, 2009, p.41]. If historical trends continue, this growth in volume must be accompanied with a decrease in the accident rate per flight in order to keep the absolute number of accidents minimum.

Human error plays an important role in aviation safety. About 56% of the accidents have humans as their main contributing factor (Figure 1.11¹⁰). The main causal contributor for accident is “cockpit crew”. Many responsible agencies have concluded that further improvements in safety would be served by a comprehensive system-wide risk model for civil aviation. This model should enable the disaggregation of fatal accidents into their causal components, including, in particular, human error.

The Netherlands ministry of Transport and Water Management commissioned a project for the realization of a causal model to be used for comparing alternatives for strengthening safety measures, for finding causes of incidents and accidents and for quantification of the probability of adverse events in the aviation system [Ale et al., 2006]. The model is being developed by a consortium including Delft University of Technology (TUD), Det Norske Veritas (DNV), the National Aerospace Laboratory (NLR) and White Queen (WQ). These organizations have been involved in the process of building the appropriate tools for the delivery of the model. The final product should be delivered in the form of a computer assisted decision tool supported by reports on the underlying technology and data [Ale et al., 2007].

Originally the The Causal Model for Air Transport Safety (CATS) comprised 3 different kinds of techniques: Fault Trees (FTs), Event Sequence Diagrams (ESDs) and BBNs. A schematic representation of the CATS model is presented in Figure 1.12.

The ESDs represent generic accident scenarios. Fault Trees link to the initiating events and pivotal events of the ESDs and describe them in a more detailed manner as a sequence of barrier failures. The base events of the fault trees include events representing human reliability, such as for instance ‘*autopilot incorrectly used by **flight crew***’, ‘*pilot disregards cross wind limit per severe wind*’, ‘*failure of*

¹⁰The human factor plays a role in the categories cockpit crew, maintenance and air traffic control.

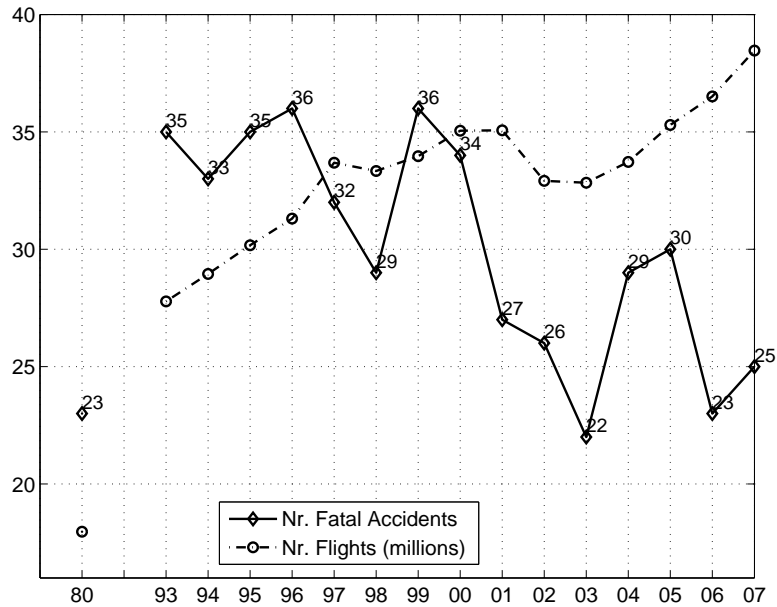


Figure 1.8: Worldwide number of flights and fatal accidents 1993-2007 CAANL [2008]. Commercial operated aircrafts with take-off weight $\geq 5,700$ kg.

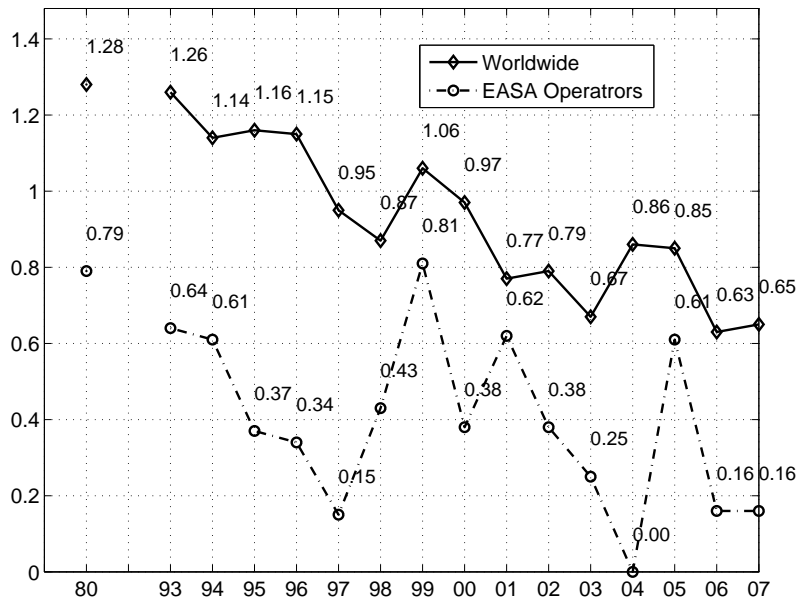


Figure 1.9: Worldwide and EASA fatal accidents per million flights 1993-2007 CAANL [2008]. Commercial operated aircrafts with take-off weight $\geq 5,700$ kg.

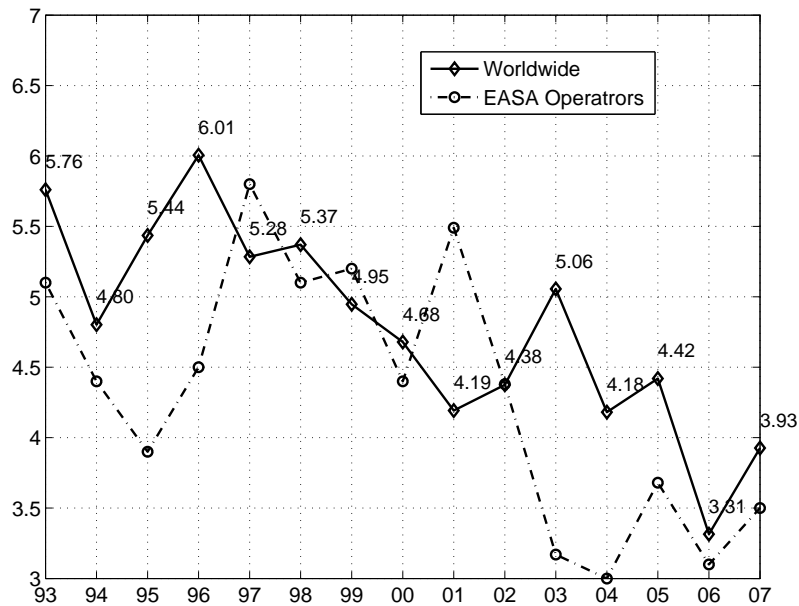


Figure 1.10: Worldwide and EASA fatal and non-fatal accidents per million flights 1993-2007 CAANL [2008]. Comm. op. aircrafts with take-off weight $\geq 5,700$ kg.

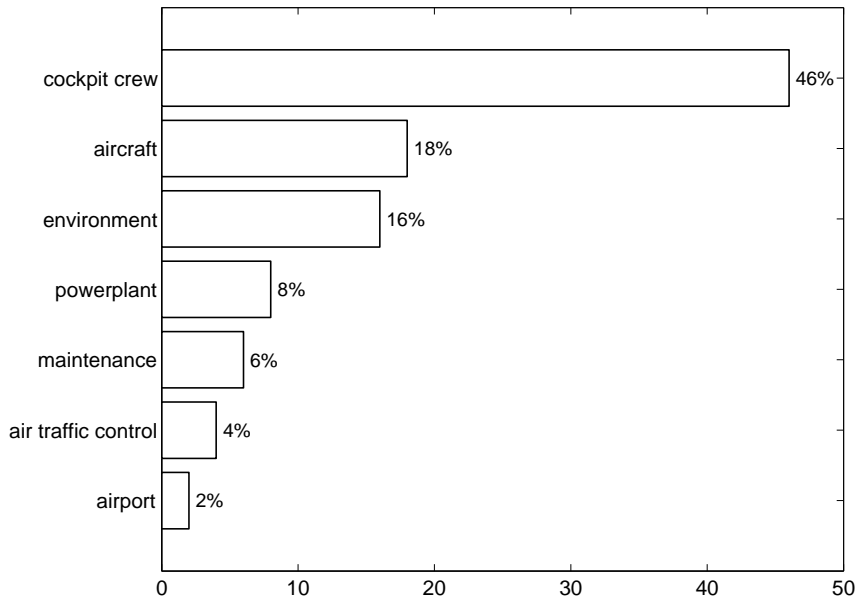


Figure 1.11: Relative importance of contributing factors in fatal accidents 1993-2007 CAANL [2008]. Commercial operated aircrafts with take-off weight $\geq 5,700$ kg.

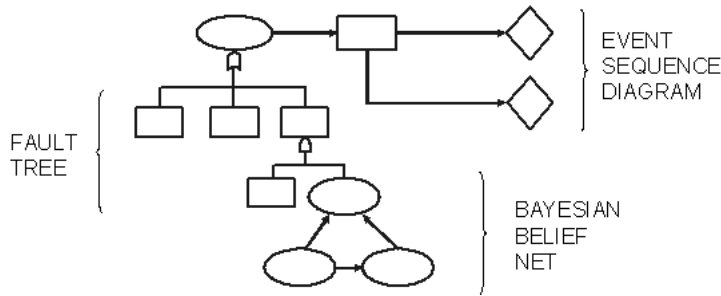


Figure 1.12: Schematic representation of the CATS model with ESDs, FTs and BBNs

air traffic controller to advise pilot per windshear on take off with LLWAS¹¹, or ‘breaks not applied correctly *by flight crew* per control following encounter with unexpected wind’. Base events involving human reliability are detailed further as Bayesian Belief Nets. BBNs are more general models than FTs and ESDs, hence ultimately these were also represented through functions as part of a large scale BBN. For this purpose UNINET [Cooke et al., 2007], a stand-alone software package is being developed at the Delft Institute of Applied Mathematics of the Delft University of Technology for dealing with large scale BBNs.

Figure 1.13 shows the BBN representing the CATS model. The graph in Figure 1.13 at the moment of publication consists of 1,504 nodes and 4,979 arcs. It is evident that the simple idea represented in Figure 1.12 becomes a very complicated graphical structure once all the elements of the model are finally quantified and integrated into a single BBN.

Building a Bayesian network with about 1.5 thousand nodes and 5 thousand arcs is a very complex task. Robinson [1977] presents results about unlabeled and labeled acyclic directed graphs. The number of unlabeled directed acyclic graphs¹² grows extremely fast with the number of nodes. Just to give an idea, the largest number of nodes for which unlabeled directed acyclic graphs has been computed is 18 and it is in the order of 1.55×10^{43} . The number of BBNs that one could construct with 1.5 thousand nodes are mind boggling. The CATS consortium brought together efforts from many professionals from different disciplines in order to construct the model shown in Figure 1.13. The major focus of this thesis is in the description of the quantification of the model in Figure 1.13. Emphasis is placed in the techniques used for the quantification of the dependence measures required by NPCDBBNs. Three human error models were quantified through structured expert judgment for the CATS model: flight crew, air traffic control and maintenance technician.

In the case of the CATS model, the distributions of the individual variables (marginal distributions) were almost all retrieved from data. The quantification and combination of dependence through expert opinion were a crucial step in

¹¹Low Level Windshear Alert System

¹²Which is a lower bound for the number of Bayesian networks possible on n nodes. An upper bound is the number of labeled acyclic directed graphs.

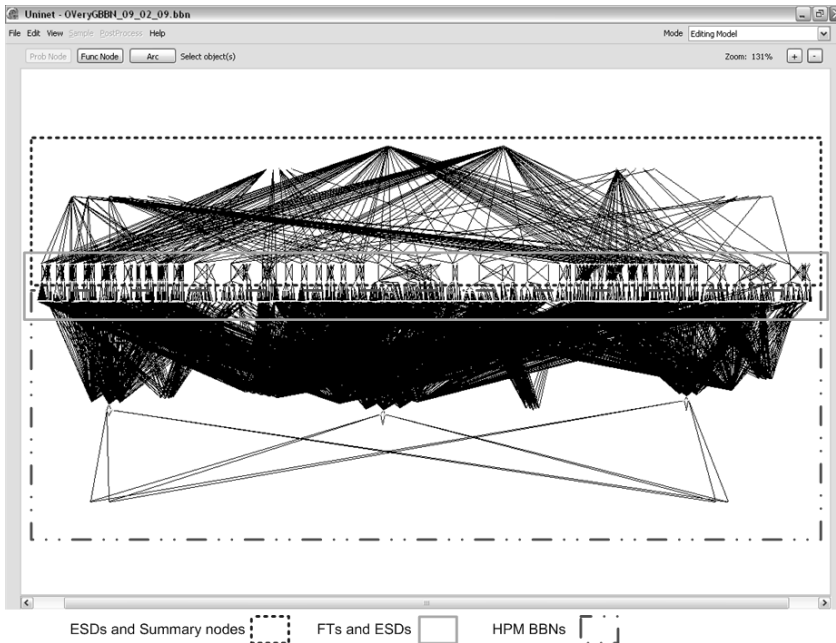


Figure 1.13: *The CATS model in UNINET.*

the CATS model as it provides a powerful tool for the analysis of the aviation system. This will be seen in the thesis through examples of model use. We can be confident that the joint distribution represents a validated expert belief about the influences of various variables on the accident probability. Further empirical validation of the CATS model should be a major goal in the future of the project.

Results from the CATS model that will be discussed in this thesis reflect the fact that the human error plays a mayor role in aviation safety. From the human actors involved in the aviation system and considered in the model, the cockpit crew and maintenance personnel appear more important than the air traffic control crew.

In particular, from the variables that measure human's performance at a basic level captain's and first officer's experience are the most highly rank correlated with accident probability (≈ -0.22 for each). This rank correlation is comparable to the correlation between aircraft generation and accident probability (-0.24). From the variables included for cockpit crew, training is the least important. The sample rank correlation between accident probability and training for both captain and first officer is close to zero (< 0.01).

Crew experience is immediately followed by maintenance technician experience in regards to absolute rank correlation with accident probability (≈ -0.21). Again this rank correlation is comparable to the one between accident probability and aircraft generation or fatigue. The rank correlation between maintenance crew and accident probability is smaller than 0.1 in absolute value for all other variables

related to maintenance crew.

In contrast with the flight crew or maintenance crew experience, the rank correlation between the accident probability and experience of air traffic controllers is about a factor 260 smaller than the correlation between accident probability and cockpit crew experience. For air traffic controllers the most important variable is the communication with cockpit crew. This is expressed through a rank correlation of 0.1 between total transmission time and accident probability. There are some applications where inferences with small correlations are not much different than those with independence. At first sight this could appear to be the case in the CATS model. However, as it will be seen next and later in chapters 3 and 7 the effect of model variables on accident probability can be large.

To illustrate the use of the model and the effect of rank correlations of the magnitude described previously, Table 1.1 is presented. Table 1.1 shows the result of conditionalizing on selected variables. For all three conditioning variables the 97th percentile if its distribution is used. Observe that though the rank correlation between accident probability and captain's experience is almost equal to the rank correlation between accident probability and maintenance technician experience the conditional distributions may differ significantly. The conditional mean of the accident probability when captain's experience is set to 17,016 *hrs.* is ≈ 3 times smaller than the original accident probability. The effect of captain's experience in accident probability is larger than the two other cases. The conditional mean of the accident probability given maintenance technician experience is 24 *yrs.* is ≈ 1.2 times smaller than the original accident probability. Finally, the conditional probability of accident given the air-ground transmission time is 100 *sec.* is ≈ 1.76 times larger than the unconditional mean. Conclusions similar to those briefly presented here are examples of possible use of the BBN representing the CATS model.

Uncond. Prob. of accident/flight	5% 8.58×10^{-8}	50% 4.59×10^{-7}	95% 8.98×10^{-6}	mean 3.18×10^{-6}
Conditioning variable	min	max	conditioned value	
Captain's experience (hrs)	3,069	27,913	17,016	
Cond. Prob. of accident/flight	5% 6.95×10^{-8}	50% 2.92×10^{-7}	95% 2.88×10^{-6}	mean 9.84×10^{-7}
Conditioning variable	min	max	conditioned value	
Maintenance technician experience (yrs)	0.6	31	24	
Cond. Prob. of accident/flight	5% 6.66×10^{-8}	50% 2.97×10^{-7}	95% 7.05×10^{-6}	mean 2.66×10^{-6}
Conditioning variable	min	max	conditioned value	
Air/ground total transmission time (sec)	17.5	306.5	100	
Cond. Prob. of accident/flight	5% 1.02×10^{-7}	50% 6.14×10^{-7}	95% 1.69×10^{-5}	mean 5.60×10^{-6}

Table 1.1: Unconditional probability of accident / flight, and conditional probability for selected variables.

The rest of the thesis is divided as follows: in chapter 2 the problem of enumerating regular vines is investigated. This section is of interest because in the last years the problem of finding an 'optimal' vine for data sets has been investigated. This requires a classification of regular vines and algorithms for generating them.

A result concerning the number of regular vines on n nodes is also discussed in chapter 2. Chapter 3 presents discrete BBNs and non parametric continuous discrete BBNs. The relationship between D-Vines and BBNs is also briefly discussed in chapter 3. The process of building the BBN from Figure 1.13 and examples of model use are presented in chapter 3 as well. In this thesis special attention is paid to the techniques for eliciting and combining rank and conditional rank correlations from domain experts as input for NPCDBBNs. This is discussed in chapter 4. The quantification of human reliability models used in the CATS model will be discussed in chapter 5. Chapter 6 presents an application of the same type of techniques used for measuring risks in the aviation system in measuring earth dams risks in Mexico. Finally, conclusions are presented in chapter 7.

CHAPTER 2

About The Number of Vines and Regular Vines on n Nodes.¹

2.1 Introduction

Man has always been fascinated by counting all sorts of different objects². The problem of counting graphs has been undertaken in the past [Harary and Palmer, 1973.]. Labeled trees find application in probability theory. Trees are the immediate ancestors of vines (section 1.3.3). These objects were first successfully counted by Cayley [1889].

Vines are graphical models that extend the idea of a tree. These objects have found application in probability theory and uncertainty analysis. More recently they are becoming popular in statistical analysis of data [Aas et al., 2009], [Aas and Berg, 2009], [Min and Czado, 2008], [Kolbjornsen and Stien, 2008], [Chollete et al., 2009].

In this chapter previous results concerning the number of trees on n nodes are briefly discussed in section 2.2. Section 2.3 presents two ways to characterize vines on n variables. The first method counts the total number of vines on n nodes and extracts regular vines by discarding those vines which are non-regular. The

¹This chapter is based on Morales-Nápoles et al. [2009a]

²Calculating prodigies have counted many things along history, Jedediah Buxton (1702) an illiterate man from Elmtun, England kept a mental record of all the free beer and ale he was given since the age of 12 and that averaged out to 5 or 6 ounces a day. When taken to see *Richard III* at the Drury Lane Playhouse in London “he declared after a fine piece of music, that the innumerable sounds produced by the instruments had perplexed him beyond measure, and he attended even to Mr. *Garrick* only to count the words that he uttered, in which, he says, he perfectly succeeded” [Smith, 1983]. Thomas Fuller, an African man shipped to America as a slave in 1724 “began his application to figures by counting to ten, and then when he was able to count a hundred, he thought himself (to use his own words) “a very clever fellow”. His first attempt after this was to count the number of hairs in a cow’s tail, which he found to be 2872” [Fauvel and Gerdes, 1990]

second method constructs all possible regular vines on n nodes using line graphs at each level in the vine. Neither method yields the number of regular vines on n nodes as a function of n .

Section 2.4 characterizes regular vines as triangular arrays, and finds the number of regular vines on n nodes by extending a regular vine on $n-1$ nodes. This enables us to express the number of regular vines on n nodes as $\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$. The results from section 2.3 may be contrasted with the result from section 2.4. For example, there are 11 unlabeled trees on 7 nodes each of which admits a number of regular vines. From these 11 trees, the one where every node has degree at most equal to 2 admits only one regular vine and can be labeled in 2,520 different ways. Other trees may be analyzed similarly to enumerate regular vines. In general for trees on seven nodes there are $2,520 \times 1 + 9 \times 2,520 + 19 \times 5,040 + 840 \times 33 + 630 \times 80 + 2,520 \times 168 + 840 \times 168 + 1,260 \times 342 + 420 \times 1,452 + 210 \times 2,928 + 7 \times 23,040 = 2,580,480 = \binom{7}{2} \times 5! \times 2^{\binom{7}{2}}$. Interestingly, the number of extensions of a regular vine on $n-1$ nodes to a regular vine on n nodes does not depend on the particular regular vine on $n-1$ nodes being extended. Section 2.5 gathers some conclusions and final comments.

2.2 Trees

A **tree** is an undirected acyclic graph. The graph isomorphism problem consist on deciding whether there exists a mapping from the nodes of one graph to the nodes of a second graph such that the edge adjacencies are preserved.

Definition 2.2.1. *Two labeled graphs $G_i = (E_i, N_i)$ and $G_j = (E_j, N_j)$ are **isomorphic** if there is a bijection $\varphi : N_i \rightarrow N_j$ such that for all pairs $(a, b) \in E_i \iff (\varphi(a), \varphi(b)) \in E_j$. If two graphs are isomorphic they are the same **unlabeled** graph.*

*A connected graph $T = (N, E)$ is called a **labeled tree** with nodes $N = \{1, 2, \dots, n\}$ and edges E , where E is a subset of pairs of N with no cycle.*

In this section labeled trees will be briefly discussed. These structures have been used to represent high dimensional probability distributions [Cooke, 1997] and they are often called dependence trees. This section however will be concerned with the properties of trees only as graphs. For an account of dependence trees see Kurowicka and Cooke [2006]. We begin our presentation with a well known result about trees.

2.2.1 The Number of Labeled Trees on n Nodes and the Prüfer Code

Two different labeled trees on 5 nodes are presented in Figures 2.1 and 2.2. The reader may observe that permuting nodes 1 and 5 in T_1 transforms it into T_2 and hence they would be the same unlabeled tree. In this section the interest will be mainly in labeled trees.

The first proof about the number of labeled trees on n nodes is due to Cayley [1889]. Since then several proofs have been presented [Moon, 1967].

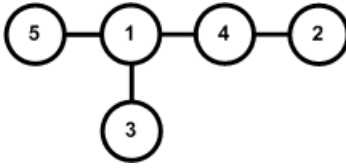


Figure 2.1: T_1 a tree on 5 nodes.

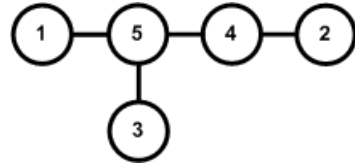


Figure 2.2: T_2 a tree on 5 nodes.

Theorem 2.2.1. *The number of labeled trees on n nodes is n^{n-2} .*

One of various proofs due to Prüfer [1918] of this theorem provides a very useful result for representing labeled trees. The argument is to notice that there is a one-to-one correspondence between the set of trees with n labeled nodes and the set of ordered $(n-2)$ -tuples $(A_1, A_2, \dots, A_{n-2})$ where each A_i is an integer not greater than n .

Definition 2.2.2. *Every sequence of numbers $R(T) = (A_1, A_2, \dots, A_{n-2})$ where each A_i is an integer not greater than n is a **Prüfer Code** for some labeled tree T on n nodes.*

In his paper Prüfer obtains the correspondence by the following procedure: For a given tree, remove the endpoint³ with the smallest label (other than the root⁴) and let A_1 be the label of the unique node which is adjacent to it. Remove the endpoint and the edge adjacent to it and a tree on $n-1$ nodes is obtained. Repeat the operation with the new tree on $n-1$ nodes to obtain A_2 and so on. The process is terminated when a tree on two nodes has been found. The reader may check that the trees from Figures 2.1 and 2.2 have Prüfer codes $R(T_1) = (4, 1, 1)$ and $R(T_2) = (5, 4, 5)$ respectively. The procedure described above may be easily reversed, that is, suppose you start with a sequence of $(n-2)$ -tuples $R(T) = (A_1, A_2, \dots, A_{n-2})$ then to obtain the only tree corresponding to the sequence one applies algorithm 2.2.1:

Algorithm 2.2.1. Decoding a Prüfer code.

1. Take a sequence $R(T_k) = (A_1, A_2, \dots, A_{n-2})$ for $k = 1, 2, \dots, n^{n-2}$ where each $A_i, i = 1, 2, \dots, n-2$ is an integer not greater than n .
2. Write the root in the right most position of $R(T_k)$. Notice that $R(T_k)$ has now length $n-1$ which is $|E|$.
3. Write another row of integers on the bottom of R_k from left to right. Each entry B_i in this new row is the smallest integer that has not been already written in this new row (the row of B'_i s) nor in the first row (the row of A'_i s) in the position exactly above it or every other position to the right.

³The endpoints are nodes with degree one in the tree, they are sometimes referred to as *leaves*.

⁴Without loss of generality we will choose node n as the root of all labeled trees on n nodes. Choosing any other node as the root makes no difference except that the algorithm and the procedure to find the Prüfer code for a given tree must be modified.

4. The resulting code $S(T_k)$ is the **Extended Prüfer Code**. Each column in the extended Prüfer code represents an arc in the unique labeled tree corresponding to it.

$$S(T_k) = \begin{pmatrix} A_1 & A_2 & A_3 & \dots & n \\ B_1 & B_2 & B_3 & \dots & B_{n-1} \end{pmatrix}$$

Take the two Prüfer codes $R(T_1) = (4, 1, 1)$ and $R(T_2) = (5, 4, 5)$. Apply algorithm 2.2.1 to decode each sequence into the extended Prüfer code. The reader may check in equation (2.2.1) that $S(T_1)$ corresponds to Figure 2.1 and $S(T_2)$ to Figure 2.2.

$$S(T_1) = \begin{pmatrix} 4 & 1 & 1 & 5 \\ 2 & 3 & 4 & 1 \end{pmatrix}, S(T_2) = \begin{pmatrix} 5 & 4 & 5 & 5 \\ 1 & 2 & 3 & 4 \end{pmatrix} \quad (2.1)$$

Prüfer then gives an induction argument to show that for each $(n - 2)$ -tuple there is some tree which determines the given sequence by the above procedure. From the code one can see that a node with degree m would occur exactly $m - 1$ times in the code. Labeled trees are interesting not only as objects that can be counted and subject of combinatorial problems. They find application in optimization, probability theory and uncertainty analysis ([Cooke, 1997], [Kurowicka and Cooke, 2006]). In next section vines will be discussed and the ideas presented in this section will be extended to deal with these graphical objects.

2.3 Vines

A vine [Cooke, 1997] is a set of nested trees. Just as labeled trees, vines have been used to represent high dimensional probability distributions [Bedford and Cooke, 2002] and [Kurowicka and Cooke, 2006] with applications in uncertainty analysis. More recently they are being applied in statistical analysis of multivariate data sets [Aas et al., 2009], [Min and Czado, 2008], [Aas and Berg, 2009] and [Chollete et al., 2009]. These last references are concerned with choosing an optimal vine to represent multivariate data sets. Algorithms for enumerating all possible regular vines on n nodes will be needed for this purpose. All trees in a vine may be thought of as labeled trees. In this section some results about the number of vines on n nodes will be presented.

2.3.1 The Number of Vines on n Nodes and the Prüfer Code

The ideas presented in section 2.2.1 can be extended to count the number of vines (and regular vines) that are possible on n variables. This will be shown in the present subsection. This subsection begins with the definitions of vine and regular vine.

Definition 2.3.1. $V(n)$ is a labeled vine on n elements if:

1. $V(n) = (T_1, T_2, T_3, T_4, \dots, T_n)$.

2. T_1 is a labeled tree with nodes $N_1 = 1, 2, \dots, n$ and edges E_1 . For $i = 2, \dots, n$, T_i is a labeled tree with nodes $N_i = E_{i-1}$. E_{i-1} has been given a unique labeling.

If in addition for $i = 2, \dots, n - 1$, if $(a, b) \in E_i$, then $|a \triangle b| = 2$, where \triangle denotes the symmetric difference, then $V(n)$ is a **labeled regular vine**. In other words, if a and b are nodes of T_i connected by an edge in T_i , where $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i . This condition is called the **proximity condition**.

The nodes reachable from a given edge in a regular vine are called the **constraint set** of that edge. When two edges are joined by an edge in tree T_i , the intersection of the respective constraint sets form the **conditioning set**. The symmetric difference of the constraint sets is the **conditioned set**. Formal definitions may be found in Kurowicka and Cooke [2006]. Vines (and regular vines) may be classified according to the unlabeled tree used at each level in the vine. For this reason the following definition is introduced.

Definition 2.3.2. If a bijection as in definition 2.2.1 may be found for each $T_i \in V_k(n)$ and $T_i \in V_j(n)$ then we speak of the same **tree-equivalent vine** and accordingly the same **tree-equivalent regular vine** when the proximity condition holds.

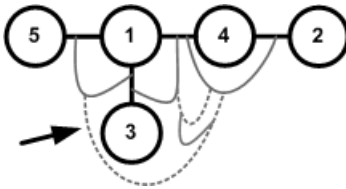


Figure 2.3: Non-regular vine on 5 nodes.

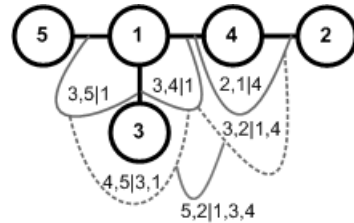


Figure 2.4: Regular vine on 5 nodes.

In Figures 2.3 and 2.4 respectively a non-regular and a regular vine on five nodes are generated. The edge that makes Figure 2.3 a non-regular vine is indicated by an arrow. The conditioned set is separated from the conditioning set by a vertical line “|” in Figure 2.4. Obviously these two vines are different labeled vines. However, according to definition 2.3.2 they are the same tree-equivalent vine. Observe that by permuting the numbers in T_1 in Figure 2.4 we would generate different labeled regular vines but according to definition 2.3.2 the same tree-equivalent regular vine.

Since every labeled tree can be represented by a Prüfer code, then every subtree in the vine may also be represented by a Prüfer code and in this way the vine may be generated. A way to write all possible vines on n nodes is presented in algorithm 2.3.1.

Algorithm 2.3.1. Constructing all possible vines on n nodes.

1. Set $i = 1$.

2. Construct all Prüfer codes possible for T_i .
3. The edges of each one of the $n^{n-(i+1)}$ trees in step 2 become nodes in T_{i+1} . Hence, for each tree in step (2):
 - (i) Label the $n - i$ edges of each tree giving the label 1 to the edge appearing in the first column in its extended Prüfer code, 2 to the edge in the second column and so on until all edges have been labeled ⁵.
 - (ii) Construct all Prüfer codes possible for T_{i+1} and connect the new labeled edges (from T_i) as nodes according to these new Prüfer codes.
4. Set $i := i + 1$ and go to step (3) until two edges must be connected in the last tree. At this point there is only one way to connect them and no Prüfer code is required.

From algorithm 2.3.1 it may be observed that to write any vine on n nodes all is required are $n - 2$ Prüfer codes. The first one of length $n - 2$, the second one of length $n - 3$ and so on until the last one of length 1. A vine on n nodes may be represented by an upper triangular array of size $(n - 2) \times (n - 2)$ whose first row represents the Prüfer code of the first tree in the vine, the second row the second tree of the vine and so on. For example $V_1(5)$ represents the vine from Figure 2.3 and $V_2(5)$ the one in 2.4 :

$$V_1(5) = \begin{pmatrix} 4 & 1 & 1 \\ & 3 & 2 \\ & & 1 \end{pmatrix}, V_2(5) = \begin{pmatrix} 4 & 1 & 1 \\ & 3 & 2 \\ & & 2 \end{pmatrix} \quad (2.2)$$

Corollary 2.3.1. *The number of vines on n nodes is $\prod_{i=1}^n i^{i-2}$.*

Proof. The proof is in fact algorithm 2.3.1. This is a consequence of theorem 2.2.1 and definition 2.3.1. \square

Regular vines are most interesting in uncertainty analysis. Implementing Algorithm 2.3.1 in a computer is very easy and it provides a simple way to construct all possible regular vines on n nodes by simply discarding those that are not regular. However, this method incurs an excessive burden of searching all vines (see table 2.1). According to corollary 2.3.1 the number of vines grows extremely fast with n and it could be very restrictive in time to find all regular vines even for a modest number of nodes (8 or 9). Another possibility to construct only regular vines will be discussed in the next subsection.

2.3.2 Regular vines and the line graph

As stated at the end of previous section, another possibility is available to produce only regular vines as opposed to producing all possible vines and discarding those

⁵This labeling is not unique and any other labeling would work equally well as long as all n^{n-2} trees are labeled in the same way.

that are not regular as in algorithm 2.3.1. The idea is to use the line graph⁶ of each tree in the vine. Harary notes in [Harary, 1969] that the concept of the line graph of a given graph is so natural that it has been rediscovered independently by many authors.

Definition 2.3.3. [Beineke, 2006] The **line graph** $LG(G)$ of a graph G has as its nodes the edges of G , with two nodes being adjacent in LG if the corresponding edges are adjacent in G .

If the edges of the first tree of Figure 2.4 are labeled according to the second step in algorithm 2.3.1 then the line graph of this tree can be found according to definition 2.3.3. This line graph corresponds to Figure 2.5. Nodes 1, 2, 3 and 4 in Figure 2.5 corresponds to edges (4,1), (1,3), (1,4) and (5,1) respectively in Figure 2.4.

If in the same way we label the nodes of the second tree in the vine in Figure 2.4 accordingly, then the line graph in Figure 2.6 may be obtained. In this new line graph, nodes 1, 2, 3 correspond respectively to nodes (2,1|4), (3,4|2) and (3,5|1) in Figure 2.4.

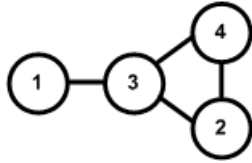


Figure 2.5: Line Graph of the first tree in Figure 2.4

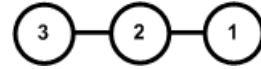


Figure 2.6: Line Graph of the second tree of the vine from Figure 2.4.

Definition 2.3.4. [Harary, 1967] A **spanning subgraph** T of a graph G is a subgraph with the same set of nodes as G . If T is a tree, it is called a **spanning tree** of G .

It is clear from definitions 2.3.3 and 2.3.4 that in order to find all regular vines on n nodes, all the spanning trees of the line graphs of all subtrees in the vine must be found. This result is summarized in algorithm 2.3.2.

Algorithm 2.3.2. Constructing all possible regular vines on n nodes.

1. Set $i = 1$.
2. Construct all Prüfer codes possible for T_i .
3. The edges of each one of the $n^{n-(i+1)}$ trees in step 2 become nodes in T_{i+1} .
Hence, for each tree in step (2):

⁶Line graphs are also known as derived graphs, interchange graphs, adjoint and edge to vertex dual[Beineke, 2006].

Nodes	Trees		Vines		
	A ^a	B ^b	C ^c	D ^d	E ^e
3	1	3	3	3	1
4	2	16	48	24	2
5	3	125	6,000	480	5
6	6	1,249	7,776,000	23,040	22
7	11	16,807	130,691,232,000	2,580,480	136
8	23	262,144	34,259,922,321,408,000	660,602,880	1,464
9	47	4,782,969	$1.63864146405703 \times 10^{23}$	380,507,258,880	24,115

Table 2.1: Number of unlabeled and labeled trees, vines, regular vines and tree-equivalent classes of regular vines in 3, 4, 5, 6, 7, 8 and 9 nodes.

^aNumber of unlabeled trees

^bNumber of labeled trees

^cNumber of labeled vines

^dNumber of labeled regular vines

^eNumber of tree-equivalent regular vine classes.

- (i) Label the edges of each tree giving label 1 to the edge appearing in the first column in its extended Prüfer code, 2 to the edge in the second column and so on until all edges have been labeled ⁷.
4. Construct the line graph of each one of the trees from step 2.
5. For each line graph from step 3 find all possible spanning trees. Connect the edges of each tree in step 1 according to all spanning trees from its line graph. This will give all possible T_{i+1} for each T_i .
6. Set $i := i + 1$ and go to step (2) until two edges must be connected in the last tree. At this point there is only one way to connect them and no Prüfer code is required.

Notice that the vines generated by this procedure may still be stored in an $(n-2) \times (n-2)$ upper triangular array as in equations (2.2) once a way of labeling the edges from each tree in the vine is specified. Algorithm 2.3.2 does not produce any irregular vine as opposed to algorithm 2.3.1. However it involves a greater programming effort and more operations as all possible spanning trees of the line graphs in all trees in the vine must be found. Several algorithms for finding all spanning trees of a given graph have been proposed and examined [Minty, 1965], [Mayeda and Seshu, 1967], [Read and Tarjan, 1975], [Smith, 1997] and [Shioura et al., 1994]. In general finding all possible spanning trees of a given graph other than a complete graph ⁸ is demanding in terms of time and space [Smith, 1997].

Table 2.1 presents a summary with the number of labeled trees, vines and regular vines on 3, 4, 5, 6, 7, 8 and 9 nodes⁹. The second column presents the

⁷As before, this labeling is not unique and any other labeling would work equally well as long as all n^{n-i+1} are uniquely labeled.

⁸For a complete graph all possible spanning trees are the n^{n-2} Prüfer codes

⁹For 1 and 2 variables there is exactly one of each object.

number of unlabeled trees on n nodes. The third column corresponds to the values obtained by applying the formula in theorem 2.2.1 and the fourth to values obtained by applying the formula in corollary 2.3.1. Algorithms 2.3.1 and 2.3.2 allow to count the number of regular vines on n nodes. The number of regular vines on up to 7 nodes was found using algorithm 2.2.1 and the values for 8 and 9 nodes using algorithm 2.3.2¹⁰. The results of counting regular vines with algorithms 2.2.1 and 2.3.2 are presented in column 5. To implement algorithm 2.3.2, MATGRAPH [Sheinerman, 2009] was used to find line graphs for each of the 23 and 47 unlabeled trees on 8 and 9 nodes. A version of the Mayeda-Seshu algorithm was used [Smith, 1997, p.10] to find all spanning trees of each of the 70 line graphs.

Column six in table 2.1 presents the number of tree-equivalent regular vines on n nodes. Also, algorithm 2.3.1 may be used to list the number of tree-equivalent vines (or tree-equivalent regular vines) on n nodes by checking for isomorphism at each level in the vine. Also, algorithm 2.3.2 can be used to count the number of tree-equivalent regular vines on n nodes by checking tree isomorphism at each level of the vine¹¹. Appendix A presents a catalogue with non-isomorphic trees on 1, 2, 3, 4, 5, 6, 7, 8 and 9 nodes and some relevant characteristics of each one. In particular an example of Prüfer code, the number of labeled trees, the number of regular vines per labeled tree and the number of tree-equivalent regular vines is shown.

A similar catalogue was presented in Moon [1967] for trees with at most five nodes. In Kasyanov and Evstigneev [2000] a catalogue of non-isomorphic trees with at most 8 nodes may be found¹². None of the above catalogues presents results for vines.

Tables A.1 to A.4 present the 48 trees on 8 nodes or less. These trees will be used to present pictures of tree-equivalent regular vines on at most 6 nodes in tables A.8 and A.9. Finally tables A.10 to A.32 present tree-equivalent regular vines on 7 and 8 nodes.

The concept of the line graph also allows to obtain bounds for the number of regular vines admissible by unlabeled trees on n nodes. These results are presented next as lemmas. Lemma 2.3.3 that is rather evident has been stated in Cooke [1997] without a proof.

Lemma 2.3.2. *If the first tree of a vine on n nodes has one node with maximal degree, then the number of labeled regular vines possible with this tree equals the number of labeled regular vines on $(n - 1)$ nodes.*

Proof. Since every edge in T_1 is adjacent to each other then the line graph

¹⁰Actually algorithm 2.3.2 does not need to be implemented completely to count the number of regular vines on 8 and 9 nodes. Observe that it is sufficient to know how many spanning trees of each unlabeled class in $n - 1$ nodes does a line graph of a tree in n nodes contain.

¹¹As for counting regular vines algorithms 2.3.1 and 2.3.2 do not need to be implemented completely to count the number of tree-equivalent regular vines on 8 and 9 nodes. Observe that it is sufficient to know how many spanning trees of each unlabeled class in $n - 1$ nodes does a line graph of a tree in n nodes contain.

¹²This catalogue repeats a tree in eight nodes neglecting another one. In the same reference tables counting the number of rooted trees on up to 26 nodes and the number of non-isomorphic trees on less than 26 nodes may be found.

of this tree is a complete graph on $(n - 1)$ nodes that has $(n - 1)^{n-3}$ possible spanning trees. These are all possible labeled trees on $n - 1$ nodes each of which admits a fixed number of labeled regular vines. \square

Lemma 2.3.3. *If the first tree of a vine on n nodes has $(n - 2)$ nodes with degree 2, then the number of regular vines possible with this tree equals 1.*

Proof. Observe that the line graph of T_1 will be also a tree on $n - 1$ nodes with $(n - 3)$ nodes with degree 2. Hence its only possible spanning tree will be itself and to preserve regularity this tree should be used in T_2 . The same argument holds for all $j \geq 2$ and hence only one regular vine is possible. \square

Lemma 2.3.3 provides a lower bound for the number of regular vines possible for a given unlabeled tree T_1 . In the same way lemma 2.3.2 provides an upper bound. This result may be observed in tables A.1 to A.11 in appendix A. A more general result for counting labeled regular vines is dealt with in next section.

In applications two kind of regular vines have been most widely used. **C-Vines** are regular vines for which each tree in the vine has one node with maximal degree. **D-Vines** are regular vines for which the first tree of the vine has $(n - 2)$ nodes with degree 2. Next results about the number of D-vines and C-vines on n nodes are presented. Both results were presented in Aas et al. [2009] with proofs that are slightly different to the ones presented here.

Lemma 2.3.4. *The number of C-vines on n nodes equals the number of D-vines on n nodes and is $(n!/2)$*

Proof. For C-vines observe that there are n possible labeled trees on n nodes for which a single node has maximal degree. Once the first tree has been fixed any of the $(n - 1)$ edges may be chosen so as to construct any of the $(n - 1)$ possible labeled trees on $(n - 1)$ nodes for which a single node has maximal degree. Any of these would preserve regularity. The same argument holds for all other trees on the vine until two edges need to be connected as nodes in T_{n-1} . Hence there are $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (3) = (n!/2)$ C-vines on n nodes.

For D-vines observe that from lemma 2.3.3, $T_1 \in V$ completely determines the vine. And since there are $(n!/2)$ ways of choosing it the result follows. \square

2.4 The Number of Regular Vines on n Nodes.

So far the number of vines has been obtained from Cayley's theorem in corollary 2.3.1. Results concerning the number of tree-equivalent and labeled regular vines on at most 8 nodes have been presented by using Prüfer codes and line graphs. This section derives a formula for the number of regular vines on n nodes.

Definition 2.4.1. *If node e is an element of node f in a regular vine, we say that e is an **m-child** of f ; similarly, if e is reachable from f via the membership relation: $e \in e_1 \in \dots \in f$, we say that e is an **m-descendent** of f .*

Lemma 2.4.1. *Kurowicka and Cooke [2006] For any node M of order $k > 0$ in a regular vine, if node i is a member of the conditioned set of M , then i is a member of the conditioned set of exactly one of the m -children of M , and the conditioning set of an m -child of M is a subset of the conditioning set of M .*

Definition 2.4.2. *If element a occurs with element b as conditioned variables in tree k , then a and b are termed **k-partners**. Nodes A and B are **siblings** if they are m -children of a common parent.*

Regularity¹³ means that every node in T_i , $i \geq n - 1$ must have a sibling and a common child with its sibling. In this section, another triangular array representing a regular vine will be introduced. In this section another triangular array representing a regular vine will be introduced. One disadvantage of using a triangular array such as the one used in section 2.3.1 is that the information regarding the label of nodes in the first tree of a regular vine is lost when assigning new labels to its edges when they become nodes of the next tree. The same happens as more trees are added to a regular vine. This means that conditioned and conditioning sets are not immediately visible anymore. The idea of the construction presented here is to preserve the information concerning the labels of the first tree as lower trees in the vine are added. In analogy to a Prüfer code a sequence of n -tuples $(A_n, A_{n-1}, \dots, A_1)$ where each A_i is an integer not greater than n will be called a **natural order**. This is defined next.

Definition 2.4.3. *A natural order of the elements of a regular vine on n elements is a sequence of numbers $NO(V(n)) = (A_n, A_{n-1}, \dots, A_1)$ where each A_i is an integer not greater than n obtained as follows: Take one conditioned element of the last tree of a regular vine (a tree with a single node and no edges) and assign it position n ; assign the other conditioned element of the top node position $(n - 1)$. Element A_{n-1} occurs in one m -child of the top node with an $(n - 1)$ -partner in the conditioned set. Give this $(n - 1)$ -partner position $(n - 2)$ and iterate this process until all elements have been assigned a position.*

Observe that there are two natural orders for every regular vine. A representation of the regular vine in Figure 2.4 using a directed graph is presented in Figure 2.7. This representation will be useful in the rest of the chapter for explaining some of the concepts introduced. The nodes of each tree in the regular vine are nodes in the directed graph. Observe that every parent node has exactly two children. The conditioned set is presented to the left of a vertical line ($|$ sign) and the conditioning set to its right.

The element in position n occurs as conditioned variable in tree T_n (this tree has one node and no edges). The element in position $(n - j)$ occurs in the unique node of tree T_{n-j} with conditioned set $\{A_{n-(j+1)}, A_{n-j}\}$. If 5 is chosen as A_n , then by definition 2.4.3 the natural order of the regular vine would be $NO_1(V_2(5)) = (5, 2, 3, 4, 1)$ for $j = 1, \dots, n - 2$. In the same way if node 2 was chosen as element A_n then the natural order would be $NO_2(V_2(5)) = (2, 5, 4, 3, 1)$. A regular vine may be coded as a lower triangular array with the natural ordering

¹³Or proximity in the language of subsection 2.3.1

on the diagonal. The natural order will be used in a triangular array similar to the one introduced in section 2.3.1 but that preserves all the information regarding conditioned and conditioning sets in the regular vine.

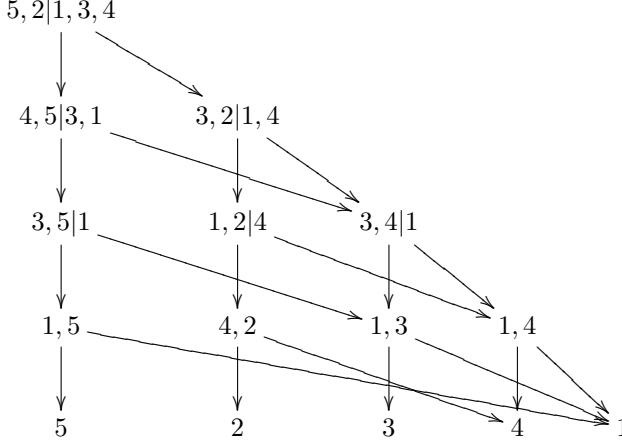


Figure 2.7: Representation of the regular vine in Figure 2.4

Definition 2.4.4. A *regular vine array* $TA(V(n)) = \{A_{i,j}\}$ for $i, j = 1, \dots, n$ and $j \geq i$ is a lower triangular matrix with elements in $\{1, \dots, n\}$ indexed in ‘reverse order’ (see equation (2.3)), where $A_{j,j}$ equals the element in position j in $NO(V(n))$ and $A_{j-1,j}$ equals the element in position $j - 1$ in the same natural order. The **echelon** of element $A_{i,j}$ is i and element $A_{i,j}$ codes the node $(A_{j,j}, A_{i,j} | A_{i-1,j}, \dots, A_{1,j})$

The regular vine array $TA(V_2(5))$ array corresponding to the regular vine $V_2(5)$ in equation (2.2) (Figures 2.4 and 2.7) using $NO_1(V_2(5))$ is presented in equation (2.3). Observe that the row and column indices are in their usual position but their sense is reversed (with respect to traditional matrix indexing) in order to facilitate adding new variables to the left. From definition 2.4.4, we may speak unambiguously of “node”, “element” or “variable” $A_{i,j}$. Thus the “node $A_{i,j}$ ” is the set of elements “ $(A_{i,j}, A_{j,j} | A_{i-1,j}, \dots, A_{1,j})$ ”, arranged to separate the conditioned elements from the conditioning elements by “|”.

$$TA(V_2(5)) = \begin{pmatrix} A_{5,5} & & & & \\ A_{4,5} & A_{4,4} & & & \\ A_{3,5} & A_{3,4} & A_{3,3} & & \\ A_{2,5} & A_{2,4} & A_{2,3} & A_{2,2} & \\ A_{1,5} & A_{1,4} & A_{1,3} & A_{1,2} & A_{1,1} \end{pmatrix} = \begin{pmatrix} 5 & & & & \\ 2 & 2 & & & \\ 4 & 3 & 3 & & \\ 3 & 1 & 4 & 4 & \\ 1 & 4 & 1 & 1 & 1 \end{pmatrix} \quad (2.3)$$

From Figure 2.7 and equation (2.3) it may be observed that a regular vine may be represented by a triangular array as described in definition 2.4.4, in which the

nodes of each tree in a regular vine have children in the immediate lower order tree. Conditions for child nodes in the triangular array are given next.

Definition 2.4.5. Node $A_{i-1,h}$ is a **child** of node $A_{i,j}$ if:

- (i) $\{A_{h,h}, A_{i-1,h}, A_{i-2,h}, \dots, A_{1,h}\} \subset \{A_{j,j}, A_{i,j}, A_{i-1,j}, A_{i-2,j}, \dots, A_{1,j}\}$
- (ii) $|\{A_{h,h}, A_{i-1,h}, A_{i-2,h}, \dots, A_{1,h}\}| = |\{A_{j,j}, A_{i,j}, A_{i-1,j}, A_{i-2,j}, \dots, A_{1,j}\}| - 1$
- (iii) $|\{A_{h,h}, A_{i-2,h}\} \cap \{A_{j,j}, A_{i-1,j}\}| = 1$

The reader may check for example that according to definition 2.4.5 $A_{2,4} = (2, 1|4)$ and $A_{2,3} = (3, 4|1)$ in 2.3 are children of $A_{3,4} = (2, 3|1, 4)$. According to definition 2.4.2, $A_{3,4} = (2, 3|1, 4)$ and $A_{3,5} = (5, 4|3, 1)$ are siblings because they are children of the common parent $A_{4,5} = (5, 2|4, 3, 1)$. Similarly $A_{2,3} = (3, 4|1)$ and $A_{2,5} = (5, 3|1)$ are children of $A_{3,5} = (5, 4|3, 1)$ and hence siblings. Other elements may be also checked by the reader. Next it will be shown that a matrix such as the one in definition 2.4.4 represents a regular vine.

We characterize first those triangular arrays which represent regular vines.

Theorem 2.4.2. $TA(V(n))$ represents a regular vine $\iff TA(V(n))$ satisfies **condition R**. That is, for all $i \geq 2$, element $A_{i,j} = A_{h,h}$ or $A_{i,j} = A_{i-1,h}$ for some h such that $i \leq h < j$ and $\{A_{j,j}, \dots, A_{i+1,j}\} \cap \{A_{i-1,h}, \dots, A_{1,h}\} = \emptyset$

Proof. \Rightarrow If $V(n)$ is a regular vine then every node $A_{i,j}$ in $TA(n)$ has two children in echelon $i - 1$ one of which is $A_{i-1,j}$. Suppose the other child is in column h , then condition *R* follows from (i), (ii), (iii) in definition 2.4.5.

\Leftarrow Let $TA(k)$ be a regular vine array satisfying condition *R*. If $k = 3$, the nodes of $TA(k)$ clearly satisfy regularity. Suppose the theorem holds for $k = n - 1$. Node $A_{n-1,n}$ satisfies regularity by definition 2.4.4. An induction will show that nodes $A_{n-1,n}, \dots, A_{1,n}$ satisfy regularity. We show first that node $A_{n-2,n}$ has a sibling and has a common child with this sibling. By condition *R* element $A_{n-2,n}$ is equal to element $A_{n-2,n-2}$ or $A_{n-3,n-2}$. In either case, node $A_{n-3,n-2}$ is a child of node $A_{n-2,n}$ and hence node $A_{n-2,n}$ satisfies regularity.

Suppose that for every $j = n - 2, \dots, k + 1$, every node $A_{j,n}$, satisfies regularity. We claim that $A_{k,n}$ must also satisfy regularity.

Node $A_{k,n}$ is a child of node $A_{k+1,n}$ and by the induction hypothesis, node $A_{k+1,n}$ satisfies regularity, therefore, it has a second child node $A_{k,h}$ and relation 2.4 must hold according to condition *R*.

$$\{A_{h,h}, A_{k,h}, A_{k-1,h}, \dots, A_{1,h}\} = \{A_{k+1,n}, A_{k,n}, \dots, A_{1,n}\} \quad (2.4)$$

Two situations are possible:

- (i) $A_{k+1,n} = A_{h,h}$ or,
- (ii) $A_{k+1,n} = A_{k,h}$

By induction node $A_{k,h}$ has two children, one of which is node $A_{k-1,h}$. It will be shown that one of these children must be a child of node $A_{k,n}$. In other words

it will be shown that $A_{k,n}$ and $A_{k,h}$ are siblings and have a common child which is the condition for regularity.

In case (i) node $A_{k-1,h}$ cannot be a child of node $A_{k,n}$ since node $A_{k-1,h}$ contains element $A_{h,h} = A_{k+1,n}$ in its conditioned set, and element $A_{k+1,n}$ cannot belong to the constraint set of node $A_{k,n}$. The other child of node $A_{k,h}$ must be node $A_{k-1,m}$ for some $k \leq m < h$. This child cannot contain element $A_{h,h}$, and:

$$\{A_{m,m}, A_{k-1,m}, \dots, A_{1,m}\} = \{A_{k,h}, A_{k-1,h}, \dots, A_{1,h}\} \quad (2.5)$$

By induction, node $A_{k,h}$ satisfies regularity; therefore either element $A_{k,h} =$ element $A_{m,m}$ or element $A_{k,h} =$ element $A_{k-1,m}$, in either case by combining 2.4 and 2.5 we see that node $A_{k-1,m}$ is a child of node $A_{k,n}$.

In case (ii) element $A_{k+1,n} \neq$ element $A_{h,h}$, and equation (2.4) must still hold and by induction $A_{k,n} = A_{h,h}$ or $A_{k,n} = A_{k-1,h}$; in either case node $A_{k-1,h}$ will be a child of node $A_{k,n}$ and the latter will satisfy regularity. \square

We now count the number of ways of extending an $n - 1$ regular vine with a fixed natural ordering. This is equivalent to adding an additional column to the left of a regular vine in the triangular array.

Evidently the top two elements of this new column are fixed, and the last element is fixed by the choices for the elements above it. If there are n elements in the new column, there are $n - 3$ elements to be chosen. It will be seen that the number of extensions is in fact 2^{n-3} regardless of the regular vine being extended.

Theorem 2.4.3. *For any vine on $n - 1$ elements, the number of regular n vines which extend this vine, preserving the natural ordering of the $n - 1$ vine is 2^{n-3} .*

Proof. Let $V(n - 1)$ be an arbitrary regular vine on $n - 1$ elements with a natural order and $TA(V(n - 1))$ its triangular array. $TA(V(n - 1))$ will be extended by adding a column of n elements to the left whose top two entries are $A_{n,n}$, $A_{n-1,n}$. The goal is to count the number of ways of adding a column to the left of $TA(V(n - 1))$, so as to preserve regularity. Node $A_{k,n}$ satisfies regularity if it has a sibling which is a child of node $A_{k+1,n}$ and has a child which is also a child of its sibling. This latter child must be a node in $V(n - 1)$. If each node $A_{k,n}$ for $k = 2, \dots, n - 2$, satisfies regularity, then $TA(V(n))$ (the extended triangular array) codes a regular vine which extends the original regular vine $V(n - 1)$.

$V(n - 1)$ has trees T_{n-1}, \dots, T_1 where T_{n-1} has one node and no edges, T_1 has $n - 1$ nodes and $n - 2$ edges; in general for $j = 1, \dots, n - 1$ tree T_{n-j} has j nodes and $j - 1$ edges. After adding node $A_{n,n}$, T_1 will have n nodes and $n - 1$ edges, T_2 will have $n - 1$ nodes and $n - 2$ edges and so on until tree n that will have a single node $A_{n-1,n} = (A_{n,n}, A_{n-1,n} | A_{n-2,n}, \dots, A_{1,n})$. This node must have two children. One child must be, evidently, node $A_{n-2,n} = (A_{n,n}, A_{n-2,n} | A_{n-3,n}, \dots, A_{1,n})$ and the other is the top node of $V(n - 1)$ which is $A_{n-2,n-1} = (A_{n-1,n-1}, A_{n-1,n-2} | A_{n-3,n-1}, \dots, A_{1,n-1})$. To satisfy regularity, nodes $A_{n-2,n}$ and $A_{n-2,n-1}$ must have a common child. This common child cannot contain element $A_{n-1,n} = A_{n-1,n-1}$ since it does not belong to node $A_{n-2,n}$ and hence the child must be of the form:

$$(A_{n-3,n-2}, A_{n-2,n-2} | A_{n-4,n-2}, \dots, A_{1,n-2})$$

The situation is pictured in Figure 2.8.

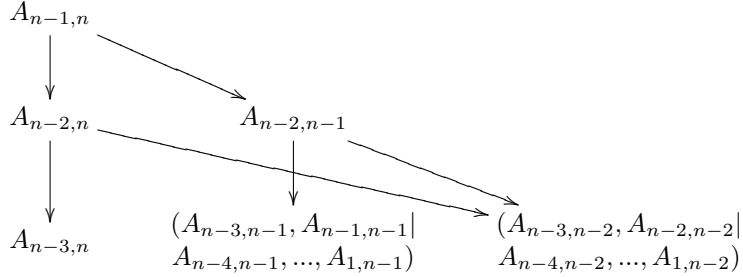


Figure 2.8: Regular Vine Growing 1.

Since element $A_{n-2,n}$ must be in exactly one of the children of the node $A_{n-2,n-1}$ it follows that element $A_{n-2,n}$ must be element $A_{n-2,n-2}$ or element $A_{n-3,n-2}$ either choice satisfying regularity.

Assume that variables $A_{n-1,n}, \dots, A_{k+1,n}$ satisfying regularity have been found. We show that variable $A_{k,n}$ can be found such that node $A_{k,n}$ satisfies regularity, and that there are exactly two choices for this element. Node $A_{k+1,n}$ may be written $A_{k+1,n} = (A_{n,n}, a | b, c, d, \dots, e)$ with children as in Figure 2.9.

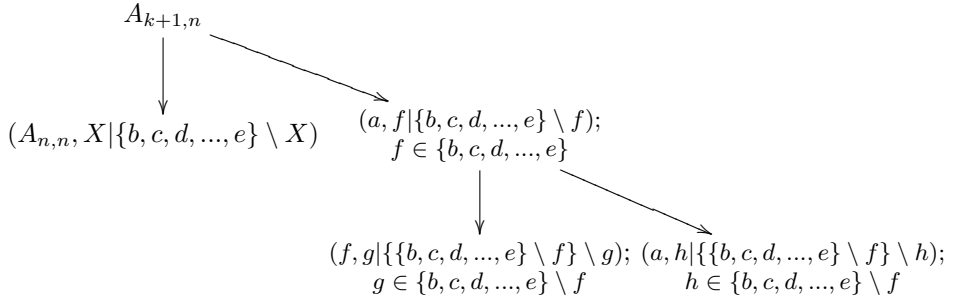


Figure 2.9: Regular Vine Growing 2.

Node $(a, f | \{b, c, d, \dots, e\} \setminus f)$ exists in the original vine $V(n-1)$ by assumption. Node $(A_{n,n}, X | \{b, c, d, \dots, e\} \setminus X)$ satisfies regularity if $X = f$ or $X = g$, either choice being possible. No other choice is possible, as no other node can have constraint set $\{b, c, d, \dots, e\}$. Note that if $k = 2$ then $\{\{b, c, d, \dots, e\} \setminus f\} \setminus g = \{\{b, c, d, \dots, e\} \setminus f\} \setminus h = \emptyset$.

It follows that for each node $A_{n-2,n}, \dots, A_{2,n}$ there is a choice among 2 alternatives. Hence there are 2^{n-3} extensions of $V(n-1)$ to a regular vine on n elements. \square

For the example from Figure 2.7 the 2^{6-3} possible extensions of the triangular array from example 2.3 are given by the tree in Figure 2.10 below. Corollary 2.4.4 follows immediately from theorem 2.4.3.

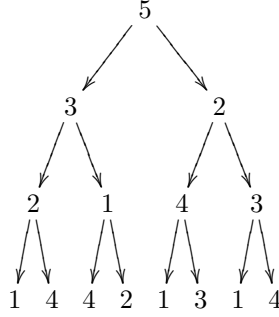


Figure 2.10: 8 Possible Extensions of $TA(V_2(5))$ in equation (2.3) Representing the Vine in Figure 2.4

Corollary 2.4.4. *The number of regular vines possible with a fixed natural order*
 $NO(n) = A_{n,n}, A_{n-1,n-1}, \dots, A_{1,1}$ *is:* $\prod_{j=1}^{n-3} 2^j = 2^{\binom{n-2}{2}}$

Proof. Start with a regular vine on three nodes with an arbitrary natural order and extend it to a regular vine on four nodes. Elements $A_{4,4}$, $A_{3,4}$ and $A_{1,4}$ are fixed by the natural order and hence only element $A_{2,4}$ may be chosen in 2 distinct ways. For each one of the 2 choices of $A_{2,4}$, from theorem 2.4.3 an extension to a regular vine on 5 nodes leaves two choices for each of the two elements $A_{3,5}$ and $A_{2,5}$. Continue this way until a regular vine on n nodes is formed and the result follows. \square

Observe that corollary 2.4.4 implies that no regular vine would be counted twice once the natural order has been fixed. Obviously two triangular arrays that are equal will represent the same vine. Once the number of regular vines that may be obtained with a given natural order is known, all that is left to know the number of regular vines on n nodes is how many natural orders are possible in order to produce all possible regular vines. Corollary 2.4.5 completes the problem of enumerating regular vines.

Corollary 2.4.5. *There are $\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$ labeled regular vines in total.*

Proof. There are $\binom{n}{2}$ ways of choosing the pair $A_{n,n}, A_{n-1,n-1}$ in a natural order and $(n-2)!$ ways of permuting elements $A_{n-2,n-2}, \dots, A_{1,1}$. By corollary 2.4.4 the proof is completed. \square

The results of corollary 2.4.5 may be observed in tables A.1 to A.7 which were obtained by the methods explained in previous sections. For example, the number of regular vines on 9 nodes is $\binom{9}{2} \times 7! \times 2^{\binom{9}{2}} = 181,440 \times 1 + 362,880 \times 69 + 362,880 \times$

$$\begin{aligned}
& 41 + 181,440 \times 13 + 181,440 \times 129 + 181,440 \times 181 + 181,440 \times 2,651 + 181,440 \times \\
& 5,390 + 90,720 \times 1,708 + 181,440 \times 1,646 + 362,880 \times 2,708 + 45,360 \times 168 + \\
& 181,440 \times 528 + 181,440 \times 887 + 181,440 \times 887 + 90,720 \times 4,202 + 181,440 \times 2,567 + \\
& 60,480 \times 528 + 181,440 \times 8,738 + 15,120 \times 18,504 + 90,720 \times 11,296 + 181,440 \times \\
& 34,417 + 45,360 \times 36,892 + 30,240 \times 72,546 + 90,720 \times 120,444 + 60,480 \times 20,904 + \\
& 181,440 \times 99,028 + 60,480 \times 34,143 + 30,240 \times 6,756 + 90,720 \times 32,812 + 90,720 \times \\
& 54,004 + 15,120 \times 32,688 + 60,480 \times 149,901 + 30,240 \times 360,084 + 30,240 \times 428,388 + \\
& 22,680 \times 680,576 + 5,040 \times 262,080 + 30,240 \times 1,232,820 + 7,560 \times 414,432 + \\
& 30,240 \times 1,919,610 + 15,120 \times 1,232,340 + 3,024 \times 1,869,120 + 7560 \times 5,255,904 + \\
& 2,520 \times 14,889,744 + 1,512 \times 23,334,480 + 504 \times 62,523,360 + 9 \times 660,602,880 = \\
& 3.8050725888 \times 10^{11}.
\end{aligned}$$

Remark. By lemma 2.3.2 and corollary 2.4.5 it may be seen that a tree with a single node with maximum degree admits $\binom{n-1}{2} \times (n-3)! \times 2^{\binom{n-3}{2}}$ regular vines.

Finally, the results of remark 2.4 may be also observed in tables A.1 to A.7. For example there are 9 trees with maximal degree on 9 nodes each of which admits $\binom{8}{2} \times (6)! \times 2^{\binom{6}{2}} = 20,160 \times 1 + 20,160 \times 11 + 40,320 \times 29 + 20,160 \times 39 + 20,160 \times 71 + 10,080 \times 820 + 5,040 \times 120 + 20,160 \times 315 + 20,160 \times 815 + 20,160 \times 423 + 5040 \times 4,520 + 6,720 \times 2181 + 10,080 \times 11,246 + 6,720 \times 315 + 20,160 \times 1,046 + 3,360 \times 3,384 + 6,720 \times 8,667 + 560 \times 89,712 + 3,360 \times 27,222 + 1,680 \times 11,160 + 840 \times 117,072 + 336 \times 279,000 + 8 \times 2,580,480 = 660,602,880$ regular vines which is exactly the total number of regular vines on 8 nodes. To finalize some conclusions are presented next.

2.5 Final Comments

This chapter investigates counting problems related to vines. Corollary 2.3.1 has been obtained from Cayley's theorem 2.2.1 to count the number of vines on n nodes. A way to efficiently code and store vines on n nodes based on the Prüfer code is proposed. This consists of an upper triangular matrix of size $(n-2) \times (n-2)$. An algorithm for building vines and two others for building regular vines on n nodes have been presented. Algorithm 2.2.1 is easy to implement and efficient if regular vines on less than 6 nodes are required. Algorithm 2.3.2 would produce only regular vines at the cost of greater programming effort and a larger number of arithmetic operations.

Table 2.1 shows the number of labeled trees, labeled vines, labeled regular vines and tree-equivalent regular vines, on up to 9 nodes. Tables A.1 to A.7 presents the number of labeled trees, regular vines per labeled tree and tree-equivalent regular vines according to unlabeled trees on n nodes. The number of ways of extending an $n-1$ vine to an n vine has been found and the number of labeled regular vines as a function of n has been presented. Future research about efficient implementation and storing of codes for producing regular vines is desirable. Vines keep a close relationship with continuous-discrete non-parametric BBNs. This will be discussed in the next chapter.

CHAPTER 3

BBNs in Aviation Safety¹

3.1 Discrete BBNs

Graphical methods for dependence modeling have become increasingly important over the past years. From the graphical methods discussed in the literature perhaps BBNs have drawn more attention from the scientific community. An overview of the development of the use of graph theory in combination with probability theory was given in chapter 1. The CATS model which is the main application driving this thesis² was briefly introduced in section 1.4. In this chapter BBNs will be presented more formally. An excellent overview of BBNs is presented in Hanea [2008]. A thorough treatment of the semantics of BBNs is presented in Pearl [1988]. The CATS model will also be explained in more detail in order to be able to show its use in risk and uncertainty analysis in later chapters.

For our purpose *Bayesian Belief Nets (BBNs)* are directed acyclic graphs whose nodes represent univariate random variables and whose arcs represent direct influences between adjacent nodes. These influences may be probabilistic or deterministic³. The graph of a BBN induces a non unique ordering of variables and stipulates that each variable is conditionally independent of its non-descendants given its parents. The parent set of variable X_i will be denoted as $Pa(i)$. Hence, to specify a joint distribution through a BBN the graph must be specified together with conditional probability functions of each variable given its parents (equation (3.1)).

$$f(X_1, \dots, X_n) = \prod_{i=1}^n f_{X_i|X_{Pa(i)}} \quad (3.1)$$

¹This chapter is based on Morales-Nápoles et al. [2008]

²And a good part of the research currently carried out in the Decision Theory group in Delft

³When an influence is deterministic, nodes will be called *functional*. The discussion presented next refers to probabilistic influences unless otherwise specified.

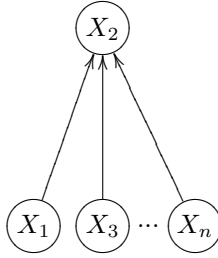


Figure 3.1: Simple example of BBN on n nodes.

If $Pa(i) = \emptyset$ then $f_{X_i|X_{Pa(i)}} = f(X_i)$. A BBN is then a concise and complete representation of the joint distribution. In the case that all nodes in the BBN are discrete then the functions to be specified are conditional probability tables (CPT) of each node given its parents. When variables are continuous, one possibility is to discretize them into a large enough number of states and use discrete BBNs. This approach might however turn out to be infeasible even for a modest sized model mainly because of the number of parameters to be specified. This idea is illustrated in example 3.1.1.

Example 3.1.1. Consider the BBN in Figure 3.1 and suppose each variable X_i has k_i states. Then $(n - 1)$ univariate marginal distributions need to be specified for each parent node of X_2 . For each of these marginal distributions $k_i - 1$ probabilities need to be assessed⁴. For X_2 a table with $k_2 \cdot k_1 \cdot k_3 \cdot \dots \cdot k_n$ conditional probabilities needs to be specified of which $k_1 \cdot k_3 \cdot \dots \cdot k_n$ are constrained by the choice of the other conditional probabilities.

In particular suppose that for the BBN in Figure 3.1 $n = 3$, $k_1 = 2$, $k_2 = 3$ and $k_3 = 3$. The states of each X_i will be $1, \dots, k_i \forall i$ then a table such as 3.1 would be required for X_2 . One cell in each row of table 3.1 is fixed by the requirement that the values in each row must sum to one. Additionally to the 12 probabilities to be assessed for X_2 , one state of X_1 and 2 of X_3 would need to be specified. \square

	$P(X_2=1 X_1=x_1, X_3=x_3)$	$P(X_2=2 X_1=x_1, X_3=x_3)$	$P(X_2=3 X_1=x_1, X_3=x_3)$
$x_1=1, x_3=1$			
$x_1=1, x_3=2$			
$x_1=1, x_3=3$			
$x_1=2, x_3=1$			
$x_1=2, x_3=2$			
$x_1=2, x_3=3$			

Table 3.1: Conditional probability table for X_2 in example 3.1.1.

In general, the number of probabilities to be assessed for a discrete BBN on n nodes with k_i states for each X_i for $i = 1, \dots, n$ is:

⁴One probability is of course determined once the others have been.

$$K = \sum_{j \in S} k_j - |S| + \sum_{l \in C} (k_l - 1) \prod_{m \in Pa(l)} k_m \quad (3.2)$$

where $S = \{X_j | Pa(j) = \emptyset\}$ and $C = \{X_l | Pa(l) \neq \emptyset\}$ and $|S| + |C| = n$.

One of the main advantages of BBNs is that they possess a graphical representation which makes them appealing for applications. Another property of discrete BBNs that makes them attractive for practitioners is that once it has been quantified, it may be used to update the joint distribution when evidence becomes available. Exact algorithms and approximation algorithms are available for this purpose. See for example Lauritzen and Spiegelhalter [1988], Cowell et al. [1999, p.123] and Pearl [1993, p.55].

It is clear from equation (3.2) that K grows rather quickly as the number of states of each X_i grow. This is one of the main drawbacks of discrete BBNs. Some of the drawbacks of discrete BBNs were discussed in Hanea et al. [2006] and Cowell et al. [1999]. We list a summary of them next:

1. K imposes an assessment burden that might lead to informal and indefensible quantification or a drastic discretization or reduction of the model.
2. Marginal distributions are often available from data. Marginal distributions for children nodes are calculated from probability tables and this could impose severe restrictions in a quantification process.
3. Discrete BBNs are flexible with respect to recalculation and updating however they are not flexible with respect to modelling changes. If a parent node is added then the child nodes must be completely re-quantified.

Continuous-discrete non-parametric BBNs (Kurowicka and Cooke [2005], Hanea et al. [2006]) have been developed to cope with some of the drawbacks that discrete (and discrete-normal) models impose. These will be discussed next.

3.2 Non-Parametric Continuous-Discrete BBNs.

Another way to deal with continuous nodes in a BBN is with the use of normal [Schachter and Kenley, 1989] or discrete-normal BBNs. For discrete-normal BBNs [Cowell et al., 1999], unconditional means and conditional variances must be assessed for each normal variable. For each arc partial regression coefficients must be assessed. In the absence of data the assessment of partial regression coefficients and conditional variances by experts is difficult if the normality assumption does not hold. More flexible models will be discussed in this section for dealing with continuous nodes.

Vines and BBNs represent a joint distribution specified by marginal distributions and conditional bivariate dependence statements. One advantage of BBNs versus vines is that the former preserve the intuitive representation of influence diagrams. This section describes the relationship between vines and non-parametric BBNs.

The graphical objects discussed in chapter 2 are used in dependence modelling. The nodes of the vine represent random variables with invertible distribution function and the edges may be used to specify conditional bivariate dependencies.

Each edge in the regular vine may be associated with a conditional rank correlation. In general these conditional rank correlations may depend on the values of the conditioning nodes, but in the present implementation, all conditional rank correlations are constant. All assignments of rank correlations to edges of a vine are consistent and each one of these correlations may be realized by a copula. A regular vine enables the construction of a joint distribution from bivariate and conditional bivariate distributions.

The reader may see in Kurowicka and Cooke [2006] how to sample a joint distribution represented by a D-vine in 4 nodes. At this point it may also be observed that a Markov-Dependence Tree is a special case of a vine where all conditional rank correlations are set to zero. In other words, in a Markov-Dependence tree the random variables that are not joined by an edge in the tree are conditionally independent given variables on the path between them. It may be observed that vines relax the assumptions about conditional independence for Markov-Dependence trees to allow for conditional dependence.

If one chooses the normal copula to realize the (conditional) rank correlations assigned to the edges of a regular vine and the marginal distributions are standard normal, then we call such vine the *standard normal vine*. The standard normal vine gives us a very convenient way of specifying a standard joint normal distribution by specifying $\binom{n}{2}$ algebraically independent numbers from $(-1, 1)$. This is in contrast to the specification of a correlation matrix that must satisfy the constraint of positive definiteness [Bedford and Cooke, 2002].

Example 3.2.1. Let us consider a standard normal D-vine on three standard normal variables and assume that the following rank correlations were specified: $r_{2,1}$, $r_{3,2}$ and $r_{3,1|2}$. The correlation matrix of the joint normal distribution corresponding to this normal vine can be calculated as follows:

- Let $\rho_{2,1}$, $\rho_{3,2}$ and $\rho_{3,1|2}$ be the product moment correlations obtained by applying equation (1.2) to $r_{2,1}$, $r_{3,2}$ and $r_{3,1|2}$ respectively.
- Since for the normal distribution partial correlation is equal to conditional correlation $\rho_{3,1|2} = \rho_{3,1;2}$, then from equation (1.1) we can compute $\rho_{3,1}$ as:

$$\rho_{3,1} = \rho_{3,1|2} \cdot ((1 - \rho_{2,1}^2)(1 - \rho_{3,2}^2))^{1/2} + \rho_{2,1}\rho_{3,2}. \quad \square$$

Non-parametric BBNs and their relationship to vines were presented in Kurowicka and Cooke [2005] and extended in Hanea et al. [2006]. A *non-parametric continuous-discrete BBN (NPCDBBN)* is a directed acyclic graph whose nodes represent continuous univariate random variables and whose arcs are associated with parent-child (un)conditional rank correlations. For each variable X_i with parents $X_j, \dots, X_{Pa(i)}$ associate the arc $X_{Pa(i)-k} \rightarrow X_i$ with the conditional rank correlation:

$$\begin{cases} r_{i,Pa(i)}, & k = 0 \\ r_{i,Pa(i)-k|Pa(i),\dots,Pa(i)-k+1}, & 1 \leq k \leq Pa(i) - 1 \end{cases} \quad (3.3)$$

The assignment is vacuous if $\{X_j, \dots, X_{Pa(i)}\} = \emptyset$. These assignments together with a copula family indexed by correlation and with conditional independence statements embedded in the graph structure of a BBN are sufficient to construct a unique joint distribution. Moreover, the conditional rank correlations in 3.3 are algebraically independent, hence any number in $(-1,1)$ can be attached to the arcs of a NPCDBBN. In Figure 3.2 it may be seen that variables X_1, \dots, X_5 are independent of each other and their dependence with variables X_6 and X_7 is described in terms of (conditional) rank correlations. Variables X_6 and X_7 are conditionally independent given X_1, \dots, X_5 .

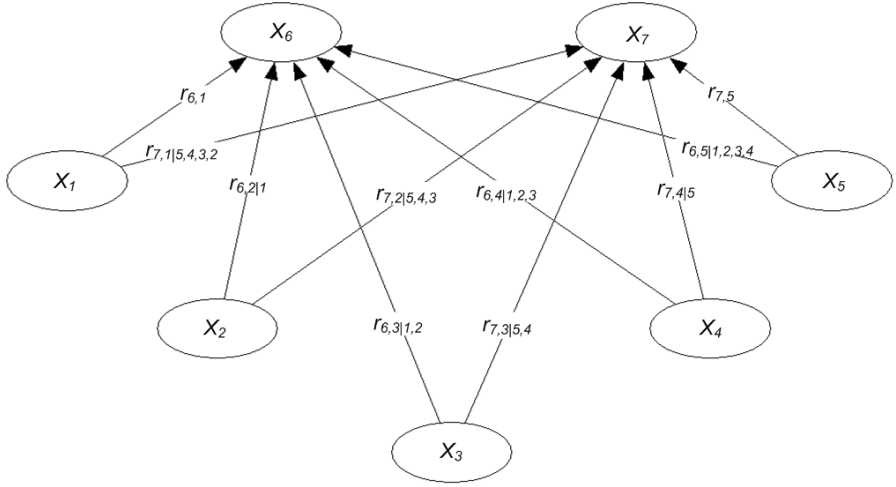


Figure 3.2: A BBN on 7 variables.

One can use the copula-vine approach [Kurowicka and Cooke, 2006] to represent the multidimensional joint distribution specified by a BBN (Kurowicka and Cooke [2005], Hanea et al. [2006]). D-Vines become an important instrument as the sampling procedure for a BBN is based on the sampling procedure for a D-vine. Some BBNs might not be represented as a single D-vine in their sampling order and it might be necessary to perform extra calculations [Hanea et al., 2006, p.716].

Any copula with an ‘easy-to-compute’ invertible conditional cumulative distribution function may be used as long as the chosen copula possesses the zero independence property⁵. Choosing the normal copula presents advantages with respect to other copulae for building the joint distribution. Observe that for the normal copula relation 1.2 holds and since conditional correlations are equal to partial correlations then a procedure similar to example 3.2.1 may be applied in the graph. Moreover since for the joint normal distribution, conditional distribu-

⁵A copula with an analytic form for the conditional and inverse conditional cumulative distribution function accelerates the sampling procedure. One example of such a copula is Frank’s copula presented in section 1.2.

tions are also normal [Tong, 1990, p.33], then analytical updating is possible by this choice [Hanea et al., 2006, p.724].

The NPCDBBN representing the CATS model (section 1.4) was implemented in UNINET [Morales-Nápoles et al., 2007] [Cooke et al., 2007]. The next section explains in more detail the implementation.

3.3 Causal Model for Air transport Safety (CATS)

As mentioned in section 1.4, the CATS model integrates ESDs, FTs, and BBNs into one single CDNPBBN. This section is devoted to a summary description of the procedure to build up the CATS model. Some of the results presented in this section are taken from Morales-Nápoles et al. [2008]. The three human reliability models that will be introduced in subsection 3.3.2 are represented for 3 different flight phases; these are Take Off (TO) En-Route (ER) and Approach & Landing (AL). In Spouge and Vernon [2008] these flight phases are considered for building up the FTs that are attached to the ESDs. The definitions of flight phases used here are equal to those in Spouge and Vernon [2008].

3.3.1 Event Sequence Diagrams & Fault Trees

An event sequence diagram is a flow chart showing a sequence of events whose happening or not happening lead to different *end states*. ESDs for the CATS model have been quantified in Roelen et al. [2007]. Since the sequence of intermediate events that must happen in order to observe the end state may be represented by logical statements, ESDs may be represented as Fault Trees. In the CATS model ESDs and FTs were modelled as a single unit. FT analysis is considered a technique which allows the analysis of a system in the context of its environment and operation to find the largest number of credible ways in which an undesired state of a system may happen [Vesely et al., 1981, p.IV-1].

Basic events in Fault Trees can be represented by a Boolean variable. In this sense, a FT may be thought of as a picture of a Boolean formula. In Boolean arithmetic, variables take only two values, usually 0 or 1. Suppose A_1 and A_2 are Boolean variables, then $A_1 \oplus A_2$ and $A_1 \otimes A_2$ are also Boolean variables, and hence take values 0 or 1. This is arranged by defining $A_1 \oplus A_2 = A_1 + A_2 - A_1 \cdot A_2$ and $A_1 \otimes A_2 = A_1 \cdot A_2$. The operators \oplus and \otimes correspond to the **AND** and **OR** operators in propositional logic. In other words, $A_1 \oplus A_2$ means “either A_1 or A_2 or both are true” and $A_1 \otimes A_2$ means “both A_1 and A_2 are true”. In Boolean arithmetic $A_1 \otimes A_1 = A_1$; this corresponds to saying that the event A_1 AND A_1 is the same as the event A_1 ⁶. Two examples of FTs are presented in Figures 3.3 and 3.4. Their usual notation for AND and OR gates is also displayed.

In most cases, we don’t know whether a given basic event occurs, we know only its probability of occurrence. Recall that $E(A_1) = P(A_1 = 1)$ where E denotes mathematical expectation. If A_1 and A_2 are independent, then $E(A_1 \otimes A_2) = E(A_1)E(A_2)$ and $E(A_1 \oplus A_2) = E(A_1) + E(A_2) - E(A_1 \cdot A_2)$.

⁶Other rules for Boolean algebra may be found in [Vesely et al., 1981, p.VII-2]

The above reasoning might suggest that we can just replace the Boolean variables at the base of a fault tree with their probabilities (i.e. their expectations) and compute the probability of the top event with ordinary arithmetic. This is not true, in general, and it may depend on how the fault tree is displayed. This is illustrated in Figures 3.3 and 3.4. Suppose that A_4 occurs when either $(A_1 \text{ AND } A_2) \text{ OR } (A_1 \text{ AND } A_3)$ occur. Observe that Figures 3.3 and 3.4 are logically equivalent fault trees.

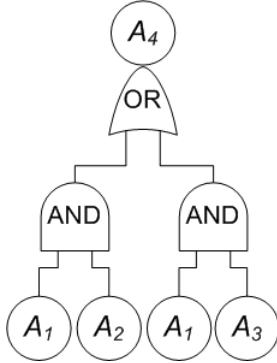


Figure 3.3

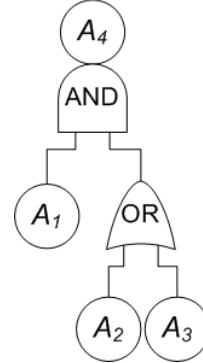


Figure 3.4

If Boolean reduction is applied to both FTs we see that they yield the same formula for A_4 . However, if we replace the variables by their expectations and apply ordinary arithmetic to the non-reduced formulae, we will get different answers. The correct calculation would be obtained from Figure 3.4 $P(A_4 = 1) = P(A_1 = 1)P(A_2 = 1) + P(A_1 = 1)P(A_3 = 1) - P(A_1 = 1)P(A_2 = 1)P(A_3 = 1)$. The problem is that when $(A_1 \text{ AND } A_2) \text{ OR } (A_1 \text{ AND } A_3)$ is computed with expectations in Figure 3.3, the term $P(A_1 = 1)$ is included twice.

In general, computing probabilities of occurrence from a Fault Tree requires some careful manipulations, before substituting Boolean variables with their expected values. However, if our Fault Trees contain no “repeated events” then we can replace Boolean variables with expectations and replace Boolean arithmetic with ordinary arithmetic. This assumes that we have captured all common cause dependencies in the Fault Tree. This means that once probabilities are assigned to the basic events, the probability of joint occurrence is computed as the product of the probabilities. In the CATS model no repeated events exist and hence we can simply replace basic events with expectations and compute with ordinary arithmetic.

The quantification of ESDs presented in Roelen et al. [2007.] is used in Spouge and Vernon [2008] to quantify the FTs that later compute the accident probability. The FTs (and consequently ESDs) presented in the appendix **DNV Collected Fault Trees (3Feb09) v7,1.xls** of Spouge and Vernon [2008] are translated into functional nodes in UNINET using ordinary arithmetic.

Translating FTs into BBNs is not new. In Bobbio et al. [1999] and Bobbio et al. [2001] FTs are translated into discrete BBNs. Our approach is different in the

sense that uncertainty analysis is carried out by sampling the probability of each base event in the FTs from a distribution (see section 3.3.3.2). The expectation of each base event distribution is the probability estimated originally for the FTs in Spouge and Vernon [2008]. By sampling each base event probability from a given distribution and computing the arithmetic operations at each level of the FT a distribution is obtained for the accident probability. Samples for each base event can be generated with some dependence structure. Most of the base events of the FTs are a result of human errors. Factors influencing human performance might induce the dependence structure for base events of the FTs. For this reason human reliability models (HRM) will be introduced next. The issue of the quantification of dependence in each HRM will be dealt with in chapter 4. In section 3.3.3 the connection between base events of the FTs and HRM will be made explicit.

3.3.2 Human Reliability Models

To a large extent, events initiating accident scenarios in the CATS model are a result of incorrect performance of humans. Models for taking into account the probability of human errors have been developed for Flight Crew Performance (FCP), Air Traffic Controller Performance (ATCP) and Maintenance Crew Performance (MNTP). These are discussed next. The quantification of the models presented in this section include field data whenever available and structured expert judgment (SEJ). Structured expert judgment is a process intended to use expert opinion in a transparent way with the purpose of treating expert judgments as scientific data [Cooke, 1991]. SEJ will be dealt with in more detail in chapter 4. In this section the models are introduced together with the data source for marginal distributions. Rank and conditional rank correlations were retrieved through SEJ in all models presented in this thesis. Techniques for eliciting such measures are also introduced in chapter 4.

3.3.2.1 FCP Model Description

The FCP model is shown in Figure 3.5. The model is described extensively in Morales-Nápoles et al. [2009b] and Roelen et al. [2007]. Variables taken into account for this model are briefly described in table 3.2 according to their labeling in Figure 3.5.

The basis for the quantification of each marginal distribution is presented in column 3. Four variables were elicited through structured expert judgment⁷ and the rest come from data. Node 14 would represent a base event in DNV's Fault Trees. Whenever the flight crew performance is of interest in the FTs an instance of node 14 will appear in the CATS model.

Each node of the BBN in Figure 3.5 shows the marginal distribution of the variables listed previously. The mean of the distribution (and the standard deviation after the \pm sign) of each variable are shown at the bottom of each node. An elicitation protocol was designed for obtaining the marginal distributions from Figure 3.5 shown in table 3.2 and the dependence information (rank and conditional

⁷See section 4.2 for an overview of structured expert judgment.

Node #	Definition	Marginal distribution.
1	Total number of hours flown (all types) for the First Officer	Data
2	Number of days since the last type recurrent training for the First Officer	Data
3	Stanford Sleepiness Scale	Data
4	Number of days since the last type recurrent training for the Captain	Data
5	Total number of hours flown (all types) for the Captain	Data
6	Likelihood that the Captain fails a proficiency check	SEJ ^a
7	Likelihood that the First Officer fails a proficiency check	SEJ
8	Rainfall rate in mm/hr	Data
9	Difference in mother tongue between Captain and First Officer per 10000 flights	SEJ
10	Likelihood that the Captain or the First Officer fail a proficiency check	SEJ
11	Aircraft generation: 1, 2, 3, or 4	Data
12	Likelihood that the flight crew needs to follow a procedure of the abnormal/emergency procedures section of the AOM ^b	Data
13	Total duration (in seconds) of the air/ground communications, per aircraft, for the approach and landing flight phase.	Data
14	Likelihood that the flight crew makes an unrecovered error that is potentially hazardous for the safety of the flight.	FT ^c

Table 3.2: *Description of variables from the model in Figure 3.5.*

^aStructured Expert Judgment

^bAircraft operations manual

^cFrom the associated Fault Tree quantified by DNV

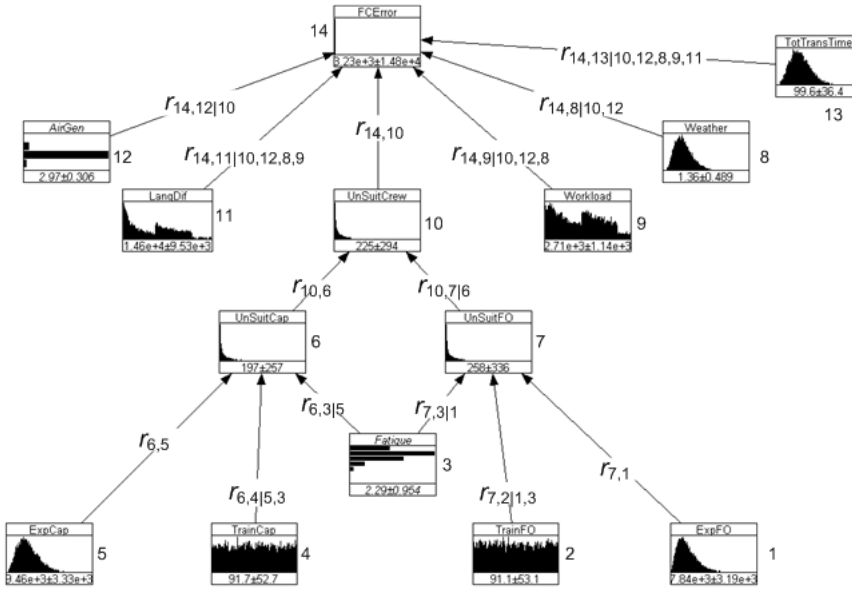


Figure 3.5: *Flight Crew Performance Model*

rank correlations) required by the model. The elicitation of rank and conditional rank correlations will be discussed in section 4.2.2 and results for the FCP model elicitation will be presented in section 5.2.

3.3.2.2 ATPC Model Description.

The Air Traffic Control Performance model (ATCP) is the second one of the generic models that has been developed to represent dependence between base events in the FTs of the CATS model. The model is discussed in details in Morales-Nápoles et al. [2009b] and Roelen et al. [2008a]. Figure 3.6 shows the BBN representing the model. Variables 1-6 are considered to be correlated to ATC error probability (variable 7) and independent of each other. Each variable considered in the model is briefly described in table 3.3 according to its labeling in Figure 3.6.

The basis for the quantification of each marginal distribution is presented in column 3. Five variables come from data and the error distribution from the quantification of FTs. Node 7 would represent base events in the Fault Trees. As with node 13 in the FCP model, whenever the air traffic error is of interest an instance of node 7 will appear in the CATS model. Results for the ATPC model elicitation will be presented in section 5.3.

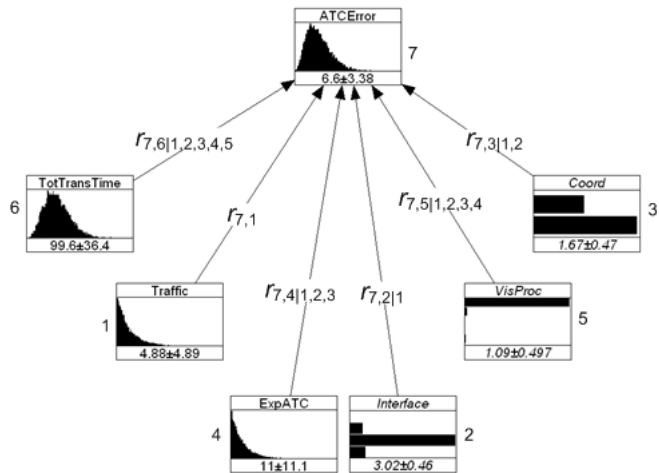


Figure 3.6: Air Traffic Controller Performance Model

Node #	Definition	Marginal distribution.
1	Number of aircraft (any type) simultaneously under control.	Data
2	Four states variable. From 1- using radio only to 4-using radio, primary and secondary radar and additional tools.	Data
3	Two states. 1 - The communication with other ATC takes place in the same room, 2 - The communication with other ATC does not take place in the same room	Data
4	Number of years working as an ATC in the same position.	Data
5	Five states variable. From 1 - normal operations to 5 - operations below 200 meters visibility.	Data
6	Total duration (in seconds) of the air/ground communications, per aircraft, for the approach and landing flight phase.	Data
7	Likelihood that the air traffic control makes an unrecovered error that is potentially hazardous for the safety of the flight.	FT ^a

Table 3.3: Description of variables from the model in Figure 3.6.

^aFrom the associated Fault Tree quantified by DNV

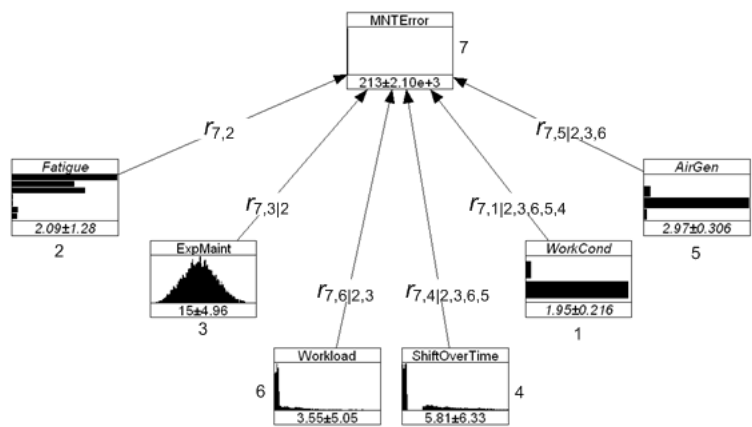


Figure 3.7: Maintenance Crew Performance Model

Node #	Definition	Marginal distribution.
1	Whether the work is performed at the ramp (outside - 1) or in the hangar (inside - 2)	SEJ ^a
2	Stanford Sleepiness Scale	SEJ
3	# of years in current position	Data
4	Time available to transfer a job (min)	SEJ
5	Aircraft generation: 1, 2, 3, or 4	Data
5	Five states variable. From 1 - normal operations to 5 - operations below 200 meters visibility.	Data
6	Estimated delay in release of the aircraft (hrs)	SEJ
7	Likelihood that the maintenance crew makes an unrecovered error that is potentially hazardous for the safety of the flight.	FT ^b

Table 3.4: Description of variables from the model in Figure 3.7.

^aStructured Expert Judgment

^bFrom the associated Fault Tree quantified by DNV

3.3.2.3 MNTP Model Description.

The maintenance crew performance model is the third and last of the generic models that have been developed to represent dependence between base events in the FTs of the CATS model. A preliminary version of the model presented here may be found in Jagielska [2007]. The MNTP model is discussed in the context of CATS in Krugła [2008] and Roelen et al. [2008b]. The model is shown in Figure 3.7 and the variables taken into account are briefly described in table 3.4 according to their labeling in Figure 3.7. Equivalently with node 13 in the FCP model and 7 in the ATPC model if a maintenance technician error is of interest an instance of node 7 will appear in the CATS model.

The three HRM briefly introduced here and the FTs in Spouge and Vernon [2008] are introduced in UNINET to build the CATS model. That process is described next.

3.3.3 The CATS Model in UNINET

3.3.3.1 ESDs & FTs for the CATS model in UNINET

In Figure 3.8 one may see ESD1 *aircraft system failure* for the TO flight phase as presented in appendix **DNV Collected Fault Trees (3Feb09) v7,1.xls** of Spouge and Vernon [2008]. In total four AND gates and four OR gates represent the FT and ESD. Fifteen base events are influenced by the MNTP model presented in section 3.3.2.3 and 3 base events by the FCP model from section 3.3.2.1. No influence of the ATPC model is observed in this particular FT. To translate this information into a BBN the process is:

1. Find a distribution of the probability of base events per demand according to their variability. The variability of each base event in this case corresponds to the different percentiles and expectation from **DNV Collected Fault Trees (3Feb09) v7,1.xls** of Spouge and Vernon [2008]. These correspond to the expectation, minimum, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 99th and maximum percentiles of each base event distribution. This information is treated as data (see Figure 3.9).
2. Connect with incoming arcs each base event to the corresponding dependence model from subsections 3.3.2.1 to 3.3.2.3 using the corresponding rank and conditional rank correlations. These nodes will be ancestors of base events in the Fault Trees.
3. Write in descendent nodes of each base event the arithmetic formulae that translate a FT into a BBN (subsection 3.3.1). These will be functional nodes in the BBN.

These steps are repeated for all ESDs presented in table 3.5.

3.3.3.2 The error distributions in UNINET

Base events in the FTs are influenced by the human performance models introduced in section 3.3.2. In fact, the probabilities presented in the FTs [Spouge and



Figure 3.9: *Distribution of base event probability*

ESD	Initiating event	Flight Phase
1	Aircraft system failure	TO
2	ATC event	TO
3	Aircraft handling by flight crew inappropriate	TO
4	Aircraft directional control related systems failure	TO
5	Operation of aircraft systems by flight crew inappropriate	TO
6	Aircraft takes off with contaminated wing	TO
7	Aircraft weight and balance outside limits	TO
8	Aircraft encounters performance decreasing windshear after rotation	TO
9	Single engine failure	TO
10	Pitch control problem	TO
11	Fire on board aircraft	ER
12	Flight crew member spatially disorientated	ER
13	Flight control system failure	ER
14	Flight crew incapacitation	ER
15	Anti-ice system not operating	ER
16	Flight instrument failure	ER
17	Aircraft encounters adverse weather	ER
18	Single engine failure	ER
19	Unstable approach	AL
21	Aircraft weight and balance outside limits	AL
23	Aircraft encounters windshear during approach/landing	AL
25	Aircraft handling by flight crew during flare inappropriate	AL
26	Aircraft handling by flight crew during roll inappropriate	AL
27	Aircraft direction control related systems failure	AL
28	Single engine failure	AL
29	Thrust reverser failure	AL
30	Aircraft encounters unexpected wind	AL
31	Aircraft are positioned on collision course	ER
32	Incorrect presence of aircraft/vehicle on runway in use	TO/AL
33	Cracks in aircraft pressure cabin	ER
35	Flight crew decision error/operation of equipment error	AL
36	Ground collision imminent	TO/AL
37	Wake vortex encounter	ER

Table 3.5: ESDs used in the CATS model.

Vernon, 2008] represent the expected probability of a given human error. Other percentiles over the distribution of error probability are also presented in appendix **DNV Collected Fault Trees (3Feb09) v7,1.xls** of Spouge and Vernon [2008]. An example of the data is presented in Figure 3.9.

Figure 3.9 shows that the minimum value that the base event probability TO01B11 can take is 0. In the same way the maximum value of TO01B11 should be equal to 6.11×10^{-5} . Other quantiles may be read in the same way. This information is used to fit a parametric distribution to represent the distribution over the base event probability. In total there are 856 basic events over all thirty five ESDs from table 3.5 in the model as presented in Figure 1.13.

With these percentiles a minimally informative distribution with respect to the log uniform measure may be found. This distribution will always comply with the percentiles provided by DNV, however, it was decided that a parametric distribution would be fit to the data provided by DNV. A parametric distribution is desired for the following reasons:

- The model is easier to maintain. The minimally informative distribution requires storing the whole distribution while a parametric distribution would require storing only a number of parameters to completely describe the distribution.
- The minimally informative solution fitted to the quantiles exemplified in Figure 3.9 will not in general preserve the expectation provided by DNV.
- The model was required to have the functionality that by specifying a mean different than the one computed by DNV and keeping the variance constant, a new distribution within a given parametric family could be obtained.

The fitting procedure to obtain the parametric distribution for each base event is described briefly next. Denote by X_i , $i = 1, \dots, 856$ the random variable described by the parametric distribution required by each base event in the FTs of the CATS model. The $m = 1, \dots, 10$ observations obtained in **DNV Collected Fault Trees (3Feb09) v7,1.xls** of Spouge and Vernon [2008] (Figure 3.9) will be denoted as $\tilde{x}_{i,q_k,m}$ for the k^{th} percentile of base event i .

Algorithm 3.3.1. Finding parametric distribution for base events in the FTs.

1. For $i = 1, \dots, 856$ find the parameters of $F_j(X_i)$ where each $j = a, \dots, g$ corresponds to one of the 7 subitems given below.
 - (a) Weibull with shift parameter equal zero,
 - (b) Weibull with shift parameter non zero,
 - (c) Gamma with shift parameter equal zero,
 - (d) Gamma with shift parameter non zero,
 - (e) Beta with parameters (0,1),
 - (f) Beta with parameters $(\tilde{x}_{i,q_{min}}, \tilde{x}_{i,q_{max}})$,
 - (g) Log-normal

such that:

- (i) $\sum_m (F_j^{-1}(q_{k,m}) - \tilde{x}_{i,q_{k,m}})^2$ is minimal
 - (ii) $\left| (E_j(X_i) - E(\tilde{X}_i)) / E(\tilde{X}_i) \right| < 1$ and,
 - (iii) $F_j^{-1}(0.9999999999999999) \leq \tilde{x}_{i,q_{max}}$.
2. Select j such that $\sum_m (F_j^{-1}(q_{k,m}) - \tilde{x}_{i,q_{k,m}})^2$ is minimal.

After applying algorithm 3.3.1 to the data no distribution of the class (a) was selected, 2 of class (b), 1 of (c) and 2 of (d) were selected. More than 99% of the base events are within classes (e) to (g). In particular a Beta with parameters (0,1) was used for 58 base events, a Beta with parameters $(\tilde{x}_{i,q_{min}}, \tilde{x}_{i,q_{max}})$ for 729 events and finally a log-normal distribution was selected for 2 base events. Whenever algorithm 3.3.1 would not find a solution, a constant equal to $E(\tilde{X}_i)$ was used. This was the case in 62 base events.

Finally, to illustrate differences in the outcomes of the fitting procedure described in algorithm 3.3.1, Figures 3.10 and 3.11 are presented. Figure 3.10 corresponds to the 5th percentile of the distribution of minimal sum of squared differences. The solution corresponds to a uniform distribution in $[1.2 \times 10^{-5}, 3.2 \times 10^{-5}]$. Figure 3.11 corresponds to the 99th percentile of the same distribution of minimal sum of squared differences. In this case algorithm 3.3.1 finds a distribution that complies with the mean specified in **DNV Collected Fault Trees (3Feb09) v7.1.xls** of Spouge and Vernon [2008]. However other percentiles are not well captured by the solution. This is due to the fact that according to DNV's data 50% of the mass is concentrated in zero and 10% in one. Though the sum of squared difference is not directly comparable across base events, Figures 3.10 and 3.11 are presented for illustration purposes.

3.3.3.3 The complete model

Once a distribution for each of the 856 base events of interest has been found, the next step is to attach the adequate dependence information between base events. From Figure 3.8 it may be observed that 18 base events represent an instance of either the FCP or the MNTP models from section 3.3.2.

Figure 3.12 shows ESD1 represented as a BBN in UNINET with a single instance of the FCP model. However, according to Figure 3.3.2 there are in total three base events in ESD1 influenced by the flight crew performance. Also from Figure 3.3.2 it may be observed that the FCP model is not the only one influencing basic events in ESD1. The fifteen base events influenced by the MNTP model should also be included in the BBN representation of ESD1. The complete representation of ESD1 is shown in Figure 3.13.

If the same process is repeated for ESD2 the model should look as in Figure 3.14. This process has to be repeated for the 35 ESDs from table 3.5. The reader should observe that some nodes change through flight phases and some do not. For example, experience in the ATPC and FCP models is consider not to change

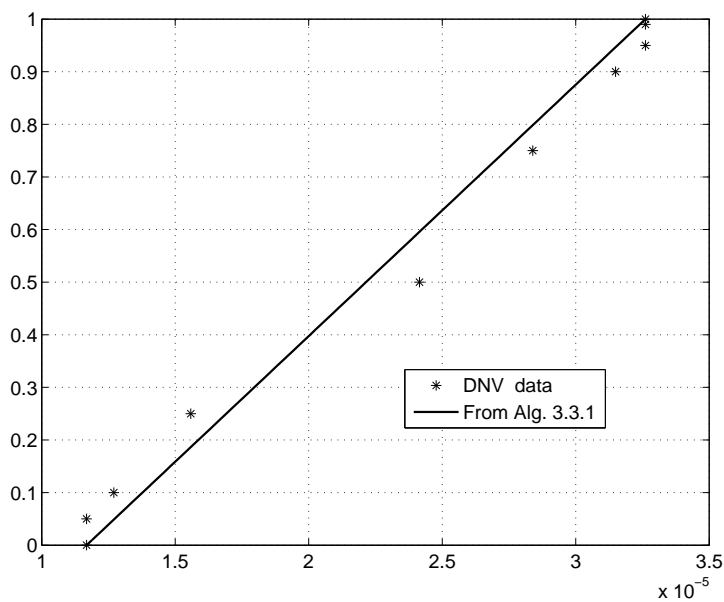


Figure 3.10: Fit of event ‘AL30B31’: No input to controls will allow the flight crew to maintain control of the aircraft per encounter with unexpected wind. $Beta(1,1,1.2\times10^{-5},3.2\times10^{-5})$, sum of squared difference = 1.1951.

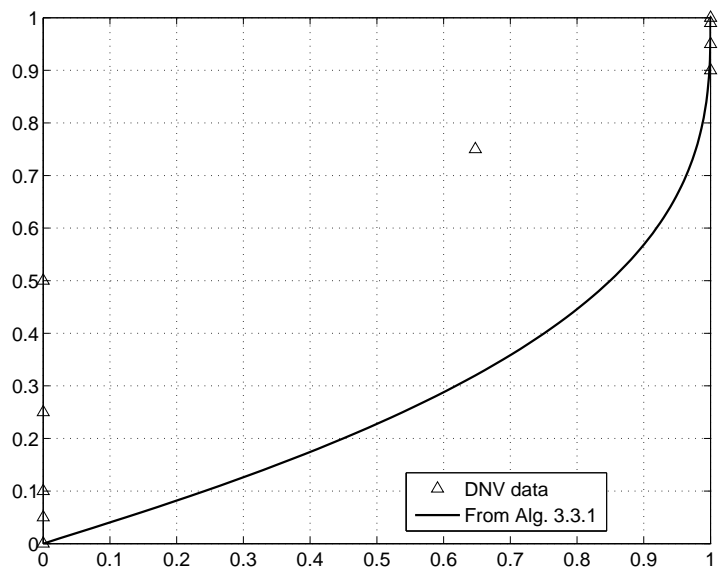


Figure 3.11: Fit of event ‘AL32B112’: ATC fails to detect a conflict and give warning due to darkness per ineffective conflict warning. $Beta(0.97,0.36,0,1)$, sum of squared difference = 1.1951).

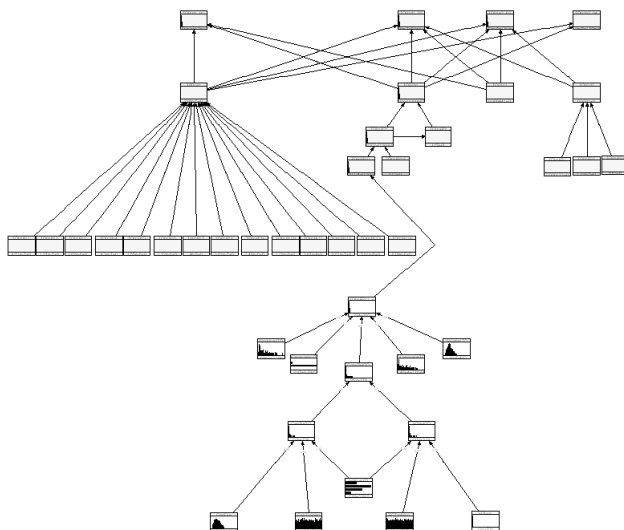


Figure 3.12: *ESD1* with a single instance of the FCP model attached to it.

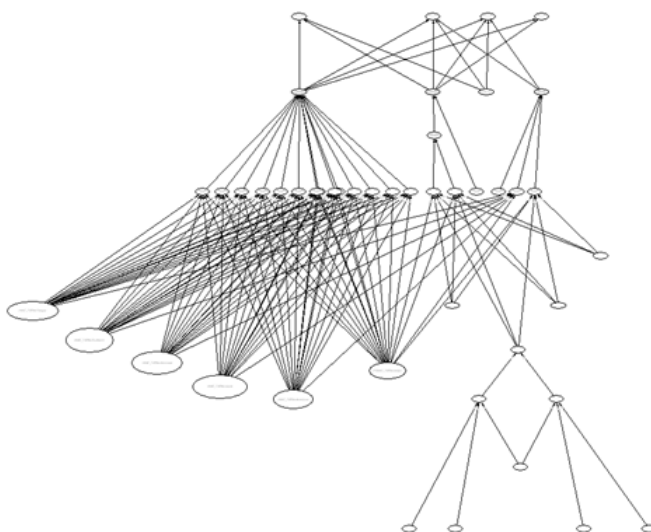


Figure 3.13: *ESD1* with 3 instances of the FCP model and 15 of the MNTP model attached to it.

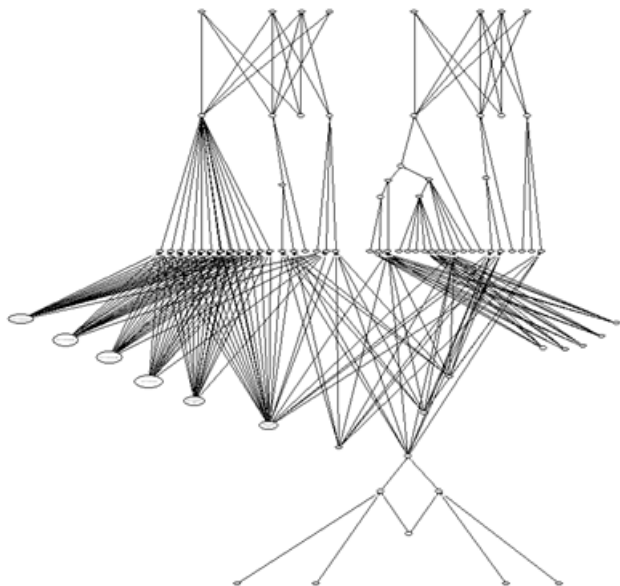


Figure 3.14: *ESD1 & ESD2 with HRMs attached.*

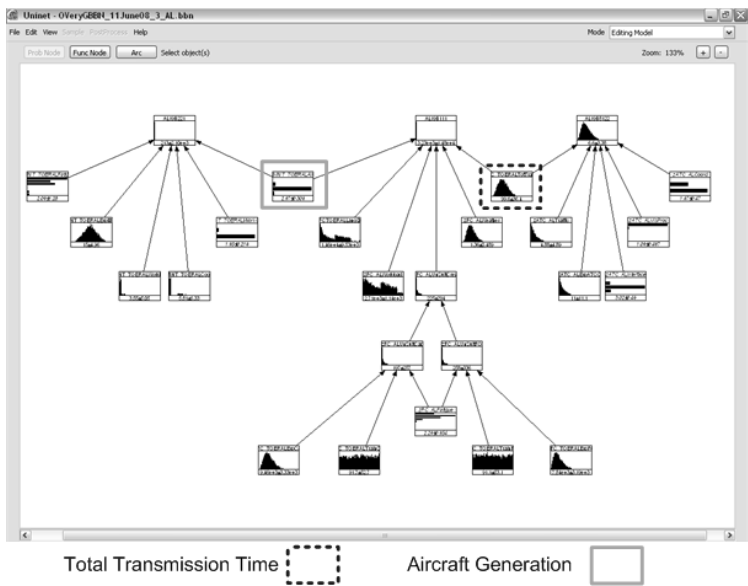


Figure 3.15: *Common nodes in the HRMs.*

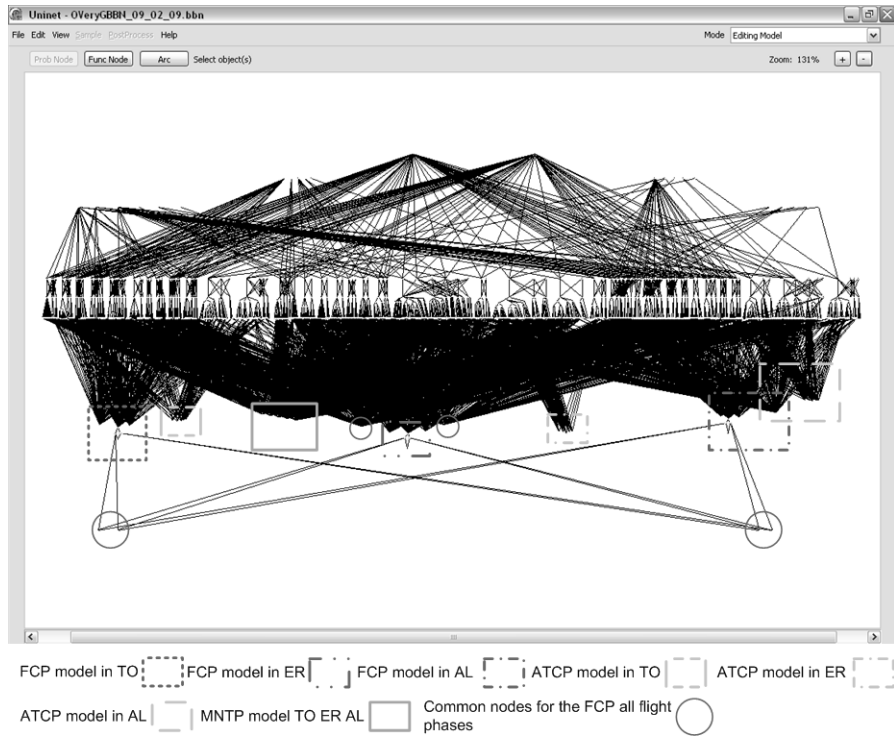


Figure 3.16: *Human performance models in CATS.*

across flight phases. On the other hand the FCP model would have one instance of weather per flight phase.

In the CATS model the FCP model and the ATPC model share in common the total transmission time in the AL flight phase. The MNTP model and the FCP model share the aircraft generation node in all flight phases. This situation is summarized in Figure 3.15.

The complete CATS model is presented in Figure 3.16. Figure 3.16 is identical to Figure 1.13 except that the human performance models used at each flight phase are indicated. The model has been integrated with the methods described in this chapter. At the date of publication of this document, the model consists of 918 probabilistic nodes, 586 functional and 4,979 arcs. Obtaining the dependence information for the models presented in sections 3.3.2.1 to 3.3.2.3 was one of the most challenging tasks in the model. Next chapter explains the methodology followed for that purpose. In next section, examples of the use of the model will be given.

3.4 Model Use

At this writing, the model presented in section 3.3.3 is still under development. However, a smaller version of the model with 834 probabilistic nodes, 532 functional and 4,756 arcs has been used [Morales-Nápoles et al., 2008]. The model in Morales-Nápoles et al. [2008] keeps the same structure as Figures 1.13 and 3.16 except that it does not include ESDs 36 and 37. In this section the version of the model from Morales-Nápoles et al. [2008] will be used. According to the model the accident probability is obtained as the expectation of the accident rate. The expectation of the baseline case is 3.18×10^{-6} . This is in line with the worldwide data tendency (see Figure 1.10). The 5^{th} and 95^{th} percentiles of the baseline accident distribution are 8.58×10^{-8} and 8.97×10^{-6} respectively.

The model was sampled in UNINET. The sample rank correlation was computed between the accident rate and each of the 45 nodes representing the HRMs from section 3.3.2 used in the CATS model as explained in section 3.3.3. The rank correlations were obtained using UNISENS [Next-Page-Software, 2009]. Results are presented in table 3.6. Other sensitivity measures may also be obtained with UNISENS [Lewandowski et al., 2007].

From table 3.6 it may be seen that variables from the FCP and the MNTTP model are most highly correlated to accident probability. Crew unsuitability is influenced by captain's and first officer's unsuitability. These are influenced in turn by experience, training and fatigue (Figure 3.5). Experience and Training are considered not to change across flight phases however fatigue does and hence 3 instances of crew unsuitability are used in the model (Figure 3.16). Crew unsuitability for each flight phase appears to be most highly correlated with accident probability ($\cong 0.3$).

Aircraft generation does not change across flight phases and is a common node for both the FCP and the MNTTP models. This variable has the second largest rank correlation in absolute value with accident probability (-0.245). After aircraft generation and excluding all variables related to unsuitability; experience for captains, first officers and maintenance personnel seem to be most influential in the accident probability. Fatigue and weather complete the top 15 most correlated variables with accident probability.

Table 3.7 presents the expectation, 5^{th} and 95^{th} percentiles of the accident rate distribution for the baseline case and two conditional distributions. The ratio of the 95^{th} to the 5^{th} percentiles in the baseline case is 104.5. The third row of table 3.7 presents data for the accident rate given the oldest type of aircraft. The expectation of the conditional distribution of accident is 16.2 times larger than the base line case. The ratio of the 95^{th} to the 5^{th} percentiles in the *conditional distribution A* is 484.3.

According to table 3.6 the next variables most highly correlated with accident probability other than unsuitability are first officers and captains experience. The model is further conditionalized on low values of experience for crew (5^{th} percentile of each experience distribution). The results are presented in the last row of table 3.7. In the *conditional distribution B* the ratio of the 95^{th} to the 5^{th} percentiles is 171.9, however the accident probability is 92 times larger than the base line case.

Rank	Variable	Model & Node #	Flight phase	Rank correlation (with accident)
1	Crew Unsuitability	FCP 10	TO	0.301
2	Crew Unsuitability	FCP 10	AL	0.300
3	Crew Unsuitability	FCP 10	ER	0.295
4	Aircraft generation	FCP\MNTP 11\5	TO\ER\AL	-0.245
5	Unsuitability (Captain)	FCP 6	AL	0.221
6	Unsuitability (FO)	FCP 7	AL	0.219
7	Unsuitability (Captain)	FCP 6	TO	0.218
8	Unsuitability (Captain)	FCP 6	ER	0.218
9	Experience (Captain)	FCP 5	TO\ER\AL	-0.217
10	Experience (FO)	FCP 1	TO\ER\AL	-0.216
11	Unsuitability (FO)	FCP 7	TO	0.216
12	Unsuitability (FO)	FCP 7	ER	0.215
13	Experience (Maintenance)	MNTP 3	TO\ER\AL	-0.208
14	Fatigue (Maintenance)	MNTP 2	TO\ER\AL	0.195
15	Weather	FCP 8	AL	0.177
16	Language difference	FCP 11	TO\ER\AL	0.160
17	Weather	FCP 8	TO	0.127
18	Weather	FCP 8	ER	0.121
19	Total transmission time	FCP\ATCP 13\6	AL	0.115
20	Workload (Maintenance)	MNTP 6	TO\ER\AL	0.108
21	Workload (Flight crew)	FCP 9	AL	0.063
22	Shift overlap time	MNTP 4	TO\ER\AL	-0.056
23	Workload (Flight crew)	FCP 9	ER	0.052
24	Workload (Flight crew)	FCP 9	TO	0.050
25	Fatigue (Flight crew)	FCP 9	AL	0.045
26	Fatigue (Flight crew)	FCP 9	ER	0.037
27	Fatigue (Flight crew)	FCP 9	TO	0.034
28	Traffic	ATCP 1	AL	-0.011
29	Training (FO)	FCP 2	TO\ER\AL	0.009
30	Interface	ATCP 2	ER	0.009
31	Working Condition	MNTP 1	TO\ER\AL	0.009
32	Traffic	ATCP 1	TO	-0.008
33	Experience (Controller)	ATCP 4	ER	0.008
34	Coordination	ATCP 3	AL	0.007
35	Visibility procedure	ATCP 5	AL	-0.006
36	Interface	ATCP 2	AL	-0.005
37	Coordination	ATCP 3	TO	0.005
38	Visibility procedure	ATCP 5	ER	0.003
39	Visibility procedure	ATCP 5	TO	0.003
40	Experience (Controller)	ATCP 4	TO	0.002
41	Interface	ATCP 2	TO	-0.002
42	Traffic	ATCP 1	ER	-0.001
43	Experience (Controller)	ATCP 4	AL	-8×10^{-4}
44	Coordination	ATCP 3	ER	-3×10^{-4}
45	Training (Captain)	FCP 4	TO\ER\AL	-2×10^{-5}

Table 3.6: Variables from the human reliability models from section 3.3.2 with highest absolute rank correlation with accident probability (32,500 samples per variable).

Model	Accident rate		
	Mean	5 th percentile	95 th percentile
Base line	3.18×10^{-6}	8.58×10^{-8}	8.97×10^{-6}
Conditional A ^a	5.14×10^{-5}	4.53×10^{-7}	2.19×10^{-4}
Conditional B ^b	2.92×10^{-4}	6.91×10^{-6}	1.20×10^{-3}

Table 3.7: *Expectation, 5th and 95th percentiles for base line accident probability and conditional distributions*

^aGiven aircraft generation = 1 (oldest kind of aircraft)

^bGiven aircraft generation = 1 (oldest kind of aircraft), captain experience = 9,467 hr (5th percentile) and first officer experience = 7,844 hr (5th percentile).

The three distributions from table 3.7 may be observed in Figure 3.17. From the picture it is immediately evident the negative effects that older aircrafts and very low experienced crews have on aviation safety.

Similar analysis to the one presented in this section is possible with the CATS model and UNINET. A single run of the CATS model in UNINET takes about 2.5 minutes on a PC with a CoreTM 2 Duo processor at 3 GHz and 3.25 GB of RAM memory. A BBN model with about 1.5 thousand nodes and a dependence structure imposed by more than 4,000 arcs may be an important source of information for the aviation system. Once the model is complete, it is the task of the users and analysts to place a major focus on the answers that should be retrieved from the model. Next chapter will focus on the methods employed to retrieve the rank and conditional rank correlations required by the models presented in section 3.3.2 from experts.

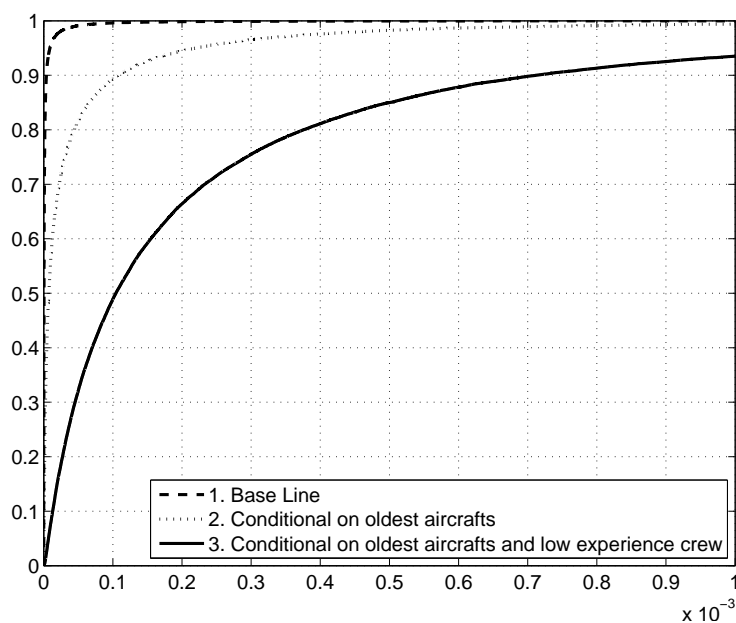


Figure 3.17: Fatal and non-fatal accident distribution from the CATS model. 1. Base line; 2. Given aircraft generation = 1; 3. Given aircraft generation = 1, captain experience = 9,467 hr. and first officer experience = 7,844 hr.

CHAPTER 4

Elicitation and Combination of Dependence¹.

4.1 Introduction

The elicitation of expert judgments for use in scientific research and decision making is a common practice nowadays. Most of the times these judgments gather information about univariate distributions of continuous uncertain quantities. The use of dependence measures between given uncertain quantities is becoming more and more important in risk and uncertainty analysis. This is true at least for the aviation industry where a large model for quantifying and analyzing risks is under development in the Netherlands (see section 1.4 and chapter 3). The use of structured expert judgment for eliciting and combining dependence measure is far less developed than the use of expert judgment for the elicitation of marginal distributions.

This chapter discusses the elicitation and combination of expert judgments in the form of rank and conditional rank correlations. These methods may be used in graphical models such as vines (chapter 2) or BBNs (chapter 3) and have been used with groups of experts for the quantification of models for the aviation industry (chapter 5) and dams safety 6. Choices of copulae for the elicitation are compared.

Whenever data are available about the joint distribution, it may be used for the quantification of a BBN. Often times, data is not available for the complete quantification of a NPCDBBN. In this case structured expert judgment provides another kind of data for model quantification. Methods for eliciting rank and conditional rank correlations have been presented before, for example in Cooke and Goossens [1999], Clemen and et al. [2000], Clemen and et al. [1999], Kraan [2002] and Morales et al. [2008].

¹This chapter is based on Morales et al. [2008] and Morales-Nápoles et al. [2009b]

The purpose of this chapter is to present examples where groups of experts have been gathered for the quantification of (un)conditional rank correlations for use in risk analysis. These measures are input for continuous-discrete non-parametric BBNs. In a project commissioned by the Dutch Ministry of Transport, Public Works and Water Management for aviation safety, a model for “*Missed Approach*” was developed and quantified with the *probabilistic method* described here. This application model will be presented in chapter 5. The probabilistic method is also used in the quantification of the FCP model. A *direct method* for the elicitation of rank correlations is also presented in the quantification of the ATPC model. The FCP and ATPC models will be further discussed in chapter 5. The combination of expert dependence measures in the form of conditional rank correlations will also be discussed and follows the ideas presented previously in Cooke and Goossens [1999] and Kraan [2002]. It is a goal of this chapter to serve as a guideline for the quantification of models similar to CATS.

4.2 Structured Expert Judgment

As stated previously in the NPCDBBN approach, nodes represent univariate random variables with invertible distribution function. Arcs represent parent-child (un)conditional rank correlations. The choice of (un)conditional rank correlations to represent influence responds to the fact that the conditional rank correlations are algebraically independent and every number in $(-1, 1)$ may be attached to the arcs of a BBN. The univariate marginal distributions represented by nodes in the BBN, together with the conditional independence statements embedded in the graph, and a copula realizing the correlations, uniquely determine the joint distribution Hanea et al. [2006]. It is important that for the chosen copula, zero correlation corresponds to the independent copula as this assures that the conditional probability statements implied by the graph are satisfied². The *normal copula* is the preferred choice since in addition to the zero independence property, computationally expensive numerical evaluations of multiple integrals are avoided. Additionally, the relationship between partial correlations and conditional rank correlations for the normal copula might be of advantage during an elicitation with experts.

The quantification of a full BBN requires marginal distributions for each node and dependence information in the form of (un)conditional rank correlations. Whenever data is not available for any of these inputs expert judgment is yet another kind of data available. The use of expert judgments in science is not new. *Structured* expert judgment has been proposed as a methodology to use expert opinion in a transparent way with the purpose of treating expert judgments as scientific data. Structured expert judgment has been widely used for the quantification of uncertain quantities in the form of subjective probability distributions [Cooke and Goossens, 2008]. The use of structured expert judgment for multivariate elicitation is less explored in the literature [O’Hagan, 2005] though some progress has been made over the past years. The use of structured expert

²This property is often referred as *the zero independence property*

judgment will be dealt with in more detail next.

4.2.1 The Classical Model for Structured Expert Judgment

The name *classical model* derives from its resemblance to classical statistical hypothesis testing. In 17 years, about 67,000 experts' subjective probability distributions have been elicited from 521 domain experts with the classical model ([Cooke and Goossens, 2008]). Fields of application include nuclear applications, chemical and gas industry, water management, aviation, health, banking, vulcanology and others.

The classical model for structured expert judgment [Cooke, 1991] is a performance based linear pooling (weighted average) model. In addition to variables of interest, experts are queried about *seed* or *calibration* variables. The latter are variables whose value is known to the analyst but not to the expert at the moment of the elicitation. Experts' performance as uncertainty assessors is measured by the *calibration* and *information* scores from seed variables. These are used to derive the weights entered in the linear pooling (Equation 4.1).

Roughly speaking, the calibration score is the probability that the divergence between the expert's assessments and the observed values on seed variables might have arisen by chance. A high score near 1 but higher than a significance level α (for instance 0.05) means that the expert's assessments are statistically supported by the set of seed variables. The second performance measure is the information score. Loosely, the information score measures the degree to which a distribution is concentrated relative to a background measure. The uniform and log uniform are most common choices for the background measures. The overall information score is the mean of information scores for each variable. The weights in the classical model are proportional to the product of statistical likelihood and information and satisfy a strictly proper scoring rule constraint [Cooke, 1991].

The linear pooling of experts' assessments is called a *Decision Maker* (DM). If $f_{e,i}$ is expert e 's density for item i then the decision maker is:

$$DM_{\alpha,i} = \sum_e w_{e,\alpha} f_{e,i} \quad (4.1)$$

The weights ($w_{e,\alpha} \geq 0$ and $\sum_e w_{e,\alpha} = 1$) are determined for each expert according to calibration and information. The value of α is chosen such that the product of the calibration and information scores of the decision maker is maximized. Any expert whose calibration score is less than α would be un-weighted in equation (4.1). Three types of DM are contained in the classical model. The equal weights decision maker (EWDM), the global weights decision maker (GWDM) and the item weights decision maker (IWDM).

The EWGD assigns equal weight to each expert and hence reduces equation (4.1) to the arithmetic mean of experts' opinions. This decision maker is not in the class of performance based DMs and hence does not implement the procedure described above to un-weight experts. The GWDM and IWDM are performance based decision makers. The GWDM determines weights per expert by each expert's calibration score and the overall information score. The IWDM determines

the weights per expert and per variable using the information score on each variable rather than the averaged information score across variables.

The classical model has been mostly used to determine information about univariate distributions. The elicitation of multivariate distributions has been a topic for more recent research. This subject will be discussed next.

4.2.2 Dependence Elicitation

The literature available to guide researchers in the elicitation of a joint distribution is much less than that available for the elicitation of univariate distributions [O'Hagan, 2005]. Previous studies have shown that eliciting dependence measures for the construction of multivariate distributions though not an easy task is still possible (Cooke and Goossens [1999], Clemen and et al. [1999], Clemen and et al. [2000]). In Kraan [2002] elicitation techniques are summarized and exemplified using the probabilistic approach. Combination schemes for experts' opinions are also proposed in Cooke and Goossens [1999] and Kraan [2002] for bivariate distributions. An extension of previous methodologies for the elicitation of conditional rank correlations is discussed in Morales et al. [2008] as input for continuous-discrete non-parametric BBNs and vines, however no discussion regarding the combination of experts' individual assessments for distributions of order higher than two was performed. In this section guidelines for the elicitation and combination of experts' estimates of rank and conditional rank correlations as input for BBNs will be discussed.

Two types of methods for the elicitation of dependence measures will be discussed. Examples of the two methods will be discussed with the BBN in Figure 4.1. This BBN corresponds to Figure 3.2 after removing node 7. The six marginal distributions (one for each node) may be computed from data from separate sources or with the classical model for expert judgment outlined in previous subsection. Variables $\{X_1 \dots X_5\}$ are independent of each other. Four conditional rank correlations and one unconditional rank correlation are required. The rank and conditional rank correlations are associated with edges according to the protocol discussed in section 3.2 [Hanea et al., 2006].

4.2.2.1 Probabilistic Approaches

Experts give probability statements such as a joint probability, a conditional probability or a probability of concordance. By making assumptions about the joint distribution, the assessments can later be translated to rank correlations. Denote $r_{6,1}^{e_i}$ the rank correlation between X_6 and X_1 for expert $i = 1, \dots, N$. Similarly the conditional rank correlation between X_6 and X_2 given X_1 will be denoted as $r_{6,2|1}^{e_i}$ for expert e_i . All other (un)conditional rank correlations in the BBN will be denoted similarly. The median value of variable X_j for expert e_i is denoted as $x_{j,q_{50}}^{e_i}$. Similarly the k^{th} percentile of variable X_j is denoted as $x_{j,q_k}^{e_i}$. The cumulative distribution function for variable X_j from expert e_i will be denoted as $F_{X_j}^{e_i}$. The probabilistic approach recommends eliciting the 5 probabilities $P_1^{e_i}, \dots, P_5^{e_i}$ to each expert to quantify the BBN in Figure 4.1 for e_i :

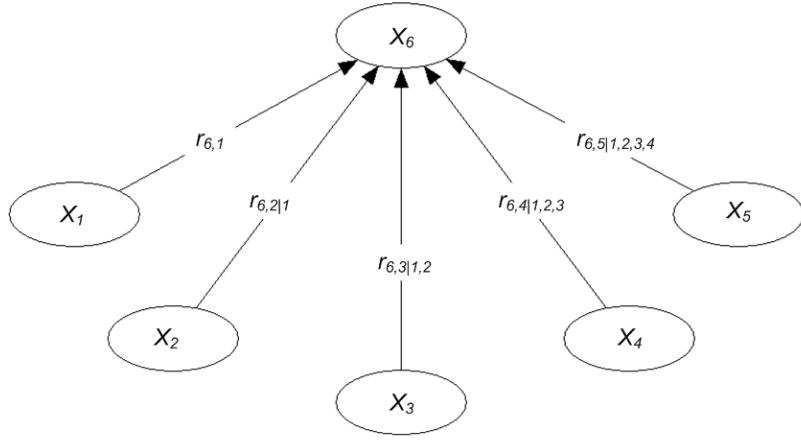


Figure 4.1: A simple example of BBN on 4 Variables.

1. $P_1^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i})$
 $= P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5)$
 2. $P_2^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}, X_2 \geq x_{2,q_{50}}^{e_i})$
 $= P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5, F_{X_2}^{e_i}(X_2) \geq 0.5)$
 3. $P_3^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}, X_2 \geq x_{2,q_{50}}^{e_i}, X_3 \geq x_{3,q_{50}}^{e_i})$
 $= P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5, \dots, F_{X_3}^{e_i}(X_3) \geq 0.5)$
 4. $P_4^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}, X_2 \geq x_{2,q_{50}}^{e_i}, X_3 \geq x_{3,q_{50}}^{e_i}, X_4 \geq x_{4,q_{50}}^{e_i})$
 $= P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5, \dots, F_{X_4}^{e_i}(X_4) \geq 0.5)$
 5. $P_5^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}, X_2 \geq x_{2,q_{50}}^{e_i}, X_3 \geq x_{3,q_{50}}^{e_i}, X_4 \geq x_{4,q_{50}}^{e_i}, X_5 \geq x_{5,q_{50}}^{e_i})$
 $= P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5, \dots, F_{X_5}^{e_i}(X_5) \geq 0.5)$
- (4.2)

The first question of the elicitation is read as: *Suppose that variable X_1 was observed above its q_k^{th} quantile. What is the probability that also X_6 will be observed above its q_k^{th} quantile?* Notice that the recommended choice for the percentile used in the probabilities stated in relation 4.2 is the median; however any other percentile $x_{j,q_k}^{e_i}$ may be used. In particular other percentiles are necessary for discrete variables. Notice also that as stated before, conditional on the BBN to be quantified, other probabilistic statements could be elicited in relation 4.2 according to the analysts preference. For example shorter conditioning sets might be considered: $P_1^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}), \dots, P_5^{e_i} = P(X_6 \geq x_{6,q_{50}}^{e_i} | X_5 \geq x_{5,q_{50}}^{e_i})$. Another option would be to elicit joint distributions, probabilities of concordance or discordance or other probabilistic statements about the

joint distribution, instead of conditional probabilities of exceedence.

Once estimates as in relation 4.2 are available to the analyst, the corresponding (un)conditional rank correlations may be computed for each expert (relation 4.3).

$$\begin{aligned}
 P_1^{e_i} &\rightarrow r_{6,1}^{e_i} \\
 P_2^{e_i} &\rightarrow r_{6,2|1}^{e_i} \\
 P_3^{e_i} &\rightarrow r_{6,3|1,2}^{e_i} \\
 P_4^{e_i} &\rightarrow r_{6,4|1,2,3}^{e_i} \\
 P_5^{e_i} &\rightarrow r_{6,5|1,2,3,4}^{e_i}
 \end{aligned} \tag{4.3}$$

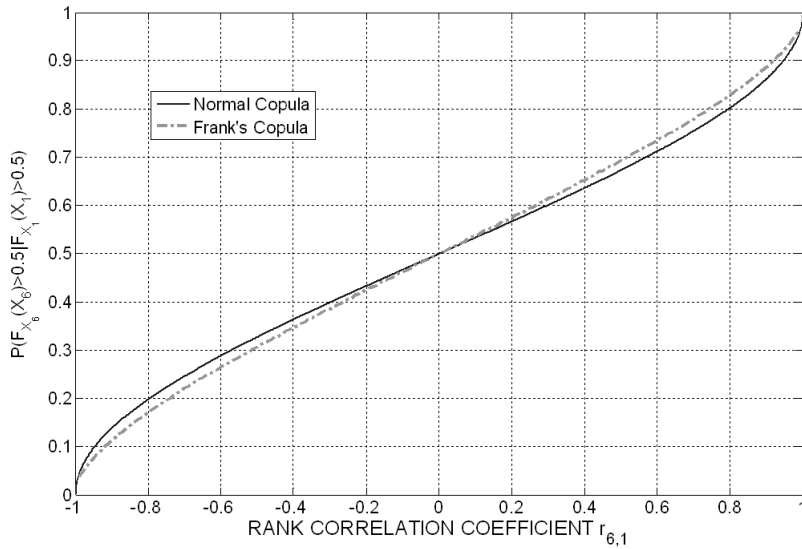


Figure 4.2: $P(X_6 \geq x_{6,q_{50}} | X_1 \geq x_{1,q_{50}})$ for the normal & Frank's copulae

The rank correlation $r_{6,1}^{e_i}$ may be obtained for each expert from their answer to $P_1^{e_i}$ in relation 4.2. The relation for each possible value of $P_1^{e_i}$ and $r_{6,1}^{e_i}$ is shown in Figure 4.2 for Frank's and the normal copulae.

For the normal copula to calculate the exceedance probability one can integrate numerically the bivariate normal density $\phi(\tilde{x}_6, \tilde{x}_1, \rho_{6,1}^{e_i})$ over the region corresponding to the quantile's exceedance region $[\Phi^{-1}(q_{50}), \infty) \times [\Phi^{-1}(q_{50}), \infty)$, where Φ^{-1} is the inverse standard normal cumulative distribution function. The analyst may use formula (4.4) where \tilde{x}_{i,q_k} is the standard normal variate transform of x_{i,q_k} . In other words \tilde{x}_{i,q_k} will be the k^{th} quantile of the corresponding standard normal distribution. The analyst then finds the ρ which satisfies the expert's conditional probability assessment and transforms this to the corresponding rank correlation using the inverse function of equation (1.2). A similar procedure can be followed using Frank's copula (equation (1.3)).

$$\frac{1}{1 - p_k} \int_{\Phi^{-1}(q_k)}^{\infty} \int_{\Phi^{-1}(q_k)}^{\infty} \phi(\tilde{x}_6, \tilde{x}_1, \rho_{6,1}^{e_i}) d\tilde{x}_6 d\tilde{x}_1 \tag{4.4}$$

For both copulae because of the zero independence property, zero correlation entails that for any k , $P_1^{e_i} = 1 - q_k$. A conditional probability value in the interval $[0, 1 - q_k)$ corresponds to negative correlation and positive correlation is attained when $P_1^{e_i} > 1 - q_k$. Choosing a value for q_k different than 0.5 makes the resulting rank correlation more dependent on the chosen of copula.

With the answer to $P_i^{e_i}$ a relationship between $P_2^{e_i}$ and $r_{6,2|1}^{e_i}$ may be computed. According to relation 4.2 experts would be queried: *Suppose that not only variable X_1 but also X_2 were observed above their medians. What is now your probability that also X_6 will be observed above its median value?*

The probability that each expert i can provide in this situation will depend on the estimate given for $P_1^{e_i}$. The reader may see this by observing that if each expert regards variables X_2 and X_6 as independent given X_1 , then their answer to $P_2^{e_i}$ is identical to the answer to question $P_1^{e_i}$ for each i . If the expert regards variables X_1 and X_6 as completely positively (negatively) correlated then he/she would have answered $P_1^{e_i} = 1$ ($P_1^{e_i} = 0$) and question 2 would not have been necessary at all, as X_6 would be completely explained by X_1 . Any answer for $P_1^{e_i}$ different than 0, 0.5 or 1 means that the expert believes that X_2 explains at least in part X_6 and hence X_2 can only explain part of the dependence that was not explained already by X_1 .

Suppose expert's 1 answer for question 1 in relation 4.2 was $P_1^{e_1} = 0.33$. In this case according to Figure 4.2 $r_{6,1}^{e_1}$ would be equal to -0.49 for the normal copula and -0.44 for the Frank's copula³. This situation is shown in Figure 4.3. For the normal copula $P_2^{e_1} \in (0, 0.65)$ and for the Frank's copula $P_2^{e_1} \in (0.16, 0.56)$. Observe that $P_2^{e_1}$ as a function of $r_{6,2|1}^{e_1}$ is highly dependent on the choice of the copula.

In the case of the normal copula, to determine the possible values for $P_2^{e_1}$ and its relationship with the conditional correlation $r_{6,2|1}^{e_1}$ we consider a normal D-vine on variables X_6 , X_1 and X_2 . As mentioned earlier, the rank correlation $r_{6,1}^{e_1}$ has been already calculated using expert's assessment in question 1. In the particular case of the BBN in Figure 3.2, variables X_1 and X_2 are independent, hence $r_{1,2}^{e_i}$ is equal to zero. Since all rank correlations specified on the BBN are algebraically independent, $r_{6,2|1}^{e_i}$ can take any value in $(-1, 1)$. The correlation matrix of the joint normal distribution corresponding to this normal vine can be found as in example 3.2.1 in chapter 3 and should have the form of equation (4.5).

$$\Sigma_{6,1,2}^{e_i} = \begin{pmatrix} \rho_{2,2}^{e_i} & \rho_{1,2}^{e_i} & \rho_{2,6}^{e_i} \\ \rho_{2,1}^{e_i} & \rho_{1,1}^{e_i} & \rho_{1,6}^{e_i} \\ \rho_{6,2}^{e_i} & \rho_{6,1}^{e_i} & \rho_{6,6}^{e_i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \rho_{2,6}^{e_i} \\ 0 & 1 & \rho_{1,6}^{e_i} \\ \rho_{2,6}^{e_i} & \rho_{1,6}^{e_i} & 1 \end{pmatrix} \quad (4.5)$$

We denote the density function of the normal distribution with the correlation matrix $\Sigma_{6,1,2}^{e_i}$ calculated from the normal vine specification as $\phi(\tilde{x}_6, \tilde{x}_1, \tilde{x}_2, \rho_{6,1}^{e_i}, \rho_{6,2|1}^{e_i})$. Hence, given the value for $r_{1,6}^{e_1}$ a relationship between $P_2^{e_1}$ and $r_{6,2|1}^{e_1}$ can be determined by transforming to $\rho_{6,2|1}^{e_1}$ using formula 1.2 and computing the triple integral (4.6):

³A similar example for the minimum information copula vs. Frank's copula is presented in Morales et al. [2008]

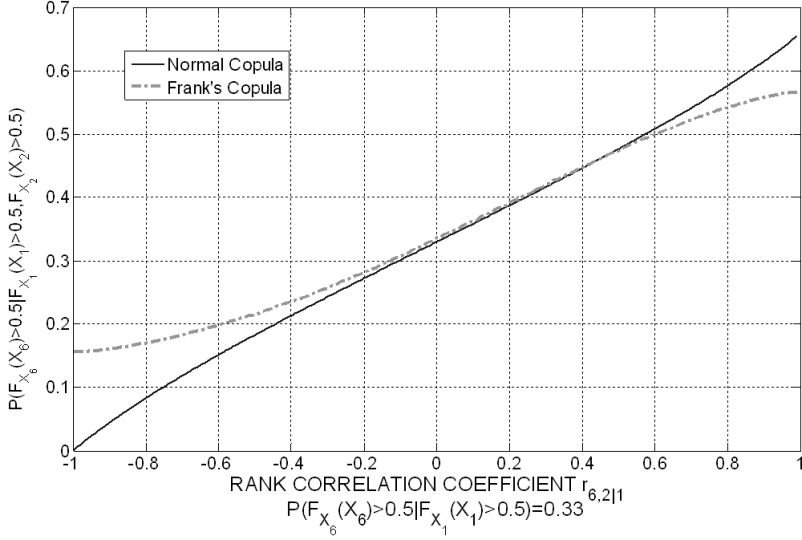


Figure 4.3: $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}, X_2 \geq x_{2,q50})$ for the normal & Frank's copulae $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}) = 0.33$

$$\frac{1}{0.5 \cdot 0.5} \int_0^\infty \int_0^\infty \int_0^\infty \phi(\tilde{x}_6, \tilde{x}_1, \tilde{x}_2, \rho_{6,1}^{e_1}, \rho_{6,2|1}^{e_1}) d\tilde{x}_6 d\tilde{x}_1 d\tilde{x}_2 \quad (4.6)$$

For other copulae (such as the Frank's copula used in this example) the relationship may be computed through simulation. Simulations were done in `matlab` R2007b following the copula-vine method [Kurowicka and Cooke, 2006, ch.6]. In both cases, if $P_2^{e_1} = P_1^{e_1}$ then the expert regards variables X_2 and X_6 as independent given X_1 and in this case $r_{6,2|1}^{e_1} = 0$.

An expert's answer for $P_2^{e_1} > P_1^{e_1}$ would correspond to $r_{6,2|1}^{e_1} > 0$ and accordingly if $P_2^{e_1} < P_1^{e_1}$ then $r_{6,2|1}^{e_1} < 0$. Notice that the fact that $r_{6,2|1}^{e_1} < 0$ (> 0) does not imply that $r_{6,2}^{e_1} < 0$ (> 0). The sign of $r_{6,2}^{e_1}$ would depend in general on the graphical structure of the BBN and the experts' previous answers. For example, suppose that $r_{1,2}^{e_1} = -0.9$ and as before $P_1^{e_1} = 0.33$. In this case for $r_{6,2|1}^{e_1} \in (-1, 1)$, $r_{6,2}^{e_1} > 0$ for both the normal and Frank's copulae. The situation may be observed in Figure 4.4.

For simplicity we go back to our example where $r_{1,2}^{e_i} = 0$. Suppose further that expert one answered $P_2^{e_1} = 0.25$ then $r_{6,2|1}^{e_1}$ equals -0.28 for the normal copula and -0.32 for Frank's copula. Next $P_3^{e_1}$ as a function of $r_{6,3|1,2}^{e_1}$ may be computed based on the expert's previous answers and the structure of the BBN. The expert would be asked: *Suppose that not only variables X_2 and X_1 but also X_3 were observed above their medians. What is now your probability that also X_4 will be observed above its median value?*

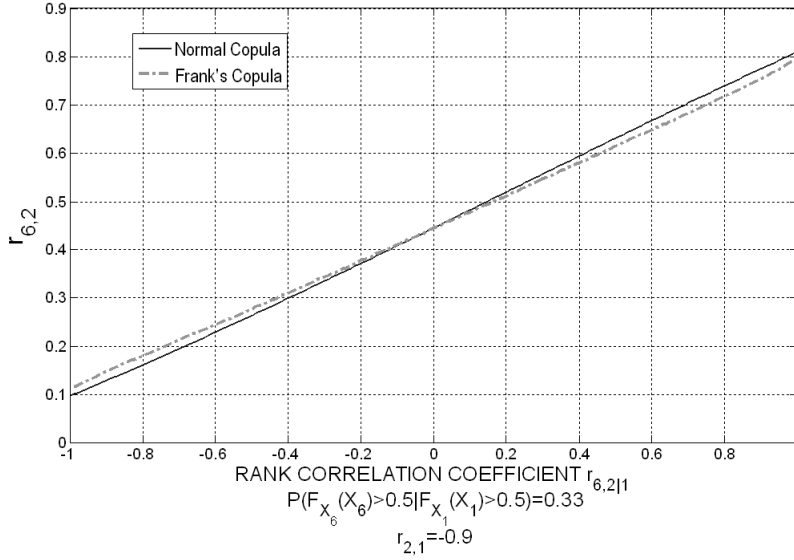


Figure 4.4: $r_{6,2}^{e_1}$ as a function of $r_{6,2|1}^{e_1}$ for the normal & Frank's copulae $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}) = 0.33$, $r_{1,2}^{e_1} = -0.9$

As before, the rank correlations $r_{1,6}^{e_1}$ and $r_{2,6}^{e_1}$ have been specified in questions 1 and 2 respectively. Again from Figure 3.2 it is observed that $r_{3,1}^{e_i}$ and $r_{2,3}^{e_i}$ are both zero. For the normal copula approach it may be observed that the correlation matrix of the joint normal distribution corresponding to the D-vine on X_1, X_2, X_3 and X_6 should look as in equation (4.7). The density of this four variate standard normal distribution will be denoted as $\phi(\tilde{x}_4, \tilde{x}_3, \tilde{x}_2, \tilde{x}_1, \rho_{1,6}^{e_1}, \rho_{2,6}^{e_1}, \rho_{6,3|1,2}^{e_1})$.

$$\Sigma_{6,3,2,1}^{e_i} = \begin{pmatrix} \rho_{3,3}^{e_i} & \rho_{2,3}^{e_i} & \rho_{1,3}^{e_i} & \rho_{3,6}^{e_i} \\ \rho_{2,3}^{e_i} & \rho_{2,2}^{e_i} & \rho_{2,1}^{e_i} & \rho_{2,6}^{e_i} \\ \rho_{1,3}^{e_i} & \rho_{2,1}^{e_i} & \rho_{1,1}^{e_i} & \rho_{1,6}^{e_i} \\ \rho_{3,6}^{e_i} & \rho_{2,6}^{e_i} & \rho_{1,6}^{e_i} & \rho_{6,6}^{e_i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \rho_{3,6}^{e_i} \\ 0 & 1 & 0 & \rho_{2,6}^{e_i} \\ 0 & 0 & 1 & \rho_{1,6}^{e_i} \\ \rho_{3,6}^{e_i} & \rho_{2,6}^{e_i} & \rho_{1,6}^{e_i} & 1 \end{pmatrix} \quad (4.7)$$

The relationship between $P_3^{e_i}$ and $r_{6,3|1,2}^{e_i}$ will be determined by transforming to the corresponding $\rho_{6,3|1,2}^{e_i}$ with formula 1.2 and computing the four dimensional integral 4.8. In the case of Frank's copula it is determined by simulation from the vine-copula method [Kurowicka and Cooke, 2006].

$$\frac{1}{0.5 \cdot 0.5 \cdot 0.5} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \phi(\tilde{x}_6, \tilde{x}_3, \tilde{x}_2, \tilde{x}_1, \rho_{6,1}^{e_i}, \rho_{6,2|1}^{e_i}, \rho_{6,3|1,2}^{e_i}) d\tilde{x}_4 d\tilde{x}_3 d\tilde{x}_2 d\tilde{x}_1 \quad (4.8)$$

This situation for expert e_1 is pictured in Figure 4.5. For both copulae if $P_3^{e_1} = P_2^{e_1}$ then the expert regards variables X_3 and X_6 as independent given

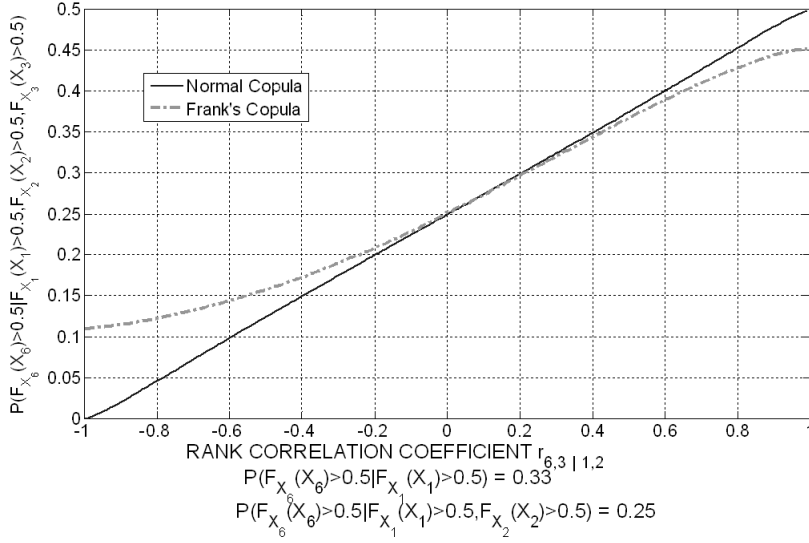


Figure 4.5: $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}, X_2 \geq x_{2,q50}, X_3 \geq x_{3,q50})$ for the normal & Frank's copulae $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}) = 0.33$, $P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}, X_2 \geq x_{2,q50}) = 0.25$

X_1 and X_2 and in this case $r_{6,3|2,1}^{e_1} = 0$. An expert's answer for $P_3^{e_1} > P_2^{e_1}$ would correspond to $r_{6,3|1,2}^{e_1} > 0$ and accordingly if $P_3^{e_1} < P_2^{e_1}$ then $r_{6,3|1,2}^{e_1} < 0$. Again the sign of $r_{6,3}^{e_1}$ would depend on the structure of the BBN and the expert's previous answers. In this case as before, the relationship between $P_3^{e_i}$ and $r_{6,3|1,2}^{e_i}$ is dependent on the choice of the copula.

Other relationships in 4.3 may be computed following the ideas discussed thus far. Extensions to other BBNs or similar graphical models follow straight away from this approach. In a real elicitation the bounds for each exceedence probability (or any other probabilistic statement chosen for the elicitation) must be computed in real time. If the expert's estimates are not consistent with the allowable bounds for each $P_j^{e_i}$ for $j = 2, \dots, Pa(n)$ for a given node n , then the estimate must be discussed with the expert and revised if necessary.

4.2.2.2 Direct Approach

Another option is to let experts directly assess a rank correlation. In particular, for each child node (X_6 in the example), we could let experts rank the parent nodes (X_1, \dots, X_5 in this case) according to rank correlation with X_6 (in absolute value). This ranking will in general be different for each expert. Experts could then be queried the following five numbers:

$$\begin{aligned}
1. \quad P_1^{e_i} &= P(X_6 \geq x_{6,q_{50}}^{e_i} | X_1 \geq x_{1,q_{50}}^{e_i}) \\
2. \quad R_2^{e_i} &= \frac{r_{6,2}^{e_i}}{r_{6,1}^{e_i}} \\
3. \quad R_3^{e_i} &= \frac{r_{6,3}^{e_i}}{r_{6,1}^{e_i}} \\
4. \quad R_4^{e_i} &= \frac{r_{6,4}^{e_i}}{r_{6,1}^{e_i}} \\
5. \quad R_5^{e_i} &= \frac{r_{6,5}^{e_i}}{r_{6,1}^{e_i}}
\end{aligned} \tag{4.9}$$

The first rank correlation is still elicited through a probabilistic statement and may be computed as described before (Figure 4.2). $R_2^{e_i}$ in relation 4.9 denotes the ratio of the second rank correlation to the largest rank correlation (in absolute value) for expert e_i . Similar notation is applied for other ratios. As before, the recommended choice for the percentile used in $P_1^{e_i}$ in relation 4.9 is the median, however any other percentile $x_{j,q_k}^{e_i}$ may be used. As stated before, other probabilistic statements could be elicited for $P_1^{e_i}$ in relation 4.9 according to the analyst's preference.

$$\begin{aligned}
P_1^{e_i} &\rightarrow r_{6,1}^{e_i} \\
R_2^{e_i} &\rightarrow r_{6,2|1}^{e_i} \\
R_2^{e_i} &\rightarrow r_{6,3|1,2}^{e_i} \\
R_2^{e_i} &\rightarrow r_{6,4|1,2,3}^{e_i} \\
R_2^{e_i} &\rightarrow r_{6,5|1,2,3,4}^{e_i}
\end{aligned} \tag{4.10}$$

Once the expert has given an estimate for $P_1^{e_i}$ the relationship between $R_2^{e_i}$ and $r_{6,2|1}^{e_i}$ may be computed. The computation of the required conditional rank correlations in relation 4.10 follows the same arguments as in section 4.2.2.1. For the normal copula the fact that conditional correlation is equal to partial correlation, the recursive formula for partial correlation and the known relationship between rank correlation and product moment correlation is sufficient to determine $R_2^{e_i}$ as a function of $r_{6,2|1}^{e_i}$ given the graph structure and the experts' assessment for $P_1^{e_i}$. For Frank's copula, as before this relation may be obtained by simulation [Kurowicka and Cooke, 2006].

In the case described thus far, the sign of $R_2^{e_1}$ depends on the experts' answer to $P_1^{e_1}$. In our example $r_{6,1}^{e_1}$ is negative. If the expert believes there is a negative correlation between X_2 and X_6 then $R_2^{e_1}$ must be positive and its value depends on the expert's belief of the distance between $r_{6,1}^{e_1}$ and $r_{6,2}^{e_1}$. Obviously if expert 1 believes that $r_{6,1}^{e_1} > r_{6,2}^{e_1}$ ($r_{6,1}^{e_1} < r_{6,2}^{e_1}$) then $R_2^{e_1} < 1$ ($R_2^{e_1} > 1$).

Assume that for a given expert $r_{6,1}^{e_1} = -0.49$. According to Figure 4.2 this would correspond to a value for $P_1^{e_1}$ equal to 0.33 for the normal copula and 0.31 for Frank's copula. The relationship between $R_2^{e_1}$ and $r_{6,2|1}^{e_1}$ is shown in Figure 4.6. In our example $R_2^{e_1} \in (-1.71, 1.71)$. In general the interval containing $R_j^{e_i}$ does not need to be symmetric about zero. This would depend on the graph structure and the expert's previous estimates (see Figure 4.4). Observe that in this case $R_2^{e_1}$ holds practically the same relationship with $r_{6,2|1}^{e_1}$ for both copulae. Suppose the expert's assessment is $R_2^{e_1} = 1.3$. This corresponds according to Figure 4.6

to a value of $r_{6,2}^{e_1} = -0.6358$ and $r_{6,2|1}^{e_1} = -0.749$ for both Frank's and normal copulae.

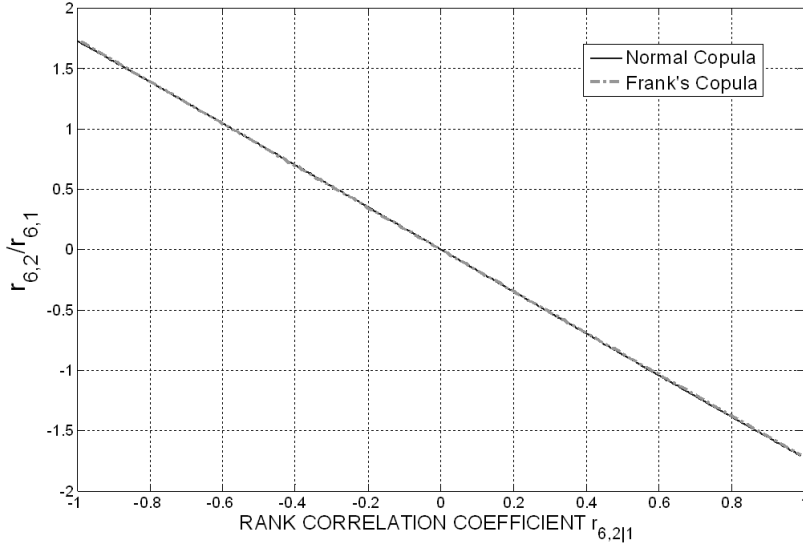


Figure 4.6: $\frac{r_{6,2}}{r_{6,1}}$ for the normal ($P(X_6 \geq x_{6,q_{50}} | X_1 \geq x_{1,q_{50}}) = 0.33$) & Frank's ($P(X_6 \geq x_{6,q_{50}} | X_1 \geq x_{1,q_{50}}) = 0.31$) copulae.

Once estimates for $P_1^{e_1}$ and $R_2^{e_1}$ are available, the relationship between $r_{6,3|1,2}^{e_1}$ and $R_3^{e_1}$ may be computed. For the example described here this relationship is shown Figure 4.7. In the example $R_3^{e_1} \in (-1.08, 1.10)$ for both the normal and Frank's copulae. Suppose that the expert would state $R_3^{e_1} = -0.8$; in this case $r_{6,3|1,2}^{e_1} \approx 0.73$ for both copulae and both unconditional correlations would be equal.

As before, in a real elicitation the bounds for each ratio of rank correlations must be computed in real time. If the experts' estimates are not consistent with the allowable bounds for each $R_j^{e_i}$ for $j = 2, \dots, Pa(n)$ for a given node n , then the estimate must be discussed with the expert and revised if necessary.

The normal copula is the preferred choice because it possess the zero independence property, it realizes a specified rank correlation without adding too much information to the independent copula [Lewandowski, 2005], its density covers the entire unit square and it offers important advantages for the computation of joint distributions specified by graphical structures such as BBNs. The use of other copulae is possible as long as they possess the zero independence property as exemplified by the Frank's copula. However the cost is a much higher computational effort⁴.

⁴For some BBNS additionally to obtaining estimates from experts through simulation, if the vine copula method [Kurowicka and Cooke, 2006] is used to update the joint distribution after computations have been done, numerical integrals might need to be calculated [Hanea et al.,

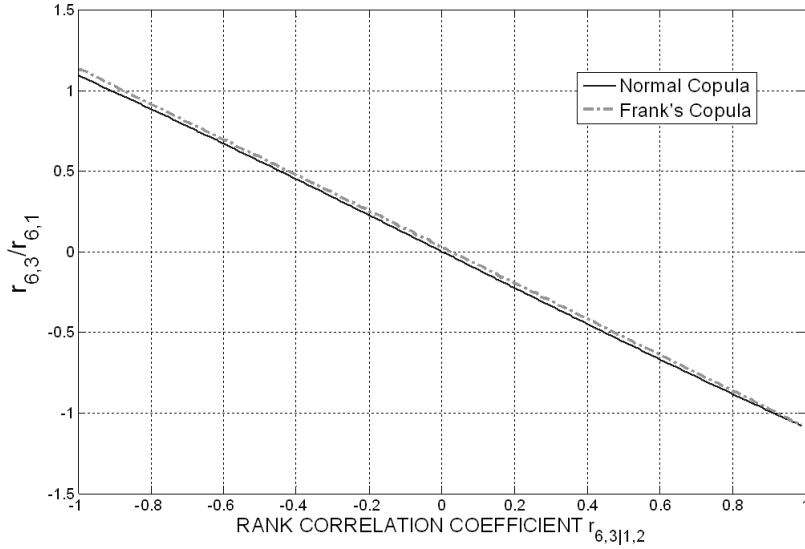


Figure 4.7: $\frac{r_{6,3}}{r_{6,1}}$ for the normal ($P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}) = 0.33$) & Frank's ($P(X_6 \geq x_{6,q50} | X_1 \geq x_{1,q50}) = 0.31$) copulae. $\frac{r_{6,2}}{r_{6,1}} = 1.3$

In chapters 5 and 6, results of the use of the techniques discussed in this section will be presented. The probabilistic method was used in eliciting the rank and conditional rank correlations required by the FCP model introduced in section 3.3.2.1. The direct method was used in the ATCP model introduced in section 3.3.2.2 and the model that will be discussed in chapter 6. One argument in favor of the elicitation of probabilistic statements is that their elicitation has proven to be feasible in previous studies [Kraan, 2002], [Morales et al., 2008] in real applications. Experts seem to be familiar with the elicitation of conditional probabilities. However, when the number of conditioning variables is large (as in relation 4.2) experts tend to object the elicitation of these exceedence probabilities. As mentioned previously this could be avoided by eliciting conditional probabilities with smaller number of conditioning variables.

The direct method combines the elicitation of one probabilistic statement with ratios of unconditional rank correlations. Based on our own experience we may say that one advantage of this method is that experts may express somewhat easier the 'relative strength' of each unconditional rank correlation (in the correlation matrix) as expressed by its absolute value. Once the correlation matrix is available for each expert any probabilistic statement may be computed (given the normal copula assumption) for each expert's estimates. The issue of combining their opinions arises once estimates from each expert are available. The combination of experts' dependence estimates is discussed next.

4.2.3 Combination of Experts' Dependence Estimates

The combination of expert's distributions for BBN's poses specific challenges. If every expert's distribution satisfies the conditional independence statements implied by the graph, then the linear pool individual densities in general will not. The reason is that conditional independence is not preserved under convex combinations of distributions. To combine the dependence information elicited from experts via conditional probabilities, it would be tempting to pool the conditional probabilities linearly to determine the conditional probability of the decision maker. This strategy would work well if the medians of all experts were the same which is not typically the case, for example if the marginal distributions come from expert judgment. In order to combine the experts' dependence information a different strategy must be taken. An example is presented with the BBN in Figure 3.2.

A strategy for combining experts' dependence estimates has been proposed in previous studies for bivariate distributions [Cooke and Goossens, 1999]. The procedure extended for multivariate distributions is presented in this section. First the individual expert judgments for marginal distributions are combined according to one of the linear pool weighting schemes [Cooke, 1991]. Later, the idea is to compute the probabilities that each expert "*would have stated*" if he/she had been asked probabilistic statements regarding the chosen quantile of the Decision Maker such that his/her estimates for the rank correlations remain unchanged (relation 4.11).

$$\begin{array}{rcl}
 r_{6,1}^{e_i} & \rightarrow & P_{1^*}^{e_i} \\
 r_{6,2|1}^{e_i} & \rightarrow & P_{2^*}^{e_i} \\
 r_{6,3|1,2}^{e_i} & \rightarrow & P_{3^*}^{e_i} \\
 r_{6,4|1,2,3}^{e_i} & \rightarrow & P_{4^*}^{e_i} \\
 r_{6,5|1,2,3,4}^{e_i} & \rightarrow & P_{5^*}^{e_i}
 \end{array} \tag{4.11}$$

First the joint distribution for each expert e_i is obtained by a procedure such as the one described in subsection 4.2.2. For each expert the joint distribution uses the estimated rank and conditional rank correlations obtained from relation 4.3 or relation 4.10 and the marginal distributions. Observe that in relation 4.10 the rank and conditional rank correlations computed for each expert could be indexed differently according to each expert.

Once the joint distribution is available for each expert, relation 4.11 says that some probabilistic statements are computed from the joint distribution of each expert. For example the exceedence probabilities in relation 4.12 below could be computed.

$$\begin{aligned}
P_{1^*}^{e_i} &= P(X_6 \geq x_{6,q50}^{DM} | X_1 \geq x_{1,q50}^{DM}) \\
&= P(F_{X_6}^{e_i}(X_6) \geq F_{X_6}^{e_i}(x_{6,q50}^{DM}) | F_{X_1}^{e_i}(X_1) \geq F_{X_1}^{e_i}(x_{1,q50}^{DM})) \\
P_{2^*}^{e_i} &= P(X_6 \geq x_{6,q50}^{DM} | X_1 \geq x_{1,q50}^{DM}, X_2 \geq x_{2,q50}^{DM}) \\
&= P(F_{X_6}^{e_i}(X_6) \geq F_{X_6}^{e_i}(x_{6,q50}^{DM}) | F_{X_1}^{e_i}(X_1) \geq F_{X_1}^{e_i}(x_{1,q50}^{DM}), F_{X_2}^{e_i}(X_2) \geq F_{X_2}^{e_i}(x_{2,q50}^{DM})) \\
P_{3^*}^{e_i} &= P(X_6 \geq x_{6,q50}^{DM} | X_1 \geq x_{1,q50}^{DM}, X_2 \geq x_{2,q50}^{DM}, X_3 \geq x_{3,q50}^{DM}) \\
&= P(F_{X_6}^{e_i}(X_6) \geq F_{X_6}^{e_i}(x_{6,q50}^{DM}) | F_{X_1}^{e_i}(X_1) \geq F_{X_1}^{e_i}(x_{1,q50}^{DM}), \dots, F_{X_3}^{e_i}(X_3) \geq F_{X_3}^{e_i}(x_{3,q50}^{DM})) \\
P_{4^*}^{e_i} &= P(X_6 \geq x_{6,q50}^{DM} | X_1 \geq x_{1,q50}^{DM}, X_2 \geq x_{2,q50}^{DM}, X_3 \geq x_{3,q50}^{DM}, X_4 \geq x_{4,q50}^{e_i}) \\
&= P(F_{X_6}^{e_i}(X_6) \geq F_{X_6}^{e_i}(x_{6,q50}^{DM}) | F_{X_1}^{e_i}(X_1) \geq F_{X_1}^{e_i}(x_{1,q50}^{DM}), \dots, F_{X_4}^{e_i}(X_4) \geq F_{X_4}^{e_i}(x_{4,q50}^{DM})) \\
P_{5^*}^{e_i} &= P(X_6 \geq x_{6,q50}^{e_i} | X_1 \geq x_{1,q50}^{DM}, X_2 \geq x_{2,q50}^{DM}, X_3 \geq x_{3,q50}^{DM}, X_4 \geq x_{4,q50}^{DM}, X_5 \geq x_{5,q50}^{DM}) \\
&= P(F_{X_6}^{e_i}(X_6) \geq F_{X_6}^{e_i}(x_{6,q50}^{DM}) | F_{X_1}^{e_i}(X_1) \geq F_{X_1}^{e_i}(x_{1,q50}^{DM}), \dots, F_{X_5}^{e_i}(X_5) \geq F_{X_5}^{e_i}(x_{5,q50}^{DM}))
\end{aligned} \tag{4.12}$$

As before the choice of the quantile of preference is the median, but any other quantile might be used as well. Other probabilistic statements might be computed in 4.12. For example conditional probabilities with smaller number of conditioning variables might be one option. In fact any probabilistic statement might be used as long as it is the same amongst experts. The reason is that these probabilistic statements will be combined later to form the DM's joint distribution⁵.

Consider the hypothetical example presented in Figure 4.8. Observe that the medians of experts 1, 2 and the *DM* disagree for variable X_1 . The medians are 50, 150 and 100 respectively for experts 1, 2 and the *DM*. It may also be observed that $F_{X_1}^{e_1}(100) = 0.66$ and $F_{X_1}^{e_2}(100) = 0.43$. For simplicity assume that all 3 experts agree on the median value of X_6 . This is the case if the marginal distribution for X_6 is obtained from data. Suppose that for the first probabilistic statement elicited experts answered as in 4.13 below.

$$\begin{aligned}
P_1^{e_1} &= P(X_6 \geq x_{6,q50} | X_1 \geq 50) = 0.75 \rightarrow r_{6,1}^{e_1} = 0.7 \\
P_1^{e_2} &= P(X_6 \geq x_{6,q50} | X_1 \geq 150) = 0.67 \rightarrow r_{6,1}^{e_2} = 0.5
\end{aligned} \tag{4.13}$$

The probabilities obtained from each expert in 4.13 cannot be combined directly. This is because, as stated previously, the median value for X_1 for each expert and the decision maker differ. In other words, the probabilities in 4.13 are taken over different events. According to 4.11 the analyst must compute a probabilistic statement for each expert taken over the same event before combining each expert's individual assessment. With the rank correlations of each expert, the analyst may compute an answer as in 4.14 below.

⁵If a method as in relation 4.10 is used, the assignment of rank and conditional rank correlations to the arcs of the BBN might not be equal across experts. However once the joint distribution is available for each expert, the same probabilistic statement may be computed for all.

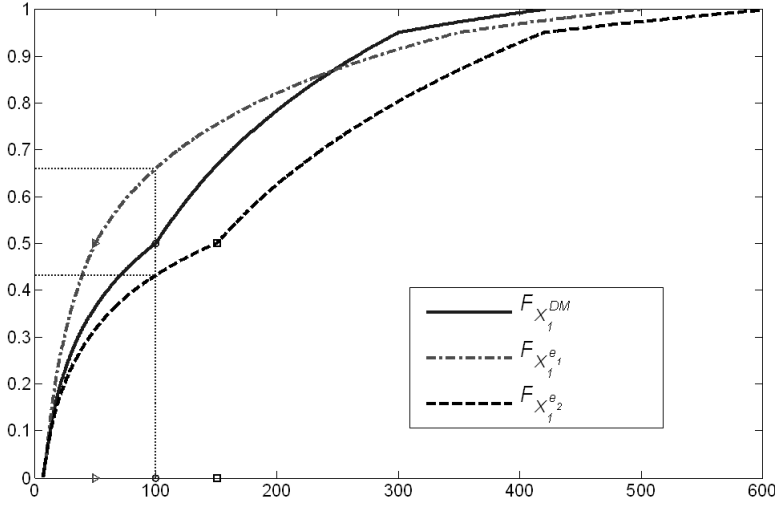


Figure 4.8: Difference in median between e_1 , e_2 and the DM for Variable X_1

$$\begin{aligned}
 r_{6,1}^{e_1} &= 0.7 \rightarrow P_{1*}^{e_1} = P(F_{X_6}^{e_1}(X_6) \geq 0.5 | F_{X_1}^{e_1}(X_1) \geq 0.66) = 0.84 \\
 r_{6,1}^{e_1} &= 0.4 \rightarrow P_{1*}^{e_2} = P(F_{X_6}^{e_1}(X_6) \geq 0.5 | F_{X_1}^{e_1}(X_1) \geq 0.43) = 0.65
 \end{aligned} \tag{4.14}$$

In Figure 4.9 the graphical representation of relation 4.14 is presented. Three probabilities are computed as a function of $r_{6,1}$. $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5)$ is represented by a solid line and it is the function from which the original estimates $r_{6,1}^{e_1}$ and $r_{6,1}^{e_2}$ in 4.13 are computed. $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_1}(X_1) \geq 0.66)$ and $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_2}(X_1) \geq 0.43)$ differ from the solid line because $F_{X_1}^{e_1}(100) = 0.66$ and $F_{X_1}^{e_2}(100) = 0.43$. The estimates in relation 4.14 are computed from the functions shown in Figure 4.9. Observe that $P_{1*}^{e_1}$ increases with respect to $P_{1*}^{e_1}$ and $P_{1*}^{e_2}$ decreases with respect to $P_{1*}^{e_2}$. Observe also that the rank correlation estimates remain equal.

When the probabilistic statements such as those suggested in relation 4.12 have been computed by the analyst, combining them is the next step. In analogy to equation (4.1) the probabilistic statement for the decision maker is computed as in 4.15.

$$P_j^{DM} = \sum_i w_{e_i} P_{j*}^{e_i} \tag{4.15}$$

The weights for each expert (w_{e_i}) may be computed from the *classical model* as described in subsection 4.2.1. Finally, as in subsection 4.2.2 the probabilistic statements obtained for the DM may be translated to the (conditional) rank correlations required by the model. In our example:

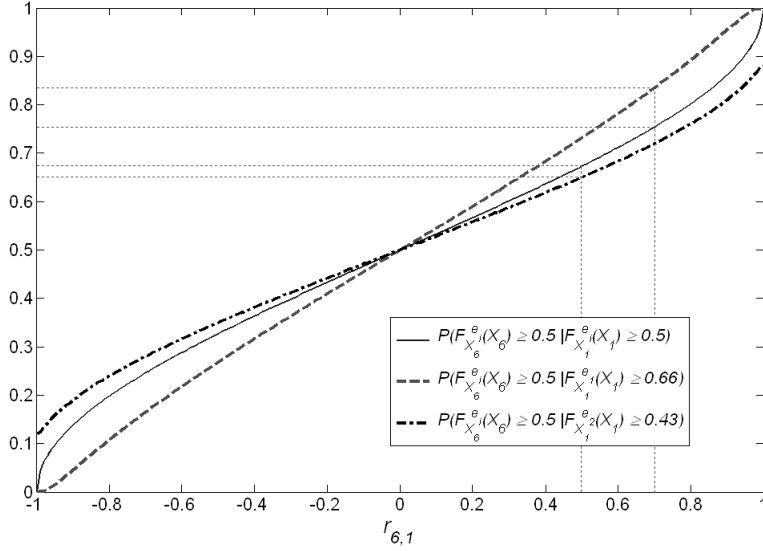


Figure 4.9: $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.5)$, $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.66)$ and $P(F_{X_6}^{e_i}(X_6) \geq 0.5 | F_{X_1}^{e_i}(X_1) \geq 0.43)$ for $r_{6,1} \in (0,1)$

$$\begin{aligned}
 P_1^{DM} &\rightarrow r_{6,1}^{DM} \\
 P_2^{DM} &\rightarrow r_{6,2|1}^{DM} \\
 P_3^{DM} &\rightarrow r_{6,3|1,2}^{DM} \\
 P_4^{DM} &\rightarrow r_{6,4|1,2,3}^{DM} \\
 P_5^{DM} &\rightarrow r_{6,5|1,2,3,4}^{DM}
 \end{aligned} \tag{4.16}$$

4.3 Final Comments

In summary, this chapter describes briefly the classical model for structured expert judgment. It is shown that optimal combination of experts' dependence estimates may be achieved by exploiting the *classical model* of expert judgments in probabilistic statements of the *DM*. For this last step adjusting each experts individual assessments to account for the same events is required.

The elicitation of rank and conditional rank correlations has been presented using probabilistic or statistical measures. The sensitivity of experts' dependence assessments to the choice of the copulae realizing the joint distribution is shown with a comparison between Frank's and the normal copulae.

One of the advantages of NPCDBBN vs. discrete BBNs is that they are more flexible with respect to modelling changes. For example when nodes are added or removed (see section 3.1, Hanea et al. [2006] and [Cowell et al., 1999]). One needs to be cautious in this respect.

Consider for example the BBN from Figure 3.1. Suppose an expert has given

estimates for the 5 rank correlations required through the direct method. She has stated that $P(X_6 \geq x_{6,q_{50}} | X_1 \geq x_{1,q_{50}}) = 0.33$ which corresponds to $r_{6,1} = -0.49$. For the second step she has answered $\frac{r_{6,2}}{r_{6,1}} = -1.68$ which corresponds to $r_{6,2|1} = 0.9297$.

Suppose further that after the elicitation, the analysts have found from data a positive rank correlation between X_1 and X_2 ($r_{1,2} = 0.1$). Of course since the rank and conditional rank correlations attached to the arcs of a NPCDBBN are algebraically independent, $r_{6,1} = -0.49$ and $r_{6,2|1} = 0.9297$ are valid choices in this new model where X_1 and X_2 are not independent. $P(X_6 \geq x_{6,q_{50}} | X_1 \geq x_{1,q_{50}}) = 0.33$ is also a valid choice since this estimate is not constraint by previous answers. However, since $r_{1,2} = 0.1$ then $\frac{r_{6,2}}{r_{6,1}} \in (-1.59, 1.83)$ and hence the estimate previously elicited from expert knowledge is not valid anymore. New estimates would be required from experts. Same kind of situations could happen regardless of the choice of the copula or the elicitation of probabilistic estimates as opposed to rank correlation ratios.

The elicitation of joint distributions by experts, as stated before, is still an issue where not much literature is available. Exploring other methods or building up in those hereby proposed and investigating the effect of assumptions made by the analysts about the models are challenges that remain for future research. The next part of this thesis deals with real applications that use the ideas expressed in the current chapter.

CHAPTER 5

Structured Expert Judgment in Aviation Safety¹

In this chapter three models related to aviation safety developed in the context of the CATS project will be discussed. First the missed approach model is presented. This model was developed by the CATS consortium with two purposes. First to aid in exploring the techniques for elicitation of rank and conditional rank correlations to be used throughout the rest of the project. Second, to be incorporated in the CATS model if required. In the final CATS model presented in Figures 1.13 and 3.16, the controlled flight into terrain or missed approach was included in ESD 35 (see table 3.5) and hence the model presented in section 5.1 was not included in the final BBN. The model is however operational and because of its relevance to the techniques discussed in previous chapters it is included here. After the missed approach model, results from the elicitation in the FCP and the ATPC models will be presented.

5.1 The Missed Approach Model

5.1.1 Introduction to the MA model.

In recent years, the Federal Aviation Authority and the Dutch Ministry of Transport have used causal modelling techniques to investigate integrated safety in air traffic. For this purpose in Roelen et al. [2002] discrete Bayesian Belief Networks (BBN) were fully quantified for the cases of *Missed approach* (MA) and *Flight crew alertness*. However, two disadvantages with discrete BBNs were encountered (see also chapter 3 sections 3.1 and 3.2):

- When variables were discretized into a number of values considered repre-

¹This chapter is based on Morales et al. [2008] and Morales-Nápoles et al. [2009b]

sensitive, the size of the conditional probability tables exploded. As a result a drastic two-valued discretization (usually OK / Not OK) was forced;

- For many variables there was extensive data from the field. When using discrete BBN's, only the nodes without parents could be quantified with field data; other nodes have their marginal distributions determined by the conditional probability tables. Finding conditional probability tables that were compatible with the existing marginal information was a daunting, sometimes hopeless task.

Because of these problems, there was interest in finding a suitable alternative to discrete BBN's. In this section we will concentrate on the model for missed approach.

5.1.2 Description of the MA model.

A missed approach should be initiated when a situation arises that would make the continuation of the approach and landing unsafe. The purpose of a missed approach is to abort a landing in unsafe circumstances to allow the crew to carry out a new approach and landing under safer circumstances. According to [Roelen et al., 2002] “the most common primal causal factor [of approach and landing accidents] was judged to be the omission of action/inappropriate action”. Hence, the missed approach model tries to capture the idea of a *Failure to execute a missed approach when conditions are present*.

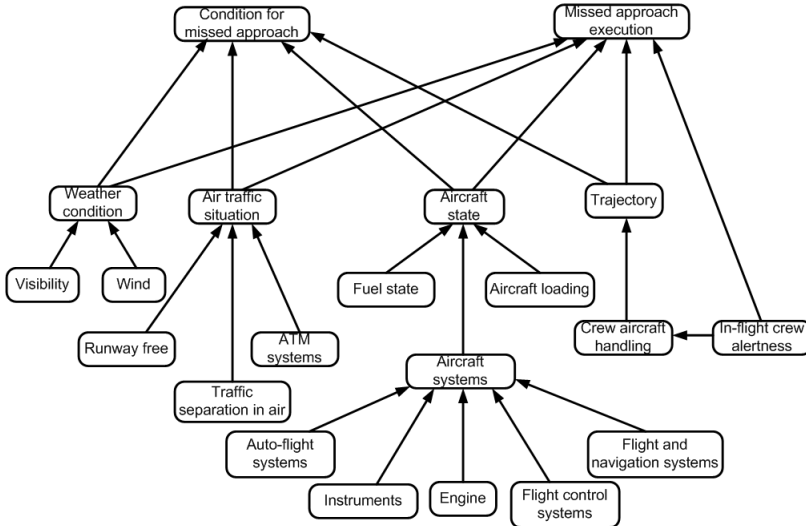


Figure 5.1: Original BBN of the Missed Approach Model.

Figure 5.1 presents the original discrete model for missed approach. All nodes in this model have two states. The top events are:

- **CONDITION FOR MISSED APPROACH** that measures whether there is a condition during the approach or landing phase that requires a missed approach according to the operator's Aircraft Operating Manual, Basic Operating Manual, and/or (inter)national regulations. The states for this node are 'yes' or 'no'. This node is a deterministic node: an unfavorable condition of either one of its parents, alone or in combination will result in a condition for missed approach.
- **MISSED APPROACH EXECUTION** that describes whether the crew executes or does not execute a missed approach under certain circumstances (states 'yes' and 'no'). Compared to the *Condition for missed approach*, this node has an extra parent. The *In-flight crew alertness node* reflects the fact that the final decision to execute a missed approach is taken by the flight crew.

These two nodes are parents to the node *Failure to execute a missed approach when conditions are present* in further modelling which takes into account a possible accident situation. As stated before, some of the variables in Figure 5.1 are more naturally modelled as continuous quantities for example: visibility, wind speed, fuel state, separation in air, etc. The variables are listed below according to their labeling in Figure 5.2. The variables were quantified using field data.

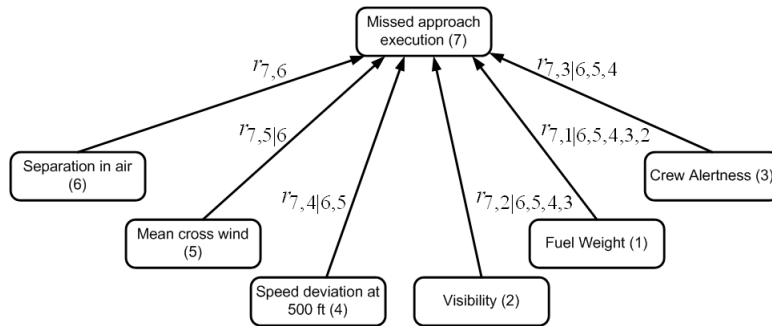


Figure 5.2: *Continuous Version of the BBN for the Missed Approach Model.*

1. **FUEL WEIGHT:** Measured in kilograms and is the remaining fuel at arrival based on data for 172 flights of a Boeing 737 at Schiphol airport.
2. **VISIBILITY:** Measured in meters and is based on a sample of 27 million observations over Europe.
3. **CREW ALERTNESS:** Measured by the Stanford Sleepiness Scale in an increasing scale from 1 to 7, where 1 signifies “*feeling active and vital; wide awake*” and 7 stands for “*almost in reverie; sleep onset soon; struggle to remain awake*” the distribution used for this study comes from field studies by the Aviation Medicine Group of TNO Human Factors in 1,295 flights.

4. SPEED DEVIATION AT 500 FT: Deviation from bug speed² at 500 ft. The data comes from 13,753 approaches of a major European airline.
5. MEAN CROSS WIND: Usually expressed as a combination of speed (in knots) and direction (compass course) of the wind at any direction not favorable for the aircraft, the cross wind distribution comes from 380,000 takeoffs and landings conducted on three large European airports.
6. SEPARATION IN AIR: Longitudinal distance (in nautical miles) between the landing aircraft and the preceding aircraft in the approach path. The distribution was retrieved from a sample size of 2,382 landings at Schiphol airport.
7. MISSED APPROACH EXECUTION: Number of missed Approach Executions per 100,000 flights at Schiphol airport. The expectation of this variable would be an estimate of the unconditional probability of executing a missed approach maneuver.

5.1.3 Expert Elicitation Results of the MA Model

Information about the marginal distributions was available from different sources and the unconditional and conditional rank correlations were elicited with the procedure from section 4.2.2.1 from a single expert at the Dutch National Aerospace Laboratory (NLR) on December 20th, 2005 in a 2.5 hours elicitation. The expert is a pilot for a major European airline and researcher at NLR, in total the expert answered 7 questions.

One marginal distribution for *Missed Approach Execution per 100,000 Flights*, one unconditional rank correlation $r_{7,6}$ and the 5 conditional rank correlations from Figure 5.2 were elicited. For the marginal distribution the expert was asked:

1. *Consider 100,000 thousand randomly chosen flights at Schiphol airport under the current conditions. On how many of these flights will a MISSED APPROACH be executed? (To capture your uncertainty please provide the 5th, 25th, 50th, 75th and 95th percentiles of your uncertainty distribution.)*

A minimal informative distribution with respect to a log uniform background measure was fit with the data provided by the expert. Next, the dependence information was queried starting with the rank correlation $r_{7,6}$ as follows³:

2. *If 50,000 of the flights from the previous question were selected at random, then the number of flights that execute a missed approach should be approximately $\frac{1}{2}$ of your median estimate from previous question. Suppose that instead of selecting those 50,000 flights at random, you select those where*

²The bug speed is the target reference speed for the approach (calculated by the aircraft crew) plus allowance for conditions such as crosswind.

³The specification of the rank correlations required in the model presented in Figure 5.2 is not unique (see equation (3.3)). For example instead of eliciting the (un)conditional rank correlations presented Figure 5.2, one could also specify $\{r_{7,5}, r_{7,6|5}, \dots\}$. In this case the order in which the variables entered the model was provided by the expert.

SEPARATION IN AIR is above its median value. What is your probability that, in this situation, the number of MISSED APPROACH executions will be larger than $\frac{1}{2}$ of your 50th percentile estimate provided in the previous question?

The assessment from question 2 is equivalent to an estimate of $P_1 = P(F_{X_7}(X_7) \geq 0.5 | F_{X_6}(X_6) \geq 0.5)$. The expert's assessment for this question was $P_1 = 0.15$ that from Figure 4.2 corresponds to $r_{7,6} = -0.88$. The conditional rank correlation $r_{7,5|6}$ was elicited as follows:

3. If 50,000 of the flights from question 1 were selected at random, then the number of flights that execute a missed approach should be approximately $\frac{1}{2}$ of your median estimate from question 1. Suppose that instead of selecting those 50,000 flights at random you select those where both SEPARATION IN AIR and MEAN CROSS WIND are both above their median values. What is your probability that, in this situation, the number of MISSED APPROACH executions will be larger than $\frac{1}{2}$ of your 50th percentile estimate provided in question 1? (bearing in mind that your new assessment should be $\in (0, 0.3)$)

The expert's assessment for question 3 is equivalent to an estimate of $P_2 = P(F_{X_7}(X_7) > 0.5 | F_{X_6}(X_6) > 0.5, F_{X_5}(X_5) > 0.5)$. The expert's answer to question 3 was $P_2 = 0.18$, and, with the methods described in 4.2.2.1 the corresponding value for $r_{7,5|6} = 0.20$ was found. The upper and lower bounds provided in question 3, i.e the interval $(0, 0.3)$ where also computed on-line with the methods described in section 4.2.2.1.

Conditional Probability		Bounds for P_i^a	Correlation	
P_1	0.15	(0, 1)	$r_{7,6}$	-0.88
P_2	0.18	(0, 0.3)	$r_{7,5 6}$	0.20
P_3	0.20	(0.01, 0.35)	$r_{7,4 6,5}$	0.12
P_4	0.24	(0.02, 0.38)	$r_{7,3 6,5,4}$	0.23
P_5	0.22	(0.04, 0.45)	$r_{7,2 6,5,4,3}$	-0.11
P_6	0.24	(0.03, 0.40)	$r_{7,1 6,5,4,3,2}$	0.11

^aEach P_i , $i = \{1, \dots, 6\}$ sequentially adds variables to the model, for instance $P_1 = P(F_{X_7}(X_7) > 0.5 | F_{X_6}(X_6) > 0.5)$, $P_2 = P(F_{X_7}(X_7) > 0.5 | F_{X_6}(X_6) > 0.5, F_{X_5}(X_5) > 0.5)$, $P_3 = P(F_{X_7}(X_7) > 0.5 | F_{X_6}(X_6) > 0.5, F_{X_5}(X_5) > 0.5, F_{X_4}(X_4) > 0.5)$, and so on.

Table 5.1: Results from Expert's Elicitation of Conditional Rank Correlations

The rest of the conditional rank correlations were elicited in a similar way by sequentially adding information about the variables entering the conditioning set. The expert was provided with the upper and lower bounds for P_i ($i = 1, \dots, 6$) at each step in the elicitation only after he had provided his estimates to check for consistency. This way of assessing conditional rank correlations helped the expert understand the meaning of dependence and increased his "buy in" in the method. The results of the elicitation for the 6 arcs in the BBN for missed approach are summarized in table 5.1.

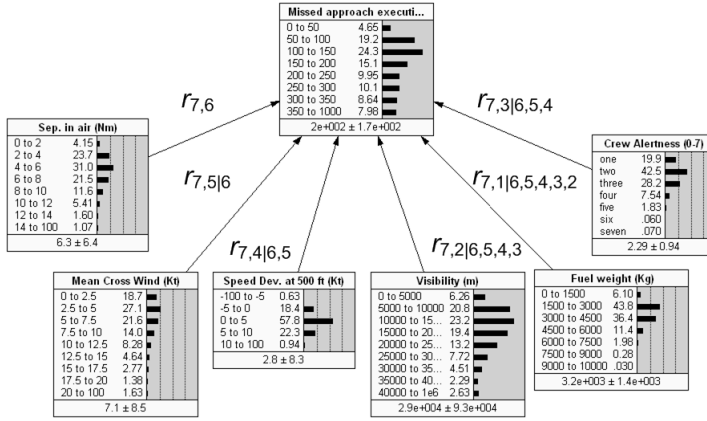


Figure 5.3: Discretized BBN of the Missed Approach Model with continuous quantities in Netica.

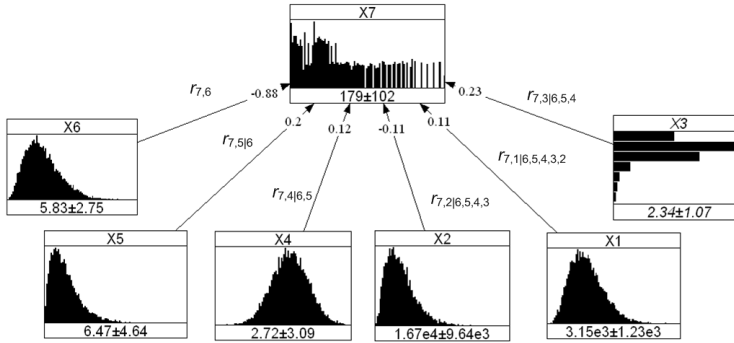


Figure 5.4: Continuous BBN of the Missed Approach Model with continuous quantities in UNINET.

5.1.4 Updating beliefs in the MA Model

In Hanea et al. [2006] techniques to efficiently deal with the joint distribution when evidence becomes available (updating the BBN) are discussed. The two possibilities are:

- **THE HYBRID METHOD.** To work with this method the information from table 5.1 together with the marginal distributions for each variable were used to create a large sample file by means of the normal copula. A discrete version of the model can be built in order to take advantage of commercial software to perform fast updating each time a new policy is evaluated.
- **THE NORMAL COPULA VINE APPROACH.** Since according to the methods described in sections 3.2 and 4.2.2 all calculations are performed on a joint normal vine, the conditional distribution can be computed analytically.

To illustrate the Hybrid Method the professional software Netica[©] will be used. For the normal copula vine approach the recently developed software application UNINET⁴ will be used. Figures 5.3 and 5.4 show the representation of the BBN for missed approach execution in Netica and UNINET respectively. The rank correlations are included to stress the fact that both versions of the model introduced in Figure 5.2 preserve the dependence structure elicited from the expert.

If instead of eliciting the 6 quantities in table 5.1, the expert would have been asked to fill in the conditional probability table for X_7 missed approach execution per 100,000 flights with the discretization of its parent variables as in Figure 5.3, then the expert would have had to provide over 1.2 million conditional probabilities (equation (3.2)) that need to be consistent with the marginal distribution from Figure 5.3 and still reflect the correct dependence information.

Figure 5.5 presents the distribution of missed approach executions per 100,000 flights given separation in air $\in (0, 2)$ Nm and the mean cross wind $\in (17.5, 20)$ Kt from Netica. The reader may compare this distribution with the unconditional distribution in Figure 5.3. The unconditional mean is 200 Missed Approach executions per 100,000 flights (standard deviation of 170), while the mean of $(X_7 | X_6 \in (0, 2), X_5 \in (17.5, 20))$ is 470 Missed Approach executions per 100,000 flights (standard deviation 290).

Figure 5.6 presents the same conditional distribution as Figure 5.5 computed analytically in UNINET. The unconditional distribution of X_7 is shown in grey behind the black histogram representing the conditional distribution of $X_7 | X_6 = 2, X_5 = 20$. In this case the conditional mean is 379 with standard deviation 47.7 missed approaches per 100,000 flights. While in Netica (Figure 5.5) one can only condition in discretized states of each variable, UNINET allows for conditioning in point values. This is the usual way in which evidence becomes available in real situations.

⁴UNINET has been developed for the CATS project commissioned by the Dutch Ministry of Transport. Currently UNINET supports both the Hybrid Method with the support of Netica and the analytical updating. The software is still under development.

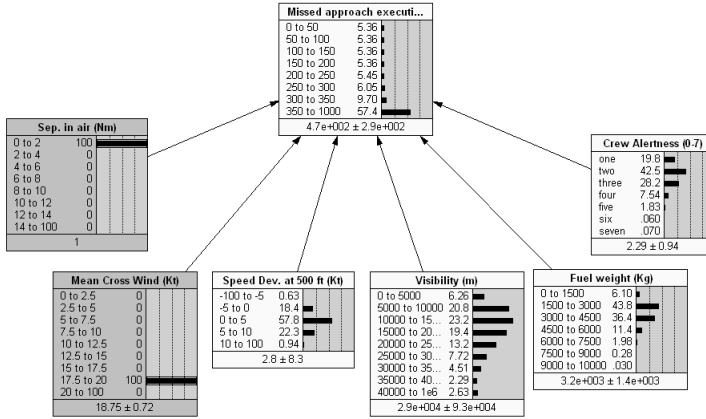


Figure 5.5: Conditional Distribution of Missed Approach Executions per 100,000 flights given $X_6 \in (0, 2)$ Nm and $X_5 \in (17.5, 20)$ Kt.

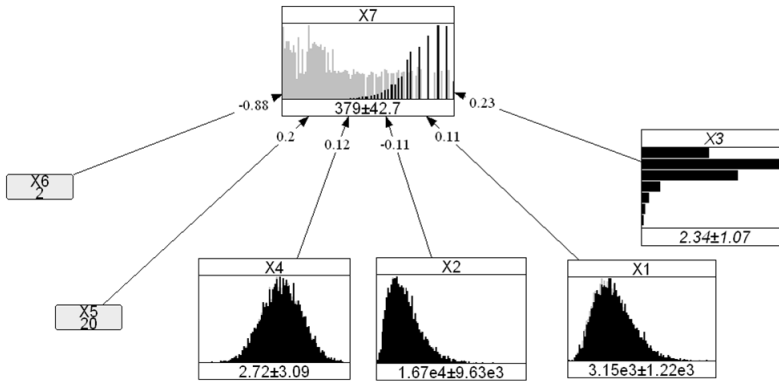


Figure 5.6: Conditional Distribution of Missed Approach Executions per 100,000 flights given $X_6 = 2$ Nm and $X_5 = 20$ Kt.

500,000 samples from the joint distribution represented by figures 5.4 and 5.6 were obtained with UNINET. The cumulative distribution function of X_7 and $X_7|X_6 = 2, X_5 = 20$ were obtained and shown in Figure 5.7. Observe that both Netica and UniNet show that $P(X_7 > 350) \approx 8\%$. In the conditional distribution computed with Netica this probability increases to $\approx 57\%$ while the analytical approach from UniNet shows that this value is as big as $\approx 75\%$.

The application to Missed Approach demonstrated that it is possible to elicit unconditional and conditional rank correlations with intuitively meaningful conditional probabilities of exceedence. The results motivate the choice of the analytical updating (UNINET) vs. the hybrid method with Netica. The next two sections present results regarding the elicitation for the human performance models used in CATS. For the two models presented next more than one expert participated in the elicitation. This is in contrast with the elicitation in the MA

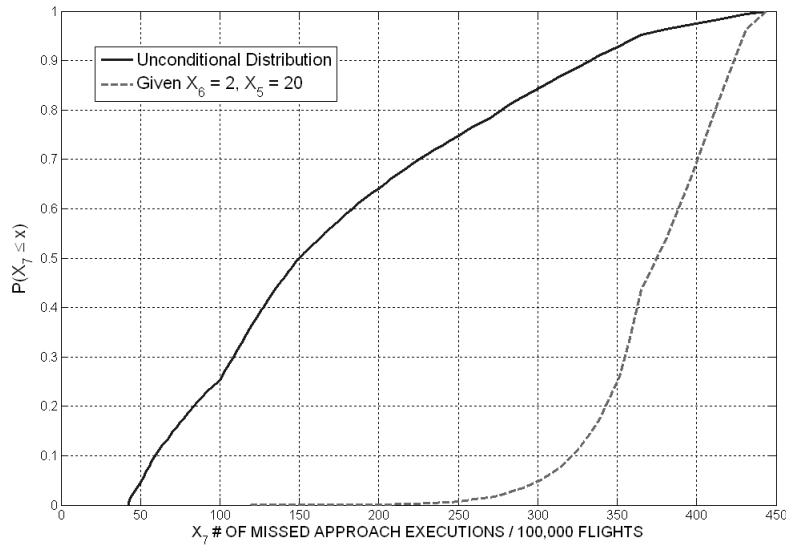


Figure 5.7: Cumulative Distribution function of X_7 and $X_7|X_6 = 2$ Nm and $X_5 = 20$ Kt.

model where as stated earlier only one expert participated.

5.2 The Flight Crew Performance Model

5.2.1 Expert Elicitation Results of the FCP Model

An elicitation protocol was designed for obtaining the marginal distributions shown in table 3.2 and the dependence information required by the model (Figure 3.5). A total of 4 marginal distributions, 11 questions for retrieving the dependence information and 8 calibration variables were asked to each expert⁵. Summary results from the classical method are presented in table 5.2. Calculations are performed with the EXCALBIUR software developed at the TU Delft.

Table 5.2 shows the resulting scores for the five experts in this study plus two *DMs*⁶. The first column gives the expert's id; the second column gives the calibration score. The ratio of highest to lowest score is about 13,000. It will be noted that experts B and D had a score corresponding to a p-value above 5%. Scores of experts E and C are marginal and for expert A rather low. Calibration scores in the order 0.001 would fail to confer the requisite level of confidence in the results.

The information scores for all items and for calibrations items are shown in

⁵In total 14 rank correlations are required in Figure 3.5 however $r_{10,6}$ and $r_{10,7|6}$ where chosen such that $r_{10,6}$ and $r_{10,7}$ would be equal, positive and as large as possible. Variable 13 was elicited with a single expert after the elicitation described in this section was performed. See Singuran [2008] for a more detailed description of node 13.

⁶The IWDM is not shown because in this case it is equal to the GWDM

Experts' Id.	Calibration score	Information score (all var.)	Information score (cal. var.)	Un-Normalized weights	Normalized weights (without DM)	Normalized weights (with DM)
A	0.02651	0.7119	0.4991	0	0	0
B	0.6638	0.95	0.574	0.381	1	0.5
C	0.001547	1.016	0.9689	0	0	0
D	0.185	1.317	1.029	0	0	0
E	5.115×10^{-05}	1.049	1.06	0	0	0
GWDM	0.6638	0.95	0.574	0.381	-	0.5
EWDM	0.2224	0.1046	0.09945	0.02212	-	-

Table 5.2: *FCPM Experts' Performance. Significance level: 0.6638 (Global Weights DM).*

columns 3 and 4 respectively. It will be noted that the overall information scores are quite similar, within a factor 2. In this case the expert with the best calibration score (B) also has one of the lowest information scores for the calibration variables which is a recurrent pattern. Weights are constructed by the product of columns 2 and 4. If these weights were normalized and used to form weighted combinations, experts A, D and B would be influential with (2.25, 32.49 and 64.98 per cent respectively).

As it may be seen in table 5.2 the EWDM is better calibrated than each expert individually except expert B. However information scores derived from the EWDM are poor. They are the lowest amongst the 5 experts in both all variables and calibration questions alone.

Table 5.2 also shows that the optimized decision maker gives all weight to expert B. The calibration score of the GWDM is about 3 times higher than the EWDM and the information score is about 9 times higher over all variables and 5.7 times higher in calibration questions alone. If no optimization was performed in the GWDM then, after normalization of the weights, experts A, D, B and the GWDM (not optimized) would be influential with 2.06, 29.66, 59.32 and 8.71 per cent respectively. Though more experts enter the pool, the calibration score of the (not optimized) GWDM is 4.58 times smaller than that of the optimized GWDM. Similarly the information scores in all variables and calibration variables are 2.57 and 1.48 times higher in the optimized GWDM. The recommended choice for the DM is the GWDM as it achieves better performance than the EWDM and the GWDM without optimization combinations. Next, results of the dependence information for the FCP model will be discussed.

(Un)Conditional Rank Correlation	Value	(Un)Conditional Rank Correlation	Value
$r_{7,1}$	-0.95	$r_{10,7 6}$	1.00
$r_{7,3 1}$	0.86	$r_{14,10}$	0.30
$r_{7,2 1,3}$	0.24	$r_{14,11 10}$	-0.32
$r_{6,5}$	-0.95	$r_{14,8 10,11}$	0.46
$r_{6,3 5}$	0.86	$r_{14,12 10,11,8}$	0.18
$r_{6,4 5,3}$	0.24	$r_{14,9 10,11,8,12}$	0.19
$r_{10,6}$	0.71	$r_{14,13 10,11,8,12,9}$	0.16

Table 5.3: *GWDM Dependence estimates for the FCP Model.*

5.2.2 Dependence in the FCP Model

To elicit the rank correlations a total of 11 questions were asked to each expert. These were similar to those in relation 4.2 in subsection 4.2.2.1. From previous subsection (5.2.1) it was observed that the global weight decision maker gave weight 1 to expert B and hence no combination was necessary. The results of the dependence elicitation are presented in table 5.3.

As explained in Hanea [2008, ch.5], the determinant of a correlation matrix is a measure of the amount of ‘linear dependence’ in a joint distribution. If variables are uncorrelated it takes value 1, and 0 when they are completely correlated. The determinants of the correlation matrix of each expert are presented in the second column of table 5.4. It may be observed that the GWDM dependence estimates shown in table 5.3 present the rank correlation matrix with the lowest value of the determinant among experts (expert B). One may see that there is a factor 70 between the highest and lowest determinant between experts.

Experts’ Id	Expert’s Determinant
A	4.936×10^{-6}
B	2.011×10^{-6}
C	7.173×10^{-4}
D	1.427×10^{-4}
E	7.562×10^{-5}

Table 5.4: *Experts’ Correlation Matrices Determinant*

From this last subsection it may be seen that the elicitation of rank and conditional rank correlations through conditional probabilities of exceedance from domain experts is possible. Next a similar model for Air Traffic Control performance will be presented.

5.3 The Air Traffic Control Performance Model

5.3.1 Expert Elicitation Results of the ATCP Model

An elicitation protocol was designed for obtaining the dependence information required by the model shown in Figure 3.6. In total 1 marginal distribution⁷, 5 questions for retrieving the dependence information and 12 calibration variables were asked to 6 experts⁸. Estimates of one expert could not be used because of inconsistent estimates (ratios outside the allowable range). Summary results from the classical method are presented in table 5.5. Calculations are performed with the EXCALIBUR software developed at the TU Delft.

Experts' Id.	Calibration score	Information score (all var.)	Information score (cal. var.)	Un-Normalized weights	Normalized weights (without DM)	Normalized weights (with DM)
A	0.1012	0.5633	0.5034	0.05095	0.5208	0.2004
B	0.04706	1.03	0.9588	0.04512	0.4612	0.1803
C	0.00131	1.423	1.349	0.001767	0.01806	0.006987
D	2.795×10^{-9}	1.669	1.655	0.0	0.0	0.0
E	2.501×10^{-6}	1.017	0.9624	0.0	0.0	0.0
GWDM	0.6827	0.3094	0.2271	0.1551	-	0.6131
IWDM	0.2441	0.4441	0.3757	0.0917	-	-
EWDM	0.1242	0.2662	0.2472	0.0307	-	-

Table 5.5: ATC Experts' Performance. Significance level: 0.00131 (Global Weights DM).

Table 5.5 shows the resulting scores for the five experts in this study plus three Decision Makers. The first column gives the expert's id; the second column gives the calibration score. The ratio of highest to lowest score among the 5 experts is about 3.62×10^7 (1.30×10^4 in the case of the FCP model experts). Only expert A had a score corresponding to a p-value above 5%. Scores of experts D and E are marginal and for expert C is rather low.

The information scores for all items and for calibrations items are shown in columns 3 and 4 respectively. It will be noted that the overall information scores are quite similar, within a factor 3. In this case (as in the FCP model) the expert with the best calibration score (A) also has the lowest information scores. The fifth column gives the “un-normalized weights”; this is the product of columns 2 and 4. If this column were normalized (among the experts) and used to form weighted combinations, experts A, B and C would be influential with (52.07, 46.11 and 1.80 per cent respectively).

In Table 5.5 the 8th expert is identified as “EWDM”. It may be observed that the EWDM is better calibrated than each expert individually. However information scores derived from the EWDM are poor. They are the lowest amongst

⁷The marginal distribution of error probability was elicited from each expert. Later on in the project, data about the marginal distribution became available and it was used instead of the one elicited from each expert.

⁸All experts are different from those participating in the FCPM. Only 10 calibration variables could be used for the combination because of lack of response from some experts.

all experts (that is including the EWDM as an expert) in both all variables and calibration questions alone.

For the GWDM all experts with a calibration score less than the significance level (0.00131) found by the optimization procedure are unweighed as reflected by the zeros in columns 5, 6 and 7 in table 5.5.

From table 5.5 one can see that after the optimization procedure is applied, 3 experts have non-zero weight. One can see that the calibration score of the GWDM is about 5.5 times higher than the EWDM. The information scores are comparable for both decision makers in both all variables and calibration variables alone. The calibration score of the GWDM is about 2.8 times larger than the IWDM. The IWDM is slightly more informative than the GWDM. However the gain in information is not a sufficient argument to justify a preference of the IWDM over the GWDM.

The recommended choice of the decision maker is the global weight decision maker as it achieves better performance than the equal weight and item weight combinations. Future analysis will be performed based on the GWDM.

5.3.2 Dependence in the ATCP Model

As stated before, to elicit the rank correlations in Figure 3.6 a total of 6 questions were asked to each expert. Experts were asked to rank each variable according to the largest unconditional rank correlation with ATC error in absolute value⁹. Then for the variable which they regarded as having the largest rank correlation in absolute value, experts would assess the usual probability of exceedence. Finally, ratios of each of the remaining rank correlations to the one assessed through a probability of exceedence were asked. This method is described in subsection 4.2.2.2 and relation 4.9.

From subsection 5.3.1 it could be observed that the GWDM was the recommended choice for combining experts' opinions in the ATC performance model. The combination of the three expert's individual assessments was done as described in section 4.2.3. The results of the combination scheme are presented in table 5.6.

(Un)Conditional Rank Correlation	Value	(Un)Conditional Rank Correlation	Value
$r_{7,1}$	-0.180	$r_{7,4 1,2,3}$	-0.060
$r_{7,2 1}$	-0.206	$r_{7,5 1,2,3,4}$	0.020
$r_{7,3 1,2}$	0.134	$r_{7,6 1,2,3,4,5}$	0.180

Table 5.6: *GWDM Dependence estimates for the ATC Model.*

The GWDM's determinant is the second largest among the 6 experts (including the DM itself) in table 5.7. This may be explained because the GWDM is dominated by experts A and B. Expert's A opinion, which has the largest deter-

⁹The ranking from each expert could be different however once the full correlation matrix of each expert is determined any probabilistic statement may be computed.

minant across experts, contributes to the GWDM's dependence estimates with 52.08% (table 5.5 column 6). Expert B, whose determinant is also large, contributes 46.12%. On the other hand expert C has the lowest determinant across experts, however his opinion contributes 1.8%. The ratio of highest to lowest determinant (column 2 in table 5.7) is about 4.5. This is comparable to the ratio of the GWDM's determinant to expert's C determinant which is 4.3. These two ratios are small compared to those observed in the FCP model where differences of the order of 70 were observed.

Experts' Id	Expert's Determinant
A	0.932304
B	0.751821
C	0.206152
D	0.344658
E	0.849824
GWDM	0.894683

Table 5.7: *Experts' Correlation Matrices Determinant & Comparison Vs. Optimized determinants.*

In summary, from this section it may be seen that the elicitation of rank and conditional rank correlations with the direct method described in section 4.2.2.2 is possible. Experts' belief that the relationship of the variables in table 3.3 to the ATC error is highly non monotonic is expressed by the high values of the determinant of the correlation matrices of each expert presented in table 5.7. It is worth noting that from comparing tables 5.7 and 5.5 one may suppose that experts tend to have a negative correlation between the determinant of the rank correlation matrix in their individual BBN and their information score. However, tables 5.4 and 5.3 show the opposite pattern.

NPCDBBNs have found application in this thesis outside the aviation sector. In the next chapter an application in earth dams safety in central Mexico will be presented. Next chapter shows the flexibility of Bayesian networks as tools for modelling risks in different sectors.

CHAPTER 6

Dams Safety in the State of Mexico¹

This chapter describes a demonstration model for earth dams safety. The aim of the project was to develop a model to investigate environmental factors that could contribute to different failure modes in earth dams in the highlands of central Mexico. This model would serve as a demonstration model for a larger model that would include larger structures and a countrywide coverage. NPCDBBNs were identified as an appropriate tool for this research. The classical method for structured expert judgment was used for model quantification in the absence of field data. The project was financed by COMECYT (State of Mexico Council for Science and Technology) for the Civil Engineering Faculty of the Autonomous University of the State of Mexico. Our role in the project was to provide technical support in the use of continuous BBNs and structured expert judgment for the development and later use of the model.

6.1 Introduction

A dam is an artificial obstruction to natural water flows constructed for one or more specific purposes such as accumulating water for farm irrigation, generating electricity, creating artificial lakes for navigation and leisure activities, supplying water to cities or industry, preventing floods, diverting river flows into canals and keeping a reserve of fresh water.

Small dams are structures of less than 15 meters height. Large dams, in contrast, are those with 15 meters or more from the foundation to the crest or, between 5 to 15 meters with a capacity of more than 3 million m^3 . Based on their structure, they can be categorized as: embankment (earth dams), gravity, arch and buttress dams [Emiroglu et al., 2002].

Regardless of their construction materials, these buildings may fail. Figure

¹This chapter is based on Delgado-Hernández et al. [2009] and Morales-Nápoles and Delgado-Hernández [2009]

6.1 shows the number of dams failures per 10 years period from 1891-1990 and the proportion of total number of failures corresponding to embankment dams [ICOLD, 1995, pp.38-45]. For every ten year period, between 50% (1891-1990) and 91.67% (1971-1980) of the failures correspond to embankment dams. However, Donnelly [2006] stated that embankment dams are the most common type of water retaining structures. In this sense, for the same data set, he noted that 2.6% of the concrete buttresses failed compared to 1.2% of embankments, 0.7% of concrete arch and 0.3% of concrete gravity dams.

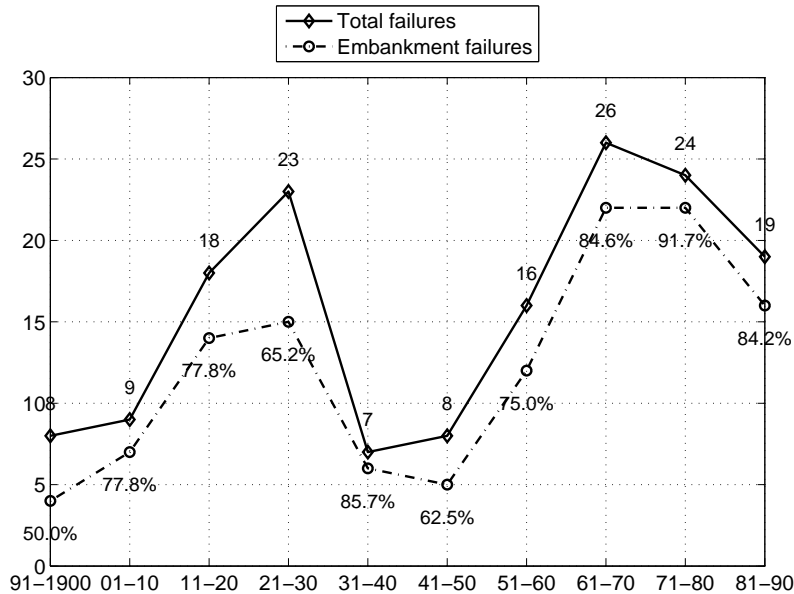


Figure 6.1: Number of dams failures per 10 years period from 1891-1990. With data from [ICOLD, 1995, pp.38-45]

The impacts of a dam collapse can be enormous, encompassing the destruction of private housing, transport and public infrastructure, industrial facilities and agricultural land. The losses may also include human harm and serious disruptions in infrastructure operation, leading to significant total economic damages.

From the end of October and up to the end of November 2007 flooding was produced in about 70% of the Tabasco flatlands affecting more than 1 million people. The main cause of the flooding in Tabasco was the severity of the runoff resulting from the uncontrolled De La Sierra basin and the coincidence and duration of intense precipitation. Because of the exceptional rainfall, the release of water through the spillway at Peñitas Dam additional to electricity generation at full had to be performed. Though, this operation was considered to be appropriate, damages were enormous. The consequences were great in part due to the vulnerability of the region and the lack of adequate and sufficient infrastructure. As part of the recommendations the use of an integrated modelling

system including hydrometeorological forecasting, rainfall-runoff relationships and *dam operation* was proposed. For more details see Aparicio et al. [2009].

Literature reports studies within the dam industry. Most are centered on the analysis of specific failure modes, and a few on mathematical models for dam risk assessment, that make use of continuous BBNs (see for example [FEMA, 2007] and [FEMA, 2008]). The central motivation for carrying out this investigation was the lack of systematic research to date within the continuous BBN framework.

Overall the study aims to develop a model to assist dam engineers, in particular those in Mexico, on their risk assessment practices. Selecting embankment dams, and more specifically earth dams on the basis of their abundance, has provided a focus. In this sense, the model is only limited to the analysis of natural events (e.g. excessive rainfall or earthquakes) and disregards those intentionally produced (e.g. terrorism or bomb attacks). It should be noted that this work is the starting point for a bigger research project to develop a comprehensive model for assessing risk in various types of dams in Mexico. The model will be referred to as the Dams Safety demonstration model or simply DS model.

The next section presents the definitions of the concepts that have been used to develop the model. Then, the selection process of seven dams in central Mexico is described. The criteria that helped create the model will also be briefly described together with its constituent elements. The application of the model in the seven earth dams located in Mexico will be illustrated as well as some final remarks and recommendations.

6.2 Earth Dams in the State of Mexico

Before we continue, we briefly introduce the components of a dam. Figure 6.2 shows a simplified graphical representation of such a structure showing its main elements. They are: crest, reservoir, upstream slope (embankment), downstream slope (embankment), river, outlet pipe, and spillway. Formal definitions may be found in FEMA [2004].

In order to develop the model some dams located in the State of Mexico were chosen. The State of Mexico is a territory in central Mexico that surrounds Mexico City to the east, north and west. The criteria for such a selection were as follows:

- (i) height: between 15 and 30 *m*
- (ii) age: more than 30 years old and,
- (iii) construction material: earth and rockfill dams.

These three conditions have significant influence in collapse events [Foster et al., 2000] and [ICOLD, 1995].

In the exercise, seven dams were identified: Embajomuy (E), San Joaquín (SJ), José Trinidad Fabela (JTF), Dolores (D), José Antonio Alzate or San Bernabé (JAA), Ignacio Ramírez or La Gavia (IR), and El Guarda (EG). Their heights range from 15 to 24 *m*, their ages from 36 to 66 years, and their capacities from

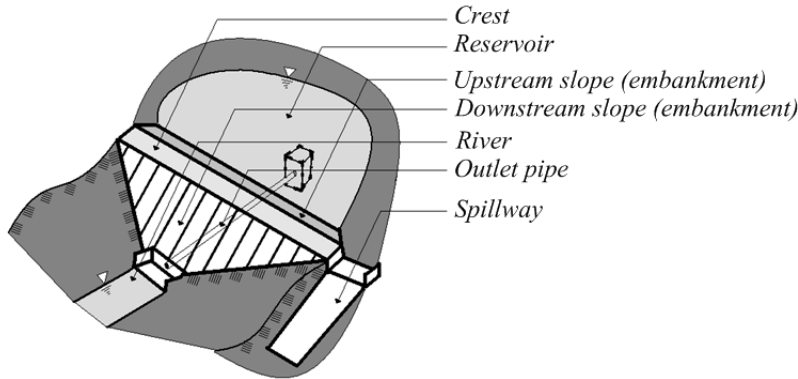


Figure 6.2: *Simplified representation of a dam with main elements.*

52,000 to 225,000 m^3 . Irrigation, flooding prevention and hydroelectric power generation, can be listed among their main purposes [SRH, 1976].

After visiting each structure, it became evident that maintenance activities are not frequent. Because of its relative location with respect to inhabited communities downstream, the JAA dam is perhaps the most important structure of the ones under study. All seven dams under study share the same basic design characteristics being the main difference amongst them the amount of people living downstream. With regard to infrastructure it is common for them to have highways, electrical transmission towers and some urban settlements downstream. In addition land used for agricultural purposes is also observed in the region of interest.

6.3 Description of the DS model.

6.3.1 Model variables & graph

Ten variables were identified as most relevant for this study. Their description, units and source of the marginal distributions is detailed next.

1. **SEISMIC FREQUENCY.** It refers to the distribution of earthquakes > 5.5 per year, in Richter magnitude scale, between 2000 and 2008 for the locations of interest. Data is available from the Mexican National Seismographic System.
2. **RAINFALL RATE.** It refers to the average value of the seven-basin (i.e. the area of influence of the 7 dams of interest) five-days moving averages in mm/day. Data is available from “ERIC” Mexican database from 1961 to 1998. A short overview of ERIC may be found in Carrera-Hernández and Gaskin [2008].

3. MAINTENANCE. Is the number of years between maintenance activities which would lead the dam to an “as good as new” condition. The marginal distribution comes from structured expert judgment.
4. OVERTOPPING. Water level from the crest during an event in which such a level may increase beyond the total embankment height (mm). Marginal distribution obtained from expert judgment.
5. LANDSLIDE. Distribution of the security factors (resisting moment/causing moment), for each of the seven dams based on their design geometrical features. The so called “Swedish method” is used for calculating such factors [SRH, 1976].
6. PIPING. Distribution of water flowing through the embankment that causes its internal erosion apart from the spillway and outlet pipe torrents (lt/sec). Data comes from expert judgment.
7. BREACHING. Refers to the average breach width i.e. the mean of both superior and inferior breach widths, due to embankment’s crest erosion (m). Calculated with the methods reported in Wahl [1998] with data from SRH [1976].
8. FLOODING. Average water level per day in the downstream flooded area during a dam failure event. Its marginal distribution is built by means of expert judgement (mm/day).
9. HUMAN COSTS. Both public and private total costs over a time period equivalent to the maximum human remaining life span, due to all possible damages, health and life losses, caused by a flooding, consequence of a dam failure. It is measured in current USD and obtained through expert judgment.
10. ECONOMIC COST. Both public and private total costs, due to all possible damages in infrastructures (e.g. schools, hospitals, bridges, roads, transport systems), fields (e.g. farms, crops), housing, supply, commercial and entertainment centers, caused by a flooding, consequence of a dam failure. It is measured in current USD and obtained through expert judgment.

Variables are broadly grouped into three categories: contributing factors (seismic frequency, rainfall rate and maintenance), failure modes (landslide, piping, overtopping and breaching), and consequences (flooding, human and economic cost). The model was built based on such configuration.

Figure 6.3 shows a scheme of the model, which includes both the requirements previously established and the variables recognized. The Figure was taken from UNINET. Arcs representing rank and conditional rank correlations between variables are shown.

Arcs between both human costs and total costs, and economic costs and total costs lack a rank correlation because the total costs are simply the sum of human

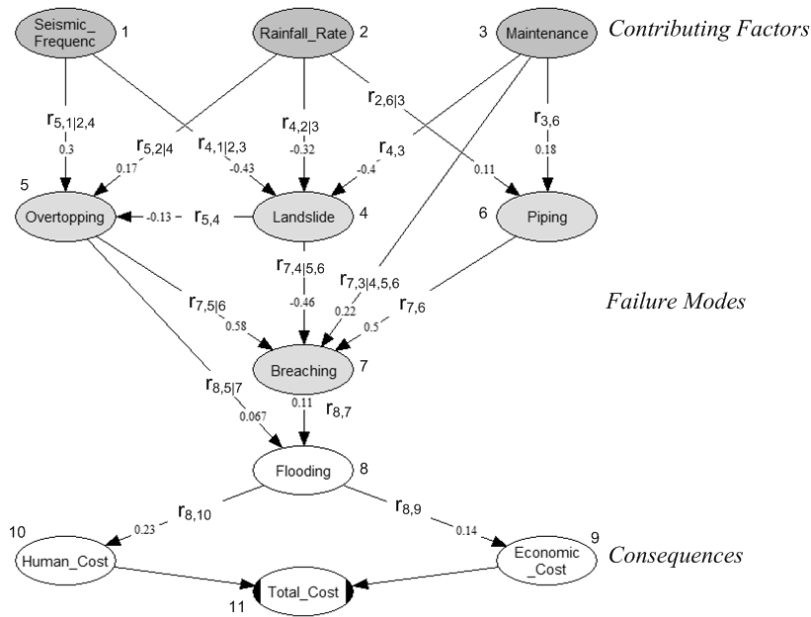


Figure 6.3: Model for Earth Dam's Risk Assessment.

and economic costs and hence the relationship is functional. To distinguish probabilistic from functional nodes in the graph, two vertical lines in the extremes of the node are drawn. It should be recognized that there might be more interactions among the nodes in Figure 6.3. However, they have not been expressed in the model for the sake of simplicity and because it is thought that their exclusion does not affect the patterns of relationship between the main variables.

6.3.2 Expert Elicitation Results of the DS Model

In total four experts participated in the elicitation. Three of the experts hold positions at the National Water Commission (CONAGUA) in the State of Mexico. The other expert holds a position in the Municipality of Zinacantepec as water manager. Two of the experts are lecturers in civil engineering at the Autonomous University of the State of Mexico (UAEM). A workshop was held on July 18 2008 in the faculty of engineering of UAEM. Individual interviews were held with each expert during the months of July and August according to experts' availability. The questionnaire included 6 questions to elicit marginal distributions (see section 6.3.1), 20 to elicit the rank and conditional rank correlations from Figure 6.3 and 20 calibration variables.

As mentioned in chapter 4 calibration variables are those known to the analyst but not to the experts at the moment of the elicitation. These are used to measure experts' performance as uncertainty assessors. One example of a calibration

variable for this elicitation is: *Consider the 7 day moving average of the daily average precipitation (mm) from the two stations related to Embajomuy Dam from January 1961 to August 1999 in ERIC II of CONAGUA [Carrera-Hernández and Gaskin, 2008]. What is the maximum moving average for the time period of reference?* In total three questions about seismicity, four over general characteristics of the 7 selected dams, nine over precipitation and two about water discharge were used as calibration variables. The results of the expert elicitation are summarized in table 6.1. Calculations are performed with the EXCALIBUR software developed at the TU Delft.

Experts' Id.	Calibration score	Information score (all var.)	Information score (cal. var.)	Un-Normalized weights	Normalized weights (without DM)	Normalized weights (with DM)
A	0.00014	0.9154	0.8259	0.0001141	0.9973	0.1404
B	3.588×10^{-14}	2.245	2.196	0	0	0
C	3.223×10^{-9}	1.507	1.576	0	0	0
D	3.57×10^{-7}	0.9291	0.8722	3.114×10^{-7}	0.0027	0.00038
GWDM	0.0009	0.8415	0.7578	0.0006981	-	0.8592
EWDM	0.07164	0.2976	0.3283	0.02352	-	-

Table 6.1: *DS Experts' Performance. Significance level: 3.57×10^{-7} (Global Weights DM).*

Table 6.1 shows the resulting scores for the four experts in this study plus two *DMs*. The first column gives the expert's id; the second column gives the calibration score. The ratio of highest to lowest score among the 4 experts is about 3.85×10^9 . For the air traffic control performance model this ratio is 3.62×10^7 and 1.30×10^4 in the case of the flight crew performance model (tables 5.3.1 and 5.2.1). In this case no individual expert had a score corresponding to a p-value above 5%.

The information scores for all items and for calibrations items are shown in columns 3 and 4 respectively. Information scores in columns 3 and 4 are within a factor 2.5 for the four experts. Expert B had the lowest calibration score, however was also the most informative. In contrast, expert A had the largest calibration score and is the least informative. This is a recurrent pattern, however low informativeness does not translate automatically into better calibration [Cooke and Goossens, 2008, p.669]. The fifth column gives the “un-normalized weights” with the GWDM² This is the product of columns 2 and 4. Experts with a calibration score less than the significance level are weighted with zero. If this column were normalized (among the experts) and used to form weighted combinations, experts A and D would be influential with 99.73% and 0.27% respectively.

The GWDM is better calibrated than each expert individually, however its information scores are lower than the information scores of each expert individually. The calibration score of the GWDM is still lower than 5% which fails to confer the requisite level of confidence for the study. Last row of table 6.1 shows

²If the GWDM without optimization would be used instead, the results in table 6.1 would be virtually unchanged. Results for the IWDM in this case are equal to the GWDM.

the EWDM. This is the only expert with a p-value above 5%. For this reason the EWDM is the recommended choice and further analysis will be conducted with this combination. The cost of this choice is in the information scores (about 3 times smaller than the GWDM). Next results about the dependence elicitation are presented.

6.3.3 Dependence in the DS Model

To elicit the rank correlations in Figure 6.3 a total of 20 questions were asked to each expert. For each child node experts were asked to rank parent variable according to the largest unconditional rank correlation with the child in absolute value. Observe that the ranking for each expert could be different however once the full correlation matrix of each expert is determined any probabilistic statement may be computed. Then for the variable which they regarded as having the largest rank correlation in absolute value, experts would assess the usual probability of exceedence [Morales et al., 2008]. Next, ratios of each of the remaining rank correlations to the one assessed through a probability of exceedence were asked. This method is described in subsection 4.2.2.2 and relation 4.9.

A convex combination of the densities realized by the BBN quantified with the individual estimates provided by each expert does not preserve the conditional independence statements embedded in the graph. Another strategy has to be considered in order to combine experts' opinions. If all experts assessed conditional probabilities of exceedence based on the same events then these probabilities may be linearly pooled to use as the DMs estimate. When the marginal distributions do not come from data then each expert provides estimates over different events. The strategy to follow is then to compute the probabilities that each expert "*would have stated*" if he/she had been asked probabilistic statements regarding a given quantile of the Decision Maker such that his/her estimates for the rank correlations remain unchanged (see relation 4.11). For a detailed explanation of the procedure for combination of dependence estimates the reader may see section 4.2.3.

From subsection 6.3.2 it could be observed that the EWDM was the recommended choice for combining experts' opinions in the DS model. The combination of the four expert's individual assessments was done as described in section 4.2.3. The quantities combined were the conditional probability of each child node given the corresponding parent. These numbers were later translated into the corresponding rank and conditional rank correlations. The results of the combination scheme are presented in table 6.2. For instance, $r_{3,6}$ stands for the rank correlation between variables $X_3 = \text{maintenance}$ and $X_6 = \text{piping}$ according to the numbering shown in Figure 6.3.

In table 6.3 the determinants of the rank correlation matrices for each expert and the DM are shown. The ratio of largest to smallest determinant is 3.95×10^5 . The EWDM's determinant is the largest among the 5 experts (including the DM itself). The reason is that marginal distributions assessed by experts differ considerably and the EWDM's combination tends to "fade away" the dependence. Example 6.3.1 gives an intuitive explanation of this remark for the case of the rank

(Un)Conditional Rank Correlation	Value	(Un)Conditional Rank Correlation	Value
$r_{3,6}$	0.1799	$r_{7,6}$	0.5025
$r_{2,6 3}$	0.1067	$r_{7,5 6}$	0.5793
$r_{4,3}$	-0.3996	$r_{7,4 5,6}$	-0.4647
$r_{4,2 3}$	-0.3164	$r_{7,3 4,5,6}$	0.2212
$r_{4,1 2,3}$	-0.4307	$r_{8,7}$	0.1135
$r_{5,4}$	-0.1278	$r_{8,5 7}$	0.0669
$r_{5,2 4}$	0.1711	$r_{10,8}$	0.1384
$r_{5,1 2,4}$	0.3025	$r_{9,8}$	0.2281

Table 6.2: EWDM Dependence estimates for the DS Model.

correlation between *Flooding* and *Economic costs*.

Experts' Id	Expert's Determinant
A	4.5703×10^{-7}
B	0.0224
C	6.2629×10^{-4}
D	0.0160
EWDM	0.1806

Table 6.3: Experts' Correlation Matrices Determinant.

Example 6.3.1. Table 6.4 presents a summary of estimates required by experts to compute $r_{9,8}^{e_i}$ in Figure 6.3. Column 1 gives the expert's id. Estimates given by each expert to the question: *Suppose that variable flooding was observed above its median value, what is the probability that also economic costs were observed above their median?* are presented in column 2 of table 6.4. The rank correlation realized by each expert's estimate is shown in column 3.

For both variables *Flooding* and *Economic costs* the EWDM's median realizes a given percentile in each experts' marginal distribution. Similarly to chapter 4, the cumulative distribution function for variable X_j from expert e_i will be denoted as $F_{X_j}^{e_i}$. The median value of variable X_j for expert e_i is denoted as $x_{j,0.5}^{e_i}$. Also, the k^{th} percentile of variable X_j is denoted as $x_{j,0.k}^{e_i}$.

For example the EWDM's median for *Economic costs* is 20.03 million usd. According to the indexing shown in Figure 6.3, *Economic costs* = X_{10} and $F_{X_{10}}^A(20.03) = 0.82$. In other words the DM's median realizes the 82th percentile in expert's A marginal distribution. In the same way the EWDM's median for *Flooding* realizes the 4th percentile in expert A's marginal distribution. Hence column 4 of table 6.4 shows the assessment that each expert "would have stated" if he/she had been asked probabilistic statements regarding the median of the Decision Maker such that his/her estimates for the rank correlations remain unchanged. For other experts the percentile realized by the EWDM's median in

each experts distribution may be read similarly. The relationship between $r_{9,8}$ and $P_{1^*}^{e_i}$ may be seen in Figure 6.4 for all four experts.

Expert	$P_1^{e_i}$	$r_{9,8}^{e_i}$	$P_{1^*}^{e_i}$
A	0.8	0.81	$P(X_9 > x_{9,0.82} X_8 > x_{8,0.04}) = 0.19$
B	0.8	0.81	$P(X_9 > x_{9,0.05} X_8 > x_{8,0.95}) = 0.99$
C	0.6	0.31	$P(X_9 > x_{9,0.16} X_8 > x_{8,0.95}) = 0.95$
D	0.7	0.59	$P(X_9 > x_{9,0.96} X_8 > x_{8,0.03}) = 0.04$
EWDM	-	0.1384	$P(X_9 > x_{9,0.50} X_8 > x_{8,0.50}) = 0.55$

Table 6.4: Combination of rank correlation $r_{9,8}$ in Figure 6.3.

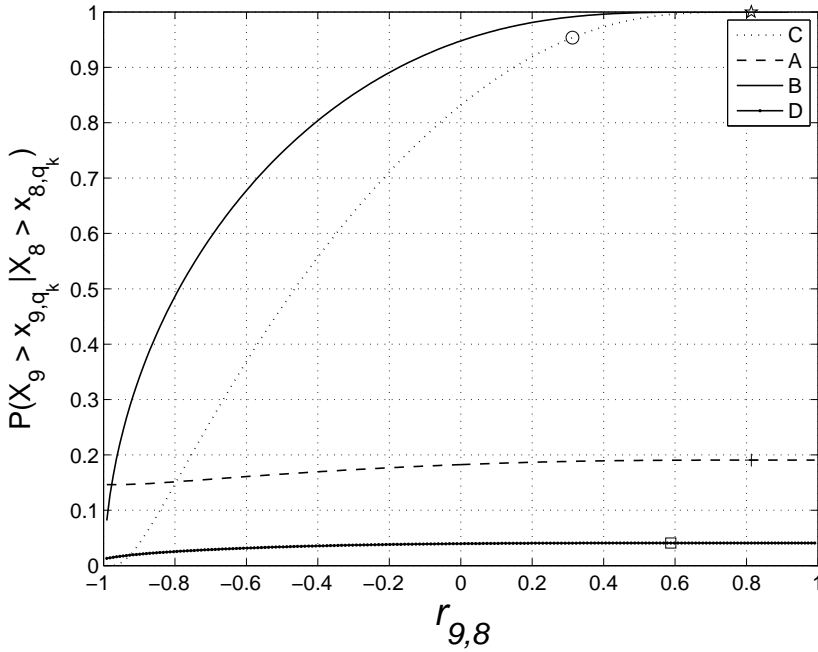


Figure 6.4: Relationship between $r_{9,8}$ and $P_{1^*}^{e_i}$ in table 6.4

The EWDM answer to $P(X_9 > x_{9,0.50} | X_8 > x_{8,0.50}) = \frac{1}{4} \sum_{i=1}^4 P_{1^*}^{e_i}$. It may be observed that $r_{9,8}^{e_i} > 0.31$ for all e_i , however because of the large differences in $P_{1^*}^{e_i}$ the $r_{9,8}^{EWDM} \approx 0.14$. \square

Other estimates in table 6.2 behave similarly to example 6.3.1 and hence the high value of the EWDM's correlation matrix determinant.

6.4 Discussion of the DS Model

One of the objectives of the model is to predict or diagnose the performance of any of the seven Mexican structures under consideration. To limit the explanation the use of the model will only be illustrated here with data from JAA.

Because of its geometry and year of construction there are two variables that can immediately be fixed for the dam under study. Variable *landslide* is a distribution over the security factor of the dams under study (see subsection 6.3.1). The security factor of JJA was calculated base on its geometry according to the so called Swedish method [SRH, 1976]. Hence $landslide = 1.95$. Secondly, the age of the dam is 46 years which can be associated with the number of years between *maintenance* activities assuming that there has not been any conservation actions since its final construction year.

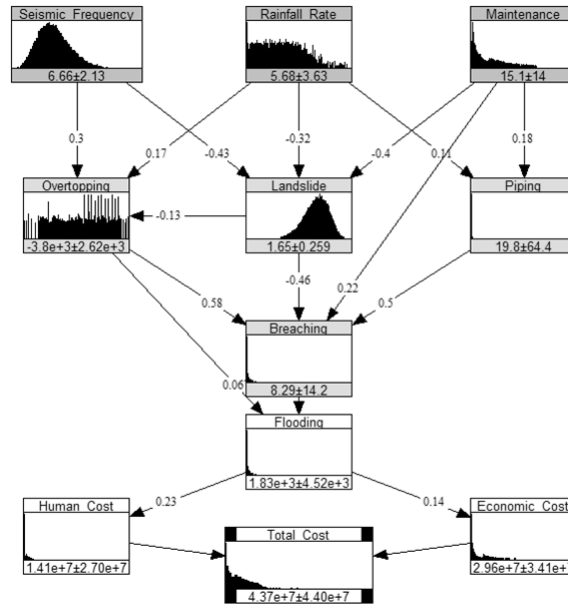


Figure 6.5: Unconditional DS model.

Figure 6.5 shows the model from Figure 6.3 with the marginal distributions from section 6.3.1. Means and standard deviations (after the \pm sign) are shown. Figure 6.6 presents the model adapted to the dam of interest (JAA). The original marginal distributions are shown in Figure 6.6 in grey while the updated belief is shown in black. According to the model, the effect of introducing evidence of $landslide = 1.95$ and $maintenance = 46$ yrs is larger in *overtopping* and *rainfall rate* than in other variables.

Suppose that additionally to the evidence previously entered, an extraordinary rainfall rate of $15 \frac{mm}{day}$ in a 7 day average is observed. Also it is known that

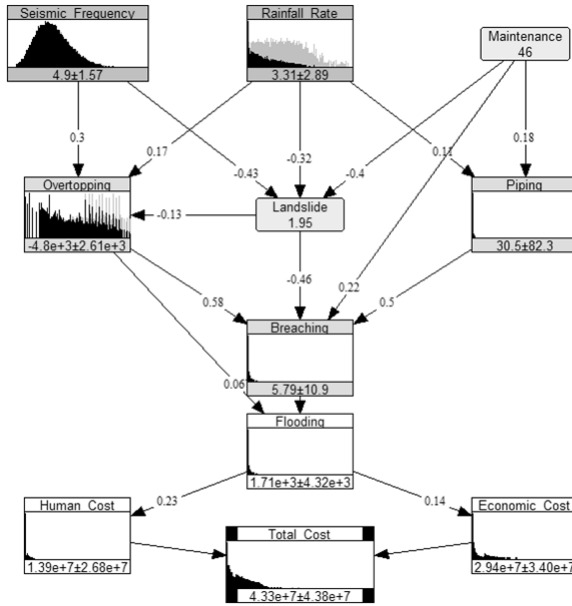


Figure 6.6: DS model given landslide = 1.95 and maintenance = 46.

the seismic frequency in this region corresponds to 8 earthquakes with intensity higher than 5.5 in Richter scale per year. Figure 6.7 shows the results of the entering additional evidence in the model. As can be seen, the anticipated flooding value has increased from $1.71 \times 10^3 \frac{mm}{day}$ (Figure 6.6) to $2.22 \times 10^3 \frac{mm}{day}$ (Figure 6.7). Similarly, the predicted human cost moved from 13.9 to 14.7 million USD, and the economic loss in turn from 29.4 to 30.1 million USD. This means that the intensification of rain and the presence of earthquakes at the same time are expected to produce higher levels of water in the potential flood area and consequently larger amounts of both human and economic losses.

Similar analysis to the one described previously was conducted for all 7 dams under study. The impact of an overtopping incident of 100 mm was employed to analyze its effects not only in the flood water level downstream, but also in human and economic costs. So for each of the seven structures three values were fixed: *landslide* (security factor), *maintenance* (dam age using the assumption above mentioned) and *overtopping* (100 mm).

Results show that given the landslide (security factor) for each dam and no maintenance performed since its construction, an overtopping of 100mm increases the expectation of a flooding by a factor 1.79 (EG) up to a factor 2.11 (SJ) Delgado-Hernández et al. [2009]. However the total costs of such an increase are not as sensitive as a flooding is (4 - 6% increase in expected costs). It may be observed that the expert combination indicate that human costs increase more than economic costs (7-9% compared to 3 or 4%); however the larger contribution of economic costs drives the total cost increase.

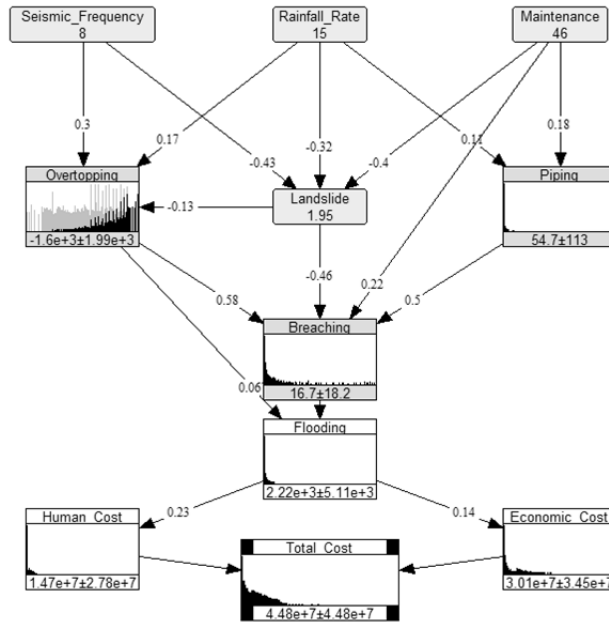


Figure 6.7: DS model given landslide = 1.95, maintenance = 46, seismic frequency = 8 and rainfall rate = 15.

Observe that differences in expected total costs are small, which can be explained by the values in the rank correlation matrix (table 6.5). Human and economic costs are obviously highly correlated with total cost ($r_{9,11} = 0.82$ and $r_{9,11} = 0.48$). Flooding is mainly correlated to human costs ($r_{8,10} = 0.23$), and total and economic costs lag behind them. However the rank correlation between flooding and total costs is still low ($r_{8,11} = 0.22$). All other variables are only weakly correlated with the total consequences. Nevertheless, in Table 6.5, flooding showed bigger variations than total costs. This means that, according to the equal weight combination of the experts' opinions used to build the model, once a dam has failed flooding variations will be more important than those related with total costs.

6.5 Final comments of the DS Model

The DS model could be used in a similar fashion as in section 6.4 to perform analysis. In fact, a wide variety of scenarios could be constructed to determine the level of impact of other particular incidents (such as piping or breaching), or a combination of them in the expected consequences.

This chapter dealt with earth dams and their failure modes, emphasizing risk assessment in a group of seven dams within the State of Mexico. The combination of BBNs and expert judgment stemmed from the recognition that Mexican dam

	X_1^a	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
X_1	1	0.00	0.00	-0.36	0.30	0.00	0.24	0.05	0.01	0.02	0.02
X_2		1	-0.01	-0.28	0.19	0.10	0.22	0.04	0.00	0.01	0.01
X_3			1	-0.41	-0.03	0.18	0.33	0.02	0.00	0.01	0.01
X_4				1	-0.12	-0.11	-0.43	-0.05	0.00	-0.02	-0.01
X_5					1	0.01	0.49	0.11	0.02	0.03	0.03
X_6						1	0.51	0.04	0.01	0.01	0.02
X_7							1	0.12	0.02	0.03	0.03
X_8								1	0.14	0.23	0.22
X_9									1	0.03	0.82
X_{10}										1	0.48
X_{11}											1

Table 6.5: Correlation matrix for the DS model.

^a X_1 = seismic frequency, X_2 = rainfall rate, X_3 = maintenance, X_4 = landslide, X_5 = overtopping, X_6 = piping, X_7 = breaching, X_8 = flooding, X_9 = economic costs, X_{10} = human costs, X_{11} = total costs

managers need simple, useful and practical tools for carrying out quantitative risk assessment, based on a solid theoretical foundation. The tools should also be applicable to the context of their structures. In an effort to fulfill such requirements, a model that considers some of the variables that have influenced dam failures globally in the past has been proposed.

While the key objectives of the study have been achieved, there were a number of limitations associated with the work. First of all, the number of experts was somewhat low mainly because there is a lack of people with the required profile to be considered as such. To find specialists aware of the current situation of the dams under study proved to be a difficult task. In the event six people were identified but only four could take part in the research.

The inclusion of more variables in the model should be considered. This is particularly relevant if some local cases have shown that other variables are important in the risk evaluation apart from those reported in international statistics.

The equal weight combination was proposed as the preferred choice for the decision maker. The choice was motivated mainly because of suboptimal performance of each individual expert. This in turn led to a suboptimal performance of the optimized decision makers. The training of experts in probabilistic assessments is fundamental for the classical method for structure expert judgment. Results from this study suggest that better training or a selection of seed variables that characterizes better the expertise in the expert panel is desired for the follow up of the project.

In spite of these observations, it is strongly believed that the methodology utilized to build the model can be applied to carry out similar exercises in different locations. Overall this research has demonstrated that the use of NPCDBBN in Mexican dams' risk assessment is not only feasible but also beneficial. Finally, it should be emphasized that this research is hoped to be the starting point of a bigger project aimed at developing a more comprehensive model applicable to different types of dams in the country.

CHAPTER 7

Conclusions

7.1 About Vines

This thesis has dealt with applications of graphical models in risk and uncertainty analysis. In particular, non-parametric continuous discrete Bayesian belief nets are used to investigate risks in the aviation system and in earth dams. The theory behind non-parametric continuous discrete Bayesian belief nets was built around vines and for that reason the study of vines is the beginning of this thesis. For the same reason some conclusions about the research presented in this thesis related to vines will be presented first.

Vines have been investigated at least since the mid 1990's. Graphical aspects of vines have been less explored than their applications in simulation, statistics and uncertainty analysis. In this thesis we have provided explicitly for the first time results concerning the number of labeled vines on n nodes and the number of labeled regular vines on n nodes. Algorithms for building both labeled vines and labeled regular vines have been proposed. Though to our knowledge, no applications have been published to this time for non regular vines, it is not immediately clear that these objects will not find application in the future.

The value of obtaining $\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$ as the number of labeled regular vines on n nodes is more clearly recognized in recent applications. Obviously this number grows extremely fast with n . Implementing the statistical techniques proposed in the literature might be restrictive even for a modest value of n . Take for example a data set with $n = 7$. According to table 2.1 there are 2,580,480 labeled regular vines of which one in principle could be the best fit to the data. This number might be too large for a personal computer to perform the job. We could think of restricting our choices to some class of tree-equivalent regular vine in order to alleviate computational burden.

According to tables A.10 and A.11 in appendix A, there are 136 tree-equivalent regular vines on 7 nodes. These are V33 (D-vine on 7 nodes) to V168 (C-vine on 7

nodes). From tables A.10 and A.11 we see that if we would wish to fit only C-vines or D-vines to our data set of 7 variables, the choices reduce to 2,520 possibilities for either one. From the same tables we can see that besides D-vines and C-vines 9 other tree-equivalent regular vines admit also 2,520 labeled regular vines. In other words, there are 11 tree equivalent regular vines that can be labeled in 2,520 different ways. These are V33 (D-vine on 7 nodes), V45, V48, V52, V86, V147, V151, V154, V164, V167 and V168 (C-vine on 7 nodes) in tables A.10 and A.11.

Similarly, from tables A.10 and A.11, we may see that there are 24 tree-equivalent regular vines that can be labeled in 5,040 different ways each. These are V34, V35, V36, V38, V42, V49, V50, V53, V57, V61, V74, V75, V76, V78, V81, V103, V107, V110, V125, V148, V149, V155, V159 and V165. If we continue in this way a distribution of tree-equivalent regular vines on 7 nodes according to the number of admissible labellings may be obtained. The resulting distribution is presented in Figure 7.1.

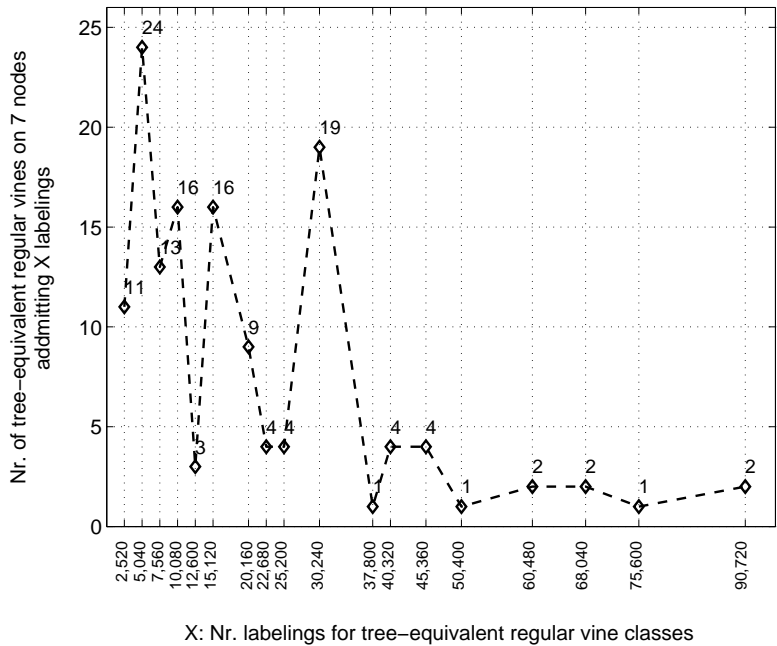


Figure 7.1: *Distribution of tree-equivalent regular vines on 7 nodes according to the number of labeled regular vines admissible.*

Of course the summation of the values observed in the vertical axis in Figure 7.1 is 136 which is the total number of tree-equivalent regular vines. As stated before, C-vines and D-vines are included in the 11 tree-equivalent regular vines that admit 2,520 labeled versions. There are however 134 other tree-equivalent regular vines of which 9 admit also 2,520 labeled regular vines. Observe that if we would choose any of V117 or V139 in in table A.11, there are 90,720 possible

labeled regular vines for each, out of which one could be selected as the best fit to the data. The choice of a subset of the 136 tree-equivalent regular vines to fit data to it is not immediately evident.

In this thesis tree-equivalent vines were studied. For $n \geq 5$ there are more equivalence classes than tree-equivalent regular vines (see Joe [2010], Morales-Nápoles [2010] and chapter 2). Equivalence classes of regular vines have recently been characterized and a formula for dimension $n \geq 5$ is presented in Joe et al. [2010]. Whether using tree-equivalent or equivalence classes of regular vines for statistical manipulation is also not entirely clear. In any case, this example shows the need to think of using tree-equivalent regular vines or equivalence classes of regular vines with criteria different than just their popularity.

In this thesis we have made a first step towards organizing vines and regular vines in a more systematic way. We believe that this task is necessary in order to progress more rapidly the space of applications for vines and make them more accessible for people interested in the subject. Hence our recommendation is to enhance efforts for a more systematic organization of vines including algorithms for generating and storing them.

7.2 About Bayesian Networks and their Applications

7.2.1 Aviation Safety

The largest part of this thesis is concerned with the application of non-parametric continuous-discrete Bayesian belief nets in aviation safety. A smaller application is also presented for earth dams safety in the State of Mexico. We begin first by discussing some conclusions about the applications presented. Then we turn our attention to conclusions relative to Bayesian networks and elicitation of dependence measures.

The Dutch ministry of transport, through the commission of a project of the magnitude of CATS, has shown the importance that safety in the aviation sector has for policy makers in the Netherlands. The CATS model can be a powerful tool for risk and uncertainty analysts in their recommendations for policy makers. One of the fundamental parts of the CATS model is the use of human reliability models. The flight crew performance, air traffic control performance and maintenance technician performance models are presented in chapter 3. Techniques for the elicitation of rank and conditional rank correlations required for these models are presented in chapter 4 and results from the actual elicitation which constitute the basis of the models' quantification are presented in chapter 5.

The techniques described in these chapters result in a large scale BBN with 1,504 nodes and 4,979 arcs. This can be readily used for risk and uncertainty analyzes. Out of the 1,504 nodes included in the model, 45 represent the 3 human reliability models introduced in this thesis. These take account of all the dependence in the model. The current version of CATS used in section 3.4 shows that the variables in the flight crew and the maintenance technician models are more highly correlated with accident probability than the variables in the air traffic controllers model. From the 20 most highly correlated variables of the three

human reliability models 16 correspond to flight crew performance, 3 to maintenance technician performance and 1 (aircraft generation) is shared by the two models. These top 20 rank correlations range from roughly 0.1 to 0.3 in absolute value. At first sight they might appear to be low values for rank correlations, however their effect on accident probability can be very large.

Take for example Figure 7.2 where the 5th, 95th and mean value of the accident distributions shown in Figure 3.17 are presented. Observe that the difference between the 5th and 95th percentiles in the 3 cases span roughly 2 orders of magnitude, hence the uncertainty over the 3 central estimates shown in Figure 7.2 is comparable. The first conditional distribution shows that the expectation of accident probability given the oldest kind of aircrafts is larger than the 95th percentile of the base line case. The expectation of the accident probability in this case would be of 5 in a 100,000 flights. For the third conditional distribution (with the additional condition of an unexperienced crew) the expectation is again larger than the 95th percentile of the accident probability distribution given old aircrafts. In this case we could expect about 3 accidents in 10,000 flights.

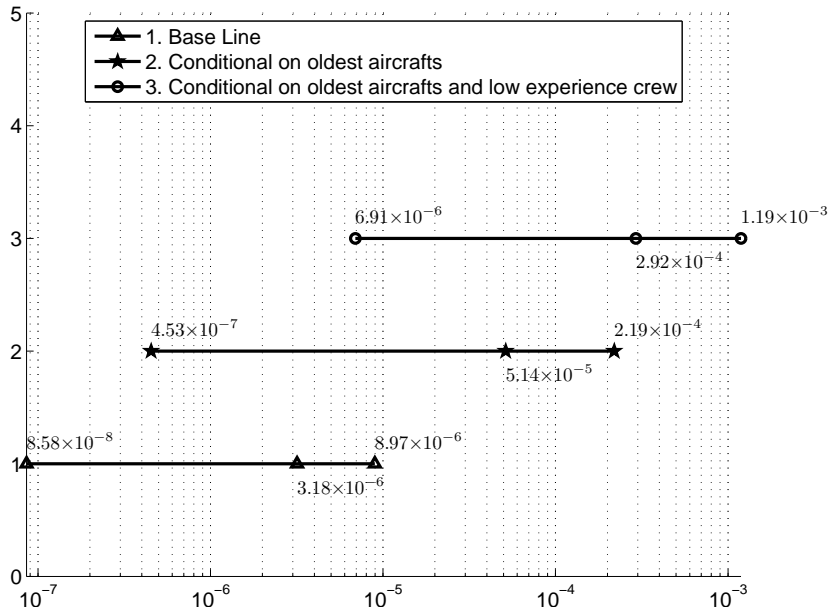


Figure 7.2: 5%-tile, mean and 95%-tile of the accident (fatal and non-fatal) distribution from the CATS model. 1. Baseline; 2. Given aircraft generation = 1; 3. Given aircraft generation = 1, captain experience = 9,467 hr. and first officer experience = 7,844 hr.

The CATS model, which represents our pool of experts opinion, says that for policy makers it would be of utmost importance to revise the number of flights currently operating where unexperienced crew is flying the oldest kind of aircrafts. Experienced pilots or new technology do not come in cheap. We can speculate that these two risky conditions meet more often in world regions where the availability of experienced pilots and new technology is scarce. Perhaps, if the aviation system ought to be more safe, investments trying to correct this difference across regions in the world should be considered.

7.2.2 Earth Dams Safety

Conclusions regarding the model for measuring earth dams risks have already been discussed in chapter 6. The most important ones are briefly repeated next: given the geometry of each of the dams under study and the assumption that no maintenance is performed since its construction, an overtopping of 10 *cm* increases the expectation of a flooding by a factor 1.79 in El Guarda up to a factor 2.11 in San Joaquín. However, the total costs of such an overtopping are not as sensitive as a flooding is (4 - 6% increase in expected costs). According to the equal weight combination of the experts' opinions used to build the model, once a dam has failed, variations in the total costs will be minimal. In other words, according to this combination of expert opinions if one or another dam under study fails with an overtopping of 10 *cm* the consequences would end up in approximately the same total costs.

This result is not strange given the fact that the 7 dams selected for the demonstration model share similar characteristics. One of the objectives of this model is to make it the basis for a larger model for investigating risks in larger dams all over the country and not only in the State of Mexico. Such a model may built significantly in the one presented in chapter 6.

7.2.3 About BBNs.

Bayesian networks have proved in this thesis to be a powerful tool for risk and uncertainty analysis. In particular the vine-copula approach in which non-parametric continuous-discrete BBNs relies require the use of rank and conditional rank correlations. In this thesis methods for the quantification of these dependence measures from experts have been proposed. Moreover, these methods have been used for the quantification of rank correlations for inducing dependence in a large scale model for air transport safety. The same kind of techniques have been used for a smaller model for earth dams risks. The techniques proposed in chapter 4 are flexible enough as to allow for some differences in the elicitation of such dependence measures.

We have shown that one of the advantages of the vine-copula approach to BBNs vs. discrete BBNs is that it makes them more flexible with respect to modelling changes. For example when nodes are added or removed less parameters might need to be re-elicited from experts. This however does not mean that in general a re-elicitation would not be necessary.

One observation that calls the attention is the fact that when marginal distributions are very different across experts, the joint distribution obtained with the method for combination described in chapter 4 of this thesis with equal weights tends to ‘fade away’ the magnitude of the dependence even if individual experts think bivariate rank correlations are of the same sign and magnitude. A similar situation could be observed in a combination within the class of performance weight combinations. More research in this direction is advised.

References

- K. Aas and D. Berg. Models for construction of multivariate dependence. *Accepted for publication in European Journal of Finance*, 2009.
- K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182 – 198, 2009.
- B. Ale, L. Bellamy, R. Cooke, L. Goossens, A. Hale, A. Roelen, and E. Smith. Towards a causal model for air transport safety an ongoing research project. *Safety Science*, 44(8):657–673, 2006.
- B. Ale, L. Bellamy, R. d. Boom, J. Cooper, R. Cooke, L. Goossens, A. Hale, D. Kurowicka, O. Morales, A. Roelen, and J. Spouge. Further developments of a causal model for air transport safety (cats); building the mathematical heart. In *ESREL.*, pages 1431–1439, 2007.
- A. Aparicio, P. Martnez-Austria, A. Gitrn, and A. Ramrez. Floods in tabasco, mexico: a diagnosis and proposal for courses of action. *Journal of Flood Risk Management*, 2009.
- G. A. Barnard and T. Bayes. Studies in the history of probability and statistics: Ix. thomas bayes’s essay towards solving a problem in the doctrine of chances. *Biometrika*, 45(3/4):293–315, 1958. ISSN 00063444. URL <http://www.jstor.org/stable/2333180>.
- M. Bartlett. Contingency table interactions. *Journal of the Royal Statistical Society Supplement*, (2):248–252, 1935.
- T. Bedford and R. Cooke. Vines - a new graphical model for dependent random variables. *Ann. of Stat.*, 30(4):1031–1068, 2002.

- L. Beineke. Derived graphs with derived complements. In *Recent Trends in Graph Theory: Proceedings of the First New York City Graph Theory Conference held on June 11, 12, and 13, 1970*. Springer, 2006.
- D. Bellhouse. The reverend thomas bayes, frs: A biography to celebrate the tercenary of his birth. *Statistical Science*, 19(1):3–43, 2004.
- N. Biggs, E. K. Lloyd, and R. J. Wilson. *Graph Theory: 1736-1936*. Clarendon Press, New York, NY, USA, 1986. ISBN 0-198-53916-9.
- A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla. Comparing fault trees and bayesian networks for dependability anlysis. In M. Felici, K. Kanoun, and A. Pasquini, editors, *SAFECOMP'99, LNCS 1698*, pages 310–322, 1999.
- A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla. Improving the analysis of dependable systems by mapping fault trees into bayesian networks. *Reliability Engineering & System Safety*, 71:249–260, 2001.
- C. Boutlier. The influence of influence diagrams on artificial intelligence. *Decision Analysis*, 2(4):229–232, 2005.
- CAANL. Veiligheidsstatistieken burgerluchtvaart (civil aviation safety data) 1993-2007. Brochure, P.O. Box 90653, 2509 LR The Hague Netherlands, 2008.
- J. Carrera-Hernández and S. Gaskin. The basin of mexico hydrogeological database (bmhdb): Implementation, queries and interaction with open source software. *Environmental Modelling & Software*, 23(11-10):1271–1279, 2008.
- A. Cayley. A theorem on trees. *The Quarterly Journal of Pure and Applied Mathematics*, 23:376–378, 1889.
- L. Chollete, A. Heinen, and A. Valdesogo. Modeling international financial returns with a multivariate regime-switching copula. Technical Report 4, Fall 2009. URL <http://ideas.repec.org/a/oup/jfinec/v7y2009i4p437-480.html>.
- C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on*, 14(3):462–467, 1968.
- G. Clemen and et al. Correlations and copulas for decision and risk analysis. *Management Science*, 45:208–224, 1999.
- G. Clemen and et al. Assesing dependencies: Some experimental results. *Management Science 2000 Informs*, 46(8):1100–1115, August 2000.
- R. Cooke. *Experts in uncertainty*. Oxford University Press, 1991.
- R. Cooke. Markov and entropy properties of tree and vine-dependent variables. In *Proceedings of the ASA Section on Bayesian Statistical Science*,, 1997.

- R. Cooke and L. Goossens. Procedures guide for structured expert judgment. Technical Report EUR18820, European Commission: Nuclear Science and Technology, Brussels-Luxemburg, July 1999.
- R. Cooke and L. Goossens. Tu delft expert judgment data base. *Reliability Engineering & System Safety*, 93:657–674, 2008.
- R. Cooke, D. Kurowicka, A. Hanea, O. Morales, B. Ababei, D.A. Ale, and A. Roelen. Continuous/discrete non parametric bayesian belief nets with unicorn and uninet. In T. Bedford, J. Quigley, L. Walls, and A. Babakalli, editors, *Proceedings of the Mathematical Methods for Reliability conference*, 2007.
- R. Cowell, A. Dawid, S. Lauritzen, and S. D.J. *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer, 1999.
- A. I. Dale. *A history of inverse probability : from Thomas Bayes to Karl Pearson*. Sources and studies in the history of mathematics and physical sciences. Springer, 2nd ed. edition, 1999.
- J. N. Darroch, S. L. Lauritzen, and T. P. Speed. Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, 8(3):522–539, 1980. ISSN 00905364. URL <http://www.jstor.org/stable/2240590>.
- F. N. David. Studies in the history of probability and statistics i. dicing and gaming (a note on the history of probability). *Biometrika*, 42(1/2):1–15, 1955. ISSN 00063444. URL <http://www.jstor.org/stable/2333419>.
- D. Delgado-Hernández, O. Morales-Nápoles, D. De-León-Escobedo, J. Rivero-Santana, D. Prez-Flores, and B. Pérez-Pliego. A model for earth dams’ risk assessment. *submitted to Journal of Geotechnical and Geoenvironmental Engineering*, 2009.
- R. Donnelly. Safe and secure: risk-based techniques for dam safety. International Water Power and Dam Construction, May 2006. <http://www.waterpowermagazine.com/story.asp?storyCode=2040340>.
- M. Emiroglu, A. Tuna, and A. Aislan. Development of an expert system for selection of dam type on alluvium foundations. *Engineering with Computers*, 18(1):24–37, 2002.
- FAA. Faa aerospace forecasts fy 2009-2025. Brochure, 800 Independence Avenue, SW Washington, DC 20591, 2009.
- J. Fauvel and P. Gerdes. African slave and calculating prodigy: Bicentenary of the death of thomas fuller. *Historia Mathematica*, (17):141–151, 1990.
- FEMA. Federal guidelines for dam safety: Glossary of terms. Glossary of Terms FEMA 148, Federal Emergency Management Agency (FEMA), April 2004.
- FEMA. The national dam safety program final report on coordination and cooperation with the european union on embankment failure analysis. REPORT FEMA 602, Federal Emergency Management Agency (FEMA), August 2007.

- FEMA. Risk prioritization tool for dams users manual. Manual FEMA P-713CD, Federal Emergency Management Agency (FEMA), March 2008.
- M. Foster, R. Fell, and M. . Spannagle. The statistics of embankment dam failures and accidents. *Canadian Geotechnical Journal*, 37(5):1000–1024, 2000.
- M. Frank. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19:194–226, 1979.
- A. Hanea. *Algorithms for Non-parametric Bayesian belief nets*. PhD thesis, TU Delft, Delft, the Netherlands, 2008.
- A. Hanea, D. Kurowicka, and R. Cooke. Hybrid method for quantifying and analyzing bayesian belief nets. *Quality and reliability Engineering International*, 22:709–729, 2006.
- F. Harary. Some theorems and concepts of graph theory. In F. Harary, editor, *A Seminar on Graph Theory*, pages 1–12, 1967.
- F. Harary. *Graph Theory*. Addison-Wesley, 1969.
- F. Harary and E. Palmer. *Graphical Enumeration*. Academic Press, 1973.
- R. Howard and J. Matheson. Influence diagrams (reprinted). *Decision Analysis*, 2(3):229–232, 1984/2005.
- ICOLD. Dam failure statistical analysis. Bulletin 99, 1995.
- I. Jagielska. Quantification of non-parametric continuous bbns with expert judgment. Master’s thesis, Delft University of Technology, The Netherlands, July 2007.
- H. Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141, 1996. ISSN 07492170. URL <http://www.jstor.org/stable/4355888>.
- H. Joe. Dependence comparisons of vine copulae in four or more variables. In D. Kurowicka and H. Joe, editors, *Dependence Modeling-Handbook on Vine Copulae*, Dependence Modeling, Scheduled Fall 2010.
- H. Joe, R. M. Cooke, and D. Kurowicka. Regular vines: Generation algorithm and number of equivalent classes. In D. Kurowicka and H. Joe, editors, *Dependence Modeling-Handbook on Vine Copulae*, Dependence Modeling, Scheduled Fall 2010.
- V. Kasyanov and V. Evstigneev. *Graph Theory for Programmers-Algorithms for Processing Trees*. Kluwer Academic Publishers, 2000.
- M. G. Kendall. Studies in the history of probability and statistics: Ii. the beginnings of a probability calculus. *Biometrika*, 43(1/2):1–14, 1956. ISSN 00063444. URL <http://www.jstor.org/stable/2333573>.

- J. H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference systems. 1983.
- O. Kolbjornsen and M. Stien. The d-vine creation of non-gaussian random fields. In *GEOSTATS*, 2008.
- B. Kraan. *Probabilistic Inversion in Uncertainty Analysis and Related Topics*. PhD thesis, Delft University of Technology, 2002.
- K. Krugła. Aviation risks with continuous/discrete non parametric bbn. Master's thesis, Delft University of Technology, The Netherlands, July 2008.
- D. Kurowicka and R. Cooke. Distribution-free continuous bayesian belief nets. In K.-M. S. Wilson A., Limnios N. and A. Y., editors, *Modern Statistical and mathematical Methods in Reliability*, pages 309–323, 2005.
- D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, 2006.
- S. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, 1996.
- S. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):157–224, 1988. ISSN 00359246. URL <http://www.jstor.org/stable/2345762>.
- D. Lewandowski. Generalized diagonal band copulas. *Insurance: Mathematics and Economics*, 37(1):49 – 67, 2005. ISSN 0167-6687. doi: DOI: 10.1016/j.insmatheco.2004.12.006. Papers presented at the DeMoSTAFI Conference, Quebec, 20-22 May 2004.
- D. Lewandowski, R. M. Cooke, and R. J. D. Tebbens. Sample-based estimation of correlation ratio with polynomial approximation. *ACM Trans. Model. Comput. Simul.*, 18(1):1–17, 2007. ISSN 1049-3301. doi: <http://doi.acm.org/10.1145/1315575.1315578>.
- W. Mayeda and S. Seshu. Generation of trees without duplication. *IEEE Transactions on Circuit Theory*, 12:181–185, 1967.
- A. Meeuwissen. *Dependent Random Variables in Uncertainty Analysis*. PhD thesis, Delft University of Technology, 1993.
- A. Meeuwissen and R. Cooke. Tree dependent random variables. Technical Report 94-28, Delft University of Technology, Dept. Mathematics, 1994.
- A. Min and C. Czado. Bayesian inference for multivariate copulas using pair copula constructions. *Submitted for publication*, 2008.
- G. Minty. A simple algorithm for listing all the trees of a graph. *IEEE Transactions on Circuit Theory*, 12:120– 120, 1965.

- J. Moon. Various proofs of cayley's formula for counting trees. In F. Harary, editor, *A Seminar on Graph Theory*, pages 70–78, 1967.
- O. Morales, D. Kurowicka, and A. Roelen. Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, 93(5):699 – 710, 2008. ISSN 0951-8320. doi: DOI: 10.1016/j.res.2007.03.020. Expert Judgement.
- O. Morales-Nápoles. Counting vines. In D. Kurowicka and H. Joe, editors, *Dependence Modeling-Handbook on Vine Copulae*, Dependence Modeling, Scheduled Fall 2010.
- O. Morales-Nápoles and D. Delgado-Hernández. Quantification of a model for earth dams' risk assessment in the state of mexico. Article, TU Delft-Universidad Autónoma del Estado de México, 2009. (in preparation).
- O. Morales-Nápoles, D. Kurowicka, R. Cooke, and D. Ababei. Continuous-discrete distribution free bayesian belief nets in aviation safety with UNINET. Technical report, TU Delft, 2007.
- O. Morales-Nápoles, R. Cooke, and D. Kurowicka. Eemcs final report for the causal modelling for air transport safety (cats) project. Report, EEMCS-TU Delft, Delft Institute of Applied Mathematics, July 2008.
- O. Morales-Nápoles, R. Cooke, and D. Kurowicka. About the number of vines and regular vines on n nodes. *Submitted to Applied Discrete Mathematics*, 2009a.
- O. Morales-Nápoles, D. Kurowicka, R. Cooke, and G. van Baren. Expert elicitation methods of rank and conditional rank correlations: An example with human reliability models in the aviation industry. *Submitted to RE&SS*, 2009b.
- H. Neils. Correlation, causation and wright's theory of "path coefficients". *Genetics*, 7:258273, 1922.
- R. B. Nelsen. *An Introduction to Copulas (Lecture Notes in Statistics)*. Springer, October 1998. ISBN 0387986235.
- Next-Page-Software. UNISENS sensitivity analysis documentation., 2009. doi: www.nextpagesoft.net.
- A. O'Hagan. Research in elicitation. In U. Singh and D. Dey, editors, *Bayesian Statistics and its applications*, pages 375–382, Anamaya, New Delhi, 2005.
- J. Pearl. Reverend bayes on inference engines: a distributed hierarchical approach. 1982.
- J. Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. 1985.
- J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

- J. Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann, September 1988.
- J. Pearl. Belief networks revisited. *Artificial Intelligence*, 59:49–56, 1993.
- J. Pearl. Influence diagrams- historical and personal perspectives. *Decision Analysis*, 2(4):232–234, 2005.
- G. Prins. *On the automorphism group of a tree*. PhD thesis, University of Michigan, 1957.
- V. H. Prüfer. Neuer beweis eines satzes über permutationen. *Arch. Math. Phys.*, (27):742–744, 1918.
- R. Read and R. Tarjan. Bounds on backtrack algorithms for listing cycles, paths, and spanning trees. *Networks*, 5:678–692, 1975.
- R. C. Read and R. J. Wilson. *An Atlas of Graphs (Mathematics)*. Oxford University Press, 2005. ISBN 0198526504.
- R. Robinson. *Counting Unlabeled Acyclic Digraphs*. In *Combinatorial Mathematics V*. Lecture Notes in Mathematics (Mathematics and Statistics). Springer Berlin / Heidelberg, November 1977. ISBN 978-3-540-08524-9. DOI 10.1007/BFb0069178.
- A. Roelen, R. Wever, R. Cooke, R. Lopuhaä, A. Hale, and L. Goossens. Causal modelling of air safety. demonstration model. Technical Report NLR-CR-2002-662, National Aerospace Laboratory, December 2002.
- A. Roelen, G. van Baren, J. Smeltink, P. Lin, and O. Morales. A generic flight crew performance model for application in a causal model of air transport. Technical Report NLR-CR-2007-562, Nationaal Lucht- en Ruimtevaartlaboratorium (National Aerospace Laboratory NLR), 2007.
- A. Roelen, B. van Doorn, J. Smeltink, M. Verbeek, and R. Wever. Quantification of event sequence diagrams for a causal risk model of commercial air transport. Report NLR-CR-2006-520, Nationaal Lucht- en Ruimtevaartlaboratorium National Aerospace Laboratory NLR., 2007.
- A. Roelen, G. van Baren, P. Lin, O. Morales, R. Cooke, and D. Kurowicka. A generic air traffic controller performance model for application in a causal model of air transport. Technical Report NLR-CR-2007-593, Nationaal Lucht- en Ruimtevaartlaboratorium (National Aerospace Laboratory NLR), 2008a.
- A. Roelen, G. van Baren, O. Morales, and K. Krugla. A generic maintenance technician performance model for application in causal model of air transport. Technical Report NLR-CR-2008-445, Nationaal Lucht- en Ruimtevaartlaboratorium (National Aerospace Laboratory NLR), 2008b.
- R. Schachter and C. Kenley. Gaussian influence diagrams. *Managment Science*, 35(5):527–550, 1989.

- E. Sheinerman. *Matgraph: A toolbox for graph theory*. Johns Hopkins University, 2009.
- O. B. Sheynin. Studies in the history of probability and statistics. xxi.: On the early history of the law of large numbers. *Biometrika*, 55(3):459–467, 1968. ISSN 00063444. URL <http://www.jstor.org/stable/2334251>.
- A. Shioura, A. Tamura, and T. Uno. An optimal algorithm for scanning all spanning trees of undirected graphs. *SIAM Journal on Computing*, 26:678–692, 1994.
- G. Singuran. System level risk analysis of new merging and spacing protocols. Master’s thesis, Delft University of Technology, The Netherlands, July 2008.
- M. Smith. Generating spanning trees. Master’s thesis, University of Victoria, 1997.
- S. B. Smith. *The Great Mental Calculators. The Psychology, Methods, and Lives of Calculating Prodigies, Past and Present*. Columbia University Press, 1983.
- T. P. Speed and H. T. Kiiveri. Gaussian markov distributions over finite graphs. *The Annals of Statistics*, 14(1):138–150, 1986. ISSN 00905364. URL <http://www.jstor.org/stable/2241271>.
- J. Spouge and G. Vernon. Fault tree modelling for the causal model of air transport safety- final report. Report DNV PROJECT NO. C21004587/3, DET NORSKE VERITAS., REVISION 1 - 28 JULY 2008.
- SRH. *Dams built in Mexico (In Spanish: Presas Construidas en Mexico)*. Mxico, 1976.
- S. M. Stigler. Who discovered bayes’s theorem? *The American Statistician*, 37(4):290–296, 1983. ISSN 00031305. URL <http://www.jstor.org/stable/2682766>.
- Y. Tong. *The multivariate Normal Distribution*. Series in Statistics. Springer, 1990.
- W. Vesely, F. Goldberg, N. Roberts, and D. Haasl. Fault tree handbook. Technical Report NUREG-0492, U.S. Nuclear regulatory Commission, 1981.
- T. Wahl. Prediction of embankment dam breach parameters. Report DSO-98-004, Dam Safety Office, US, 1998.
- J. Whittaker. *Graphical Models in Applied Multivariate Statistics (Wiley Series in Probability & Statistics)*. John Wiley & Sons, March 1990. ISBN 0471917508.
- S. Wright. Correlation and causation. *Journal of Agricultural Research*, XX(7): 557–585, 1921.
- G. Yule and M. Kendall. *An introduction to the theory of statistics*. Charles Griffin & Co., Belmont, California., 14th edition, 1965.

APPENDIX A

Regular Vines Catalogue.

Lets go now from the zoo of reality to the zoo of mythologies, the garden whose fauna is not of lions but of sphinxes, griffins and centaurs. The population of the second garden should exceed that of the first; since a monster is no other thing than a combination of elements of real beings and the possibilities of the combinatorial art border with the infinite.

Manual de zoología fantástica
J.L. BORGES

Catalogues of trees on at most twelve nodes have been presented before. In Moon [1967] pictures for trees with at most five nodes are presented. In Kasyanov and Evstigneev [2000] a catalogue of trees with at most 8 nodes may be found¹. Harary [1969] presents trees on at most 10 nodes². The 987 trees on at most 12 vertices (together with about 10,000 other graphs and many tables of interest for graph theorists) may be found in Read and Wilson [2005].

Tables A.1 to A.4 presents the 48 trees on 8 nodes or less. These trees will be used to present the tree sequences of tree-equivalent regular vines on at most 8 nodes. The purpose of this catalogue is to classify regular vines according to their graphical structure. we hope that this catalogue will help researchers interested in regular vines with their investigations. Like the authors of [Read and Wilson, 2005] this author has “tried that the data is free of errors, but accept[s] no responsibility for any loss of time, money, patience or temper occurring as a result of any mistakes that may have crept into the pages of this [catalogue].

¹This catalogue repeats a tree in eight nodes neglecting another one. In the same reference tables with the number of non-isomorphic trees on less than 26 nodes may be found.

²Harary refers to Prins [1957] for diagrams of trees with at most 12 nodes. However this reference is not available to the author at the moment of the publication of this catalogue.

Furthermore, [the author] wishes it to be understood that any mistakes are entirely the fault of the other author.”

Vines will be presented by pictures in next section and the names of the trees from table A.1 and A.2 used in each level of each regular vine in tables A.8 to A.11 will be displayed in order after the + sign. There is one tree-equivalent regular vine on 3 nodes $V3 = T3 + T2 + T1$. Every regular vine on n nodes for $n > 3$ must necessarily use $V3$ in its construction. For this reason $T3 + T2 + T1$ will be omitted when indicating the sequence of trees used in the construction of different tree-equivalent regular vines. For example the D-vine on 4 nodes will be $V4 = T4 + V3 = T4$. Next the catalogue is presented.



















Prüfer code example			1	12	11	123
						
	T1	T2	T3	T4	T5	T6
# Labeled Trees	1	1	3	12	4	60
# Regular Vines per labeled tree	1	1	1	1	3	1
# Tree-Equivalent Reg. Vines / tree	1	1	1	1	1	1
Prüfer code example	112	111	1234	1123	1213	2244
						
	T7	T8	T9	T10	T11	T12
# Labeled Trees	60	5	360	360	360	90
# Regular Vines per labeled tree	5	24	1	7	11	48
# Tree-Equivalent Reg. Vines / tree	2	2	1	3	3	5
Prüfer code example	1112	1111	12345	12344	12234	12324
						
	T13	T14	T15	T16	T17	T18
# Labeled Trees	120	6	2,520	2,520	5,040	840
# Regular Vines per labeled tree	75	480	1	9	19	33
# Tree-Equivalent Reg. Vines / tree	5	5	1	4	7	3

Table A.1: Trees with at most 7 nodes.








Prüfer code example	11233	11223	11123	12223
				
	T19	T20	T21	T22
# Labeled Trees	630	2,520	840	1,260
# Regular Vines per labeled tree	80	168	168	342
# Tree-Equivalent Reg. Vines / tree	9	17	12	17
Prüfer code example	11122	11112	11111	
				
	T23	T24	T25	
# Labeled Trees	420	210	7	
# Regular Vines per labeled tree	1,452	2,928	23,040	
# Tree-Equivalent Reg. Vines / tree	22	22	22	

Table A.2: *Trees with at most 7 nodes (Continuation).*













Prüfer code example	123456  T26	123455  T27	122345  T28	123345  T29	123435  T30	112324  T31
# Labeled Trees	20,160	20,160	40,320	20,160	20,160	10,080
# Regular Vines per labeled tree	1	11	29	39	71	820
# Tree-Equivalent Reg. Vines / tree	1	5	12	8	10	44
Prüfer code example	112344  T32	122344  T33	122334  T34	123344  T35	112233  T36	122324  T37
# Labeled Trees	5,040	20,160	20,160	20,160	5,040	6,720
# Regular Vines per labeled tree	120	315	815	423	4,520	2,181
# Tree-Equivalent Reg. Vines / tree	14	38	55	41	72	44

Table A.3: Trees with at most 8 nodes.




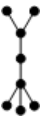
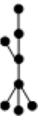






Prüfer code example	244466	123444	123334	112333	122333	111222
						
	T38	T39	T40	T41	T42	T43
# Labeled Trees	10,080	6,720	20,160	3,360	6,720	560
# Regular Vines per labeled tree	11,246	315	1,046	3,384	8,667	89,712
# Tree-Equivalent Reg. Vines / tree	114	24	61	72	111	133
Prüfer code example	122223	123333	112222	122222	222222	
						
	T44	T45	T46	T47	T48	
# Labeled Trees	3,360	1,680	840	336	8	
# Regular Vines per labeled tree	27,222	11,160	117,072	279,000	2,580,480	
# Tree-Equivalent Reg. Vines / tree	114	83	136	136	136	

Table A.4: Trees with at most 8 nodes (Continuation).



















Prüfer code example	2345678  T49	2345578  T50	2345668  T51	2345677  T52	2345658  T53	2345478  T54
# Labeled Trees	181,440	362,880	362,880	181,440	181,440	181,440
# Regular Vines on each tree	1	69	41	13	129	181
# Tree-Equivalent Reg. Vines / tree	1	21	18	6	22	18
Prüfer code example	2345477  T55	2335658  T56	2343677  T57	2335668  T58	2344668  T59	2245677  T60
# Labeled Trees	181,440	181,440	90,720	181,440	362,880	45,360
# Regular Vines on each tree	2,651	5,390	1,708	1,646	2,708	168
# Tree-Equivalent Reg. Vines / tree	164	203	104	125	221	20
Prüfer code example	2335677  T61	2344677  T62	2345577  T63	2344478  T64	2345558  T65	2345666  T66
# Labeled Trees	181,440	181,440	181,440	90,720	181,440	60,480
# Regular Vines on each tree	528	887	887	4,202	2,567	528
# Tree-Equivalent Reg. Vines / tree	70	105	91	147	162	42

Table A.5: Trees with 9 nodes.

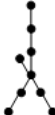

















Prüfer code example	2345448	2343638	2245577	2335577	2245477	2344438
						
	T67	T68	T69	T70	T71	T72
# Labeled Trees	181,440	15,120	90,720	181,440	45,360	30,240
# Regular Vines on each tree	8,738	18,504	11,296	34,417	36,892	72,546
# Tree-Equivalent Reg. Vines / tree	275	99	287	628	350	428
Prüfer code example	2343377	2225668	2333668	2344666	2225677	2333677
						
	T73	T74	T75	T76	T77	T78
# Labeled Trees	90,720	60,480	181,440	60,480	30,240	90,720
# Regular Vines on each tree	120,444	20,904	99,028	34,143	6,756	32,812
# Tree-Equivalent Reg. Vines / tree	724	332	840	439	166	516
Prüfer code example	2344477	2345555	2344448	2333637	2244666	2244477
						
	T79	T80	T81	T82	T83	T84
# Labeled Trees	90,720	15,120	60,480	30,240	30,240	22,680
# Regular Vines on each tree	54,004	32,688	149,901	360,084	428,388	680,576
# Tree-Equivalent Reg. Vines / tree	607	245	765	724	980	1,034

Table A.6: Trees with 9 nodes (Continuation).












Prüfer code example	2225666  T85	2333666  T86	2245555  T87	2333377  T88
# Labeled Trees	5,040	30,240	7,560	30,240
# Regular Vines on each tree	262,080	1,232,820	414,432	1,919,610
# Tree-Equivalent Reg. Vines / tree	465	1,328	735	1,328
Prüfer code example	2335555  T89	2344444  T90	2333338  T91	2225555  T92
# Labeled Trees	15,120	3,024	7,560	2,520
# Regular Vines on each tree	1,232,340	1,869,120	5,255,904	14,889,744
# Tree-Equivalent Reg. Vines / tree	1,195	901	1,328	1,464
Prüfer code example	2244444  T93	2333333  T94	1111111  T95	
# Labeled Trees	1,512	504	9	
# Regular Vines on each tree	23,334,480	62,523,360	660,602,880	
# Tree-Equivalent Reg. Vines / tree	1,464	1,464	1,464	

Table A.7: *Trees with 9 nodes (Continuation).*

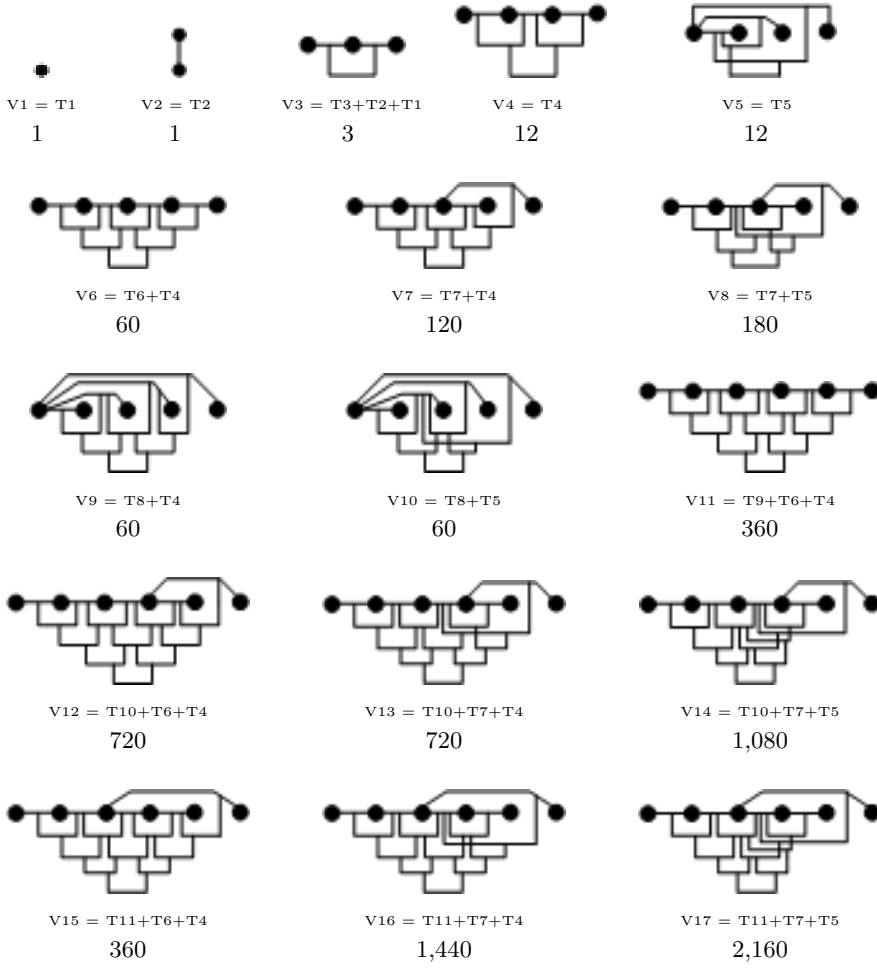


Table A.8: *Tree-equivalent regular vines with at most 6 nodes.*


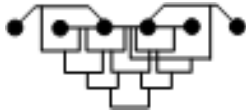








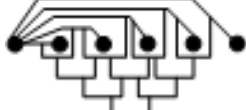




		
V18 = $T_{12}+T_6+T_4$ 360	V19 = $T_{12}+T_7+T_4$ 720	V20 = $T_{12}+T_7+T_5$ 1,080
		
V21 = $T_{12}+T_8+T_4$ 1,080	V22 = $T_{12}+T_8+T_5$ 1,080	V23 = $T_{13}+T_6+T_4$ 720
		
V24 = $T_{13}+T_7+T_4$ 2,160	V25 = $T_{13}+T_7+T_5$ 3,240	V26 = $T_{13}+T_8+T_4$ 1,440
		
V27 = $T_{13}+T_8+T_5$ 1,440	V28 = $T_{14}+T_6+T_4$ 360	V29 = $T_{14}+T_7+T_4$ 720
		
V30 = $T_{14}+T_7+T_5$ 1,080	V31 = $T_{14}+T_8+T_4$ 360	V32 = $T_{14}+T_8+T_5$ 360

Table A.9: Tree-equivalent regular vines with at most 6 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V33 = T15+T9+T6+T4	2,520	V68 = T20+T12+T8+T5	30,240
V34 = T16+T9+T6+T4	5,040	V69 = T20+T13+T6+T4	15,120
V35 = T16+T10+T6+T4	5,040	V70 = T20+T13+T7+T4	45,360
V36 = T16+T10+T7+T4	5,040	V71 = T20+T13+T7+T5	68,040
V37 = T16+T10+T7+T5	7,560	V72 = T20+T13+T8+T4	30,240
V38 = T17+T9+T6+T4	5,040	V73 = T20+T13+T8+T5	30,240
V39 = T17+T10+T6+T4	10,080	V74 = T21+T9+T6+T4	5,040
V40 = T17+T10+T7+T4	10,080	V75 = T21+T10+T6+T4	5,040
V41 = T17+T10+T7+T5	15,120	V76 = T21+T10+T7+T4	5,040
V42 = T17+T11+T6+T4	5,040	V77 = T21+T10+T7+T5	7,560
V43 = T17+T11+T7+T4	20,160	V78 = T21+T11+T6+T4	5,040
V44 = T17+T11+T7+T5	30,240	V79 = T21+T11+T7+T4	20,160
V45 = T18+T11+T6+T4	2,520	V80 = T21+T11+T7+T5	30,240
V46 = T18+T11+T7+T4	10,080	V81 = T21+T13+T6+T4	5,040
V47 = T18+T11+T7+T5	15,120	V82 = T21+T13+T7+T4	15,120
V48 = T19+T9+T6+T4	2,520	V83 = T21+T13+T7+T5	22,680
V49 = T19+T10+T6+T4	5,040	V84 = T21+T13+T8+T4	10,080
V50 = T19+T10+T7+T4	5,040	V85 = T21+T13+T8+T5	10,080
V51 = T19+T10+T7+T5	7,560	V86 = T22+T9+T6+T4	2,520
V52 = T19+T12+T6+T4	2,520	V87 = T22+T10+T6+T4	10,080
V53 = T19+T12+T7+T4	5,040	V88 = T22+T10+T7+T4	10,080
V54 = T19+T12+T7+T5	7,560	V89 = T22+T10+T7+T5	15,120
V55 = T19+T12+T8+T4	7,560	V90 = T22+T11+T6+T4	7,560
V56 = T19+T12+T8+T5	7,560	V91 = T22+T11+T7+T4	30,240
V57 = T20+T9+T6+T4	5,040	V92 = T22+T11+T7+T5	45,360
V58 = T20+T10+T6+T4	15,120	V93 = T22+T12+T6+T4	10,080
V59 = T20+T10+T7+T4	15,120	V94 = T22+T12+T7+T4	20,160
V60 = T20+T10+T7+T5	22,680	V95 = T22+T12+T7+T5	30,240
V61 = T20+T11+T6+T4	5,040	V96 = T22+T12+T8+T4	30,240
V62 = T20+T11+T7+T4	20,160	V97 = T22+T12+T8+T5	30,240
V63 = T20+T11+T7+T5	30,240	V98 = T22+T13+T6+T4	15,120
V64 = T20+T12+T6+T4	10,080	V99 = T22+T13+T7+T4	45,360
V65 = T20+T12+T7+T4	20,160	V100 = T22+T13+T7+T5	68,040
V66 = T20+T12+T7+T5	30,240	V101 = T22+T13+T8+T4	30,240
V67 = T20+T12+T8+T4	30,240	V102 = T22+T13+T8+T5	30,240

Table A.10: Tree-equivalent regular vines with 7 nodes.

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V103 = T23+T9+T6+T4	5,040	V136 = T24+T12+T8+T5	30,240
V104 = T23+T10+T6+T4	10,080	V137 = T24+T13+T6+T4	20,160
V105 = T23+T10+T7+T4	10,080	V138 = T24+T13+T7+T4	60,480
V106 = T23+T10+T7+T5	15,120	V139 = T24+T13+T7+T5	90,720
V107 = T23+T11+T6+T4	5,040	V140 = T24+T13+T8+T4	40,320
V108 = T23+T11+T7+T4	20,160	V141 = T24+T13+T8+T5	40,320
V109 = T23+T11+T7+T5	30,240	V142 = T24+T14+T6+T4	12,600
V110 = T23+T12+T6+T4	5,040	V143 = T24+T14+T7+T4	25,200
V111 = T23+T12+T7+T4	10,080	V144 = T24+T14+T7+T5	37,800
V112 = T23+T12+T7+T5	15,120	V145 = T24+T14+T8+T4	12,600
V113 = T23+T12+T8+T4	15,120	V146 = T24+T14+T8+T5	12,600
V114 = T23+T12+T8+T5	15,120	V147 = T25+T9+T6+T4	2,520
V115 = T23+T13+T6+T4	20,160	V148 = T25+T10+T6+T4	5,040
V116 = T23+T13+T7+T4	60,480	V149 = T25+T10+T7+T4	5,040
V117 = T23+T13+T7+T5	90,720	V150 = T25+T10+T7+T5	7,560
V118 = T23+T13+T8+T4	40,320	V151 = T25+T11+T6+T4	2,520
V119 = T23+T13+T8+T5	40,320	V152 = T25+T11+T7+T4	10,080
V120 = T23+T14+T6+T4	25,200	V153 = T25+T11+T7+T5	15,120
V121 = T23+T14+T7+T4	50,400	V154 = T25+T12+T6+T4	2,520
V122 = T23+T14+T7+T5	75,600	V155 = T25+T12+T7+T4	5,040
V123 = T23+T14+T8+T4	25,200	V156 = T25+T12+T7+T5	7,560
V124 = T23+T14+T8+T5	25,200	V157 = T25+T12+T8+T4	7,560
V125 = T24+T9+T6+T4	5,040	V158 = T25+T12+T8+T5	7,560
V126 = T24+T10+T6+T4	15,120	V159 = T25+T13+T6+T4	5,040
V127 = T24+T10+T7+T4	15,120	V160 = T25+T13+T7+T4	15,120
V128 = T24+T10+T7+T5	22,680	V161 = T25+T13+T7+T5	22,680
V129 = T24+T11+T6+T4	7,560	V162 = T25+T13+T8+T4	10,080
V130 = T24+T11+T7+T4	30,240	V163 = T25+T13+T8+T5	10,080
V131 = T24+T11+T7+T5	45,360	V164 = T25+T14+T6+T4	2,520
V132 = T24+T12+T6+T4	10,080	V165 = T25+T14+T7+T4	5,040
V133 = T24+T12+T7+T4	20,160	V166 = T25+T14+T7+T5	7,560
V134 = T24+T12+T7+T5	30,240	V167 = T25+T14+T8+T4	2,520
V135 = T24+T12+T8+T4	30,240	V168 = T25+T14+T8+T5	2,520

Table A.11: *Tree-equivalent regular vines with 7 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V169 = T26+T15+T9+T6+T4	20,160	V204 = T30+T17+T9+T6+T4	40,320
V170 = T27+T16+T10+T7+T5	60,480	V205 = T31+T22+T11+T6+T4	60,480
V171 = T27+T16+T10+T7+T4	40,320	V206 = T31+T22+T11+T7+T4	241,920
V172 = T27+T16+T10+T6+T4	40,320	V207 = T31+T22+T11+T7+T5	362,880
V173 = T27+T16+T9+T6+T4	40,320	V208 = T31+T22+T9+T6+T4	20,160
V174 = T27+T15+T9+T6+T4	40,320	V209 = T31+T22+T10+T7+T5	120,960
V175 = T28+T16+T9+T6+T4	80,640	V210 = T31+T22+T10+T7+T4	80,640
V176 = T28+T16+T10+T6+T4	80,640	V211 = T31+T22+T10+T6+T4	80,640
V177 = T28+T16+T10+T7+T4	80,640	V212 = T31+T22+T13+T7+T4	362,880
V178 = T28+T16+T10+T7+T5	120,960	V213 = T31+T22+T13+T7+T5	544,320
V179 = T28+T17+T10+T6+T4	80,640	V214 = T31+T22+T13+T6+T4	120,960
V180 = T28+T17+T10+T7+T4	80,640	V215 = T31+T22+T13+T8+T5	241,920
V181 = T28+T17+T10+T7+T5	120,960	V216 = T31+T22+T13+T8+T4	241,920
V182 = T28+T17+T9+T6+T4	40,320	V217 = T31+T22+T12+T6+T4	80,640
V183 = T28+T17+T11+T7+T4	161,280	V218 = T31+T22+T12+T7+T5	241,920
V184 = T28+T17+T11+T7+T5	241,920	V219 = T31+T22+T12+T7+T4	161,280
V185 = T28+T17+T11+T6+T4	40,320	V220 = T31+T22+T12+T8+T5	241,920
V186 = T28+T15+T9+T6+T4	40,320	V221 = T31+T22+T12+T8+T4	241,920
V187 = T29+T17+T10+T6+T4	80,640	V222 = T31+T20+T11+T7+T4	161,280
V188 = T29+T17+T10+T7+T4	80,640	V223 = T31+T20+T11+T7+T5	241,920
V189 = T29+T17+T10+T7+T5	120,960	V224 = T31+T20+T11+T6+T4	40,320
V190 = T29+T17+T11+T7+T4	161,280	V225 = T31+T20+T10+T7+T5	181,440
V191 = T29+T17+T11+T7+T5	241,920	V226 = T31+T20+T10+T7+T4	120,960
V192 = T29+T17+T11+T6+T4	40,320	V227 = T31+T20+T10+T6+T4	120,960
V193 = T29+T17+T9+T6+T4	40,320	V228 = T31+T20+T9+T6+T4	40,320
V194 = T29+T15+T9+T6+T4	20,160	V229 = T31+T20+T13+T7+T4	362,880
V195 = T30+T18+T11+T6+T4	60,480	V230 = T31+T20+T13+T7+T5	544,320
V196 = T30+T18+T11+T7+T4	241,920	V231 = T31+T20+T13+T6+T4	120,960
V197 = T30+T18+T11+T7+T5	362,880	V232 = T31+T20+T13+T8+T5	241,920
V198 = T30+T17+T11+T7+T4	161,280	V233 = T31+T20+T13+T8+T4	241,920
V199 = T30+T17+T11+T7+T5	241,920	V234 = T31+T20+T12+T7+T5	241,920
V200 = T30+T17+T11+T6+T4	40,320	V235 = T31+T20+T12+T7+T4	161,280
V201 = T30+T17+T10+T7+T5	120,960	V236 = T31+T20+T12+T6+T4	80,640
V202 = T30+T17+T10+T7+T4	80,640	V237 = T31+T20+T12+T8+T5	241,920
V203 = T30+T17+T10+T6+T4	80,640	V238 = T31+T20+T12+T8+T4	241,920

Table A.12: *Tree-equivalent regular vines with 8 nodes.*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V239 = T31+T18+T11+T7+T5	362,880	V274 = T33+T16+T10+T7+T4	120,960
V240 = T31+T18+T11+T7+T4	241,920	V275 = T33+T16+T10+T7+T5	181,440
V241 = T31+T18+T11+T6+T4	60,480	V276 = T33+T20+T12+T7+T5	241,920
V242 = T31+T17+T11+T7+T4	161,280	V277 = T33+T20+T12+T7+T4	161,280
V243 = T31+T17+T11+T7+T5	241,920	V278 = T33+T20+T12+T6+T4	80,640
V244 = T31+T17+T11+T6+T4	40,320	V279 = T33+T20+T12+T8+T5	241,920
V245 = T31+T17+T10+T7+T5	120,960	V280 = T33+T20+T12+T8+T4	241,920
V246 = T31+T17+T10+T7+T4	80,640	V281 = T33+T20+T10+T6+T4	120,960
V247 = T31+T17+T10+T6+T4	80,640	V282 = T33+T20+T10+T7+T4	120,960
V248 = T31+T17+T9+T6+T4	40,320	V283 = T33+T20+T10+T7+T5	181,440
V249 = T32+T19+T10+T7+T5	60,480	V284 = T33+T20+T9+T6+T4	40,320
V250 = T32+T19+T10+T7+T4	40,320	V285 = T33+T20+T13+T8+T5	241,920
V251 = T32+T19+T10+T6+T4	40,320	V286 = T33+T20+T13+T8+T4	241,920
V252 = T32+T19+T9+T6+T4	20,160	V287 = T33+T20+T13+T7+T4	362,880
V253 = T32+T19+T12+T7+T5	60,480	V288 = T33+T20+T13+T7+T5	544,320
V254 = T32+T19+T12+T7+T4	40,320	V289 = T33+T20+T13+T6+T4	120,960
V255 = T32+T19+T12+T6+T4	20,160	V290 = T33+T20+T11+T7+T4	161,280
V256 = T32+T19+T12+T8+T5	60,480	V291 = T33+T20+T11+T7+T5	241,920
V257 = T32+T19+T12+T8+T4	60,480	V292 = T33+T20+T11+T6+T4	40,320
V258 = T32+T16+T9+T6+T4	40,320	V293 = T33+T17+T10+T6+T4	80,640
V259 = T32+T16+T10+T6+T4	40,320	V294 = T33+T17+T10+T7+T4	80,640
V260 = T32+T16+T10+T7+T4	40,320	V295 = T33+T17+T10+T7+T5	120,960
V261 = T32+T16+T10+T7+T5	60,480	V296 = T33+T17+T9+T6+T4	40,320
V262 = T32+T15+T9+T6+T4	20,160	V297 = T33+T17+T11+T7+T4	161,280
V263 = T33+T19+T10+T7+T5	241,920	V298 = T33+T17+T11+T7+T5	241,920
V264 = T33+T19+T10+T7+T4	161,280	V299 = T33+T17+T11+T6+T4	40,320
V265 = T33+T19+T10+T6+T4	161,280	V300 = T33+T15+T9+T6+T4	40,320
V266 = T33+T19+T9+T6+T4	80,640	V301 = T34+T20+T11+T7+T4	322,560
V267 = T33+T19+T12+T7+T5	241,920	V302 = T34+T20+T11+T7+T5	483,840
V268 = T33+T19+T12+T7+T4	161,280	V303 = T34+T20+T11+T6+T4	80,640
V269 = T33+T19+T12+T6+T4	80,640	V304 = T34+T20+T10+T7+T5	362,880
V270 = T33+T19+T12+T8+T5	241,920	V305 = T34+T20+T10+T7+T4	241,920
V271 = T33+T19+T12+T8+T4	241,920	V306 = T34+T20+T10+T6+T4	241,920
V272 = T33+T16+T9+T6+T4	120,960	V307 = T34+T20+T9+T6+T4	80,640
V273 = T33+T16+T10+T6+T4	120,960	V308 = T34+T20+T13+T7+T4	725,760

Table A.13: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V309 = T34+T20+T13+T7+T5	1,088,640	V344 = T34+T22+T11+T6+T4	120,960
V310 = T34+T20+T13+T6+T4	241,920	V345 = T34+T22+T11+T7+T4	483,840
V311 = T34+T20+T13+T8+T5	483,840	V346 = T34+T22+T11+T7+T5	725,760
V312 = T34+T20+T13+T8+T4	483,840	V347 = T34+T22+T9+T6+T4	40,320
V313 = T34+T20+T12+T7+T5	483,840	V348 = T34+T17+T10+T6+T4	80,640
V314 = T34+T20+T12+T7+T4	322,560	V349 = T34+T17+T10+T7+T4	80,640
V315 = T34+T20+T12+T6+T4	161,280	V350 = T34+T17+T10+T7+T5	120,960
V316 = T34+T20+T12+T8+T5	483,840	V351 = T34+T17+T9+T6+T4	40,320
V317 = T34+T20+T12+T8+T4	483,840	V352 = T34+T17+T11+T7+T4	161,280
V318 = T34+T19+T10+T7+T5	241,920	V353 = T34+T17+T11+T7+T5	241,920
V319 = T34+T19+T10+T7+T4	161,280	V354 = T34+T17+T11+T6+T4	40,320
V320 = T34+T19+T10+T6+T4	161,280	V355 = T34+T15+T9+T6+T4	20,160
V321 = T34+T19+T9+T6+T4	80,640	V356 = T35+T20+T12+T7+T5	241,920
V322 = T34+T19+T12+T7+T5	241,920	V357 = T35+T20+T12+T7+T4	161,280
V323 = T34+T19+T12+T7+T4	161,280	V358 = T35+T20+T12+T6+T4	80,640
V324 = T34+T19+T12+T6+T4	80,640	V359 = T35+T20+T12+T8+T5	241,920
V325 = T34+T19+T12+T8+T5	241,920	V360 = T35+T20+T12+T8+T4	241,920
V326 = T34+T19+T12+T8+T4	241,920	V361 = T35+T20+T10+T6+T4	120,960
V327 = T34+T16+T9+T6+T4	80,640	V362 = T35+T20+T10+T7+T4	120,960
V328 = T34+T16+T10+T6+T4	80,640	V363 = T35+T20+T10+T7+T5	181,440
V329 = T34+T16+T10+T7+T4	80,640	V364 = T35+T20+T13+T8+T5	241,920
V330 = T34+T16+T10+T7+T5	120,960	V365 = T35+T20+T13+T8+T4	241,920
V331 = T34+T22+T13+T7+T4	725,760	V366 = T35+T20+T13+T7+T4	362,880
V332 = T34+T22+T13+T7+T5	1,088,640	V367 = T35+T20+T13+T7+T5	544,320
V333 = T34+T22+T13+T6+T4	241,920	V368 = T35+T20+T13+T6+T4	120,960
V334 = T34+T22+T13+T8+T5	483,840	V369 = T35+T20+T11+T7+T4	161,280
V335 = T34+T22+T13+T8+T4	483,840	V370 = T35+T20+T11+T7+T5	241,920
V336 = T34+T22+T10+T6+T4	161,280	V371 = T35+T20+T11+T6+T4	40,320
V337 = T34+T22+T10+T7+T4	161,280	V372 = T35+T20+T9+T6+T4	40,320
V338 = T34+T22+T10+T7+T5	241,920	V373 = T35+T17+T10+T6+T4	161,280
V339 = T34+T22+T12+T7+T5	483,840	V374 = T35+T17+T10+T7+T4	161,280
V340 = T34+T22+T12+T7+T4	322,560	V375 = T35+T17+T10+T7+T5	241,920
V341 = T34+T22+T12+T6+T4	161,280	V376 = T35+T17+T11+T7+T4	322,560
V342 = T34+T22+T12+T8+T5	483,840	V377 = T35+T17+T11+T7+T5	483,840
V343 = T34+T22+T12+T8+T4	483,840	V378 = T35+T17+T11+T6+T4	80,640

Table A.14: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V379 = T35+T17+T9+T6+T4	80,640	V414 = T36+T23+T14+T8+T5	604,800
V380 = T35+T21+T13+T8+T5	241,920	V415 = T36+T23+T14+T8+T4	604,800
V381 = T35+T21+T13+T8+T4	241,920	V416 = T36+T23+T14+T7+T5	1,814,400
V382 = T35+T21+T13+T7+T4	362,880	V417 = T36+T23+T14+T7+T4	1,209,600
V383 = T35+T21+T13+T7+T5	544,320	V418 = T36+T23+T14+T6+T4	604,800
V384 = T35+T21+T13+T6+T4	120,960	V419 = T36+T21+T10+T6+T4	120,960
V385 = T35+T21+T11+T7+T4	483,840	V420 = T36+T21+T10+T7+T4	120,960
V386 = T35+T21+T11+T7+T5	725,760	V421 = T36+T21+T10+T7+T5	181,440
V387 = T35+T21+T11+T6+T4	120,960	V422 = T36+T21+T9+T6+T4	120,960
V388 = T35+T21+T10+T7+T5	181,440	V423 = T36+T21+T11+T7+T4	483,840
V389 = T35+T21+T10+T7+T4	120,960	V424 = T36+T21+T11+T7+T5	725,760
V390 = T35+T21+T10+T6+T4	120,960	V425 = T36+T21+T11+T6+T4	120,960
V391 = T35+T21+T9+T6+T4	120,960	V426 = T36+T21+T13+T7+T4	362,880
V392 = T35+T16+T10+T7+T5	60,480	V427 = T36+T21+T13+T7+T5	544,320
V393 = T35+T16+T10+T7+T4	40,320	V428 = T36+T21+T13+T6+T4	120,960
V394 = T35+T16+T10+T6+T4	40,320	V429 = T36+T21+T13+T8+T5	241,920
V395 = T35+T16+T9+T6+T4	40,320	V430 = T36+T21+T13+T8+T4	241,920
V396 = T35+T15+T9+T6+T4	40,320	V431 = T36+T20+T11+T7+T4	161,280
V397 = T36+T23+T12+T7+T5	362,880	V432 = T36+T20+T11+T7+T5	241,920
V398 = T36+T23+T12+T7+T4	241,920	V433 = T36+T20+T11+T6+T4	40,320
V399 = T36+T23+T12+T6+T4	120,960	V434 = T36+T20+T9+T6+T4	40,320
V400 = T36+T23+T12+T8+T5	362,880	V435 = T36+T20+T10+T7+T5	181,440
V401 = T36+T23+T12+T8+T4	362,880	V436 = T36+T20+T10+T7+T4	120,960
V402 = T36+T23+T10+T6+T4	241,920	V437 = T36+T20+T10+T6+T4	120,960
V403 = T36+T23+T10+T7+T4	241,920	V438 = T36+T20+T13+T7+T4	362,880
V404 = T36+T23+T10+T7+T5	362,880	V439 = T36+T20+T13+T7+T5	544,320
V405 = T36+T23+T9+T6+T4	120,960	V440 = T36+T20+T13+T6+T4	120,960
V406 = T36+T23+T13+T8+T5	967,680	V441 = T36+T20+T13+T8+T5	241,920
V407 = T36+T23+T13+T8+T4	967,680	V442 = T36+T20+T13+T8+T4	241,920
V408 = T36+T23+T13+T7+T4	1,451,520	V443 = T36+T20+T12+T7+T5	241,920
V409 = T36+T23+T13+T7+T5	2,177,280	V444 = T36+T20+T12+T7+T4	161,280
V410 = T36+T23+T13+T6+T4	483,840	V445 = T36+T20+T12+T6+T4	80,640
V411 = T36+T23+T11+T7+T4	483,840	V446 = T36+T20+T12+T8+T5	241,920
V412 = T36+T23+T11+T7+T5	725,760	V447 = T36+T20+T12+T8+T4	241,920
V413 = T36+T23+T11+T6+T4	120,960	V448 = T36+T19+T10+T7+T5	60,480

Table A.15: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V449 = T36+T19+T10+T7+T4	40,320	V484 = T37+T22+T12+T8+T5	483,840
V450 = T36+T19+T10+T6+T4	40,320	V485 = T37+T22+T12+T8+T4	483,840
V451 = T36+T19+T9+T6+T4	20,160	V486 = T37+T20+T11+T7+T4	322,560
V452 = T36+T19+T12+T7+T5	60,480	V487 = T37+T20+T11+T7+T5	483,840
V453 = T36+T19+T12+T7+T4	40,320	V488 = T37+T20+T11+T6+T4	80,640
V454 = T36+T19+T12+T6+T4	20,160	V489 = T37+T20+T10+T7+T5	362,880
V455 = T36+T19+T12+T8+T5	60,480	V490 = T37+T20+T10+T7+T4	241,920
V456 = T36+T19+T12+T8+T4	60,480	V491 = T37+T20+T10+T6+T4	241,920
V457 = T36+T16+T9+T6+T4	40,320	V492 = T37+T20+T9+T6+T4	80,640
V458 = T36+T16+T10+T6+T4	40,320	V493 = T37+T20+T13+T7+T4	725,760
V459 = T36+T16+T10+T7+T4	40,320	V494 = T37+T20+T13+T7+T5	1,088,640
V460 = T36+T16+T10+T7+T5	60,480	V495 = T37+T20+T13+T6+T4	241,920
V461 = T36+T17+T10+T7+T4	80,640	V496 = T37+T20+T13+T8+T5	483,840
V462 = T36+T17+T10+T7+T5	120,960	V497 = T37+T20+T13+T8+T4	483,840
V463 = T36+T17+T10+T6+T4	80,640	V498 = T37+T20+T12+T7+T5	483,840
V464 = T36+T17+T9+T6+T4	40,320	V499 = T37+T20+T12+T7+T4	322,560
V465 = T36+T17+T11+T7+T4	161,280	V500 = T37+T20+T12+T6+T4	161,280
V466 = T36+T17+T11+T7+T5	241,920	V501 = T37+T20+T12+T8+T5	483,840
V467 = T36+T17+T11+T6+T4	40,320	V502 = T37+T20+T12+T8+T4	483,840
V468 = T36+T15+T9+T6+T4	20,160	V503 = T37+T17+T9+T6+T4	40,320
V469 = T37+T22+T11+T7+T5	725,760	V504 = T37+T17+T10+T6+T4	80,640
V470 = T37+T22+T11+T7+T4	483,840	V505 = T37+T17+T10+T7+T4	80,640
V471 = T37+T22+T11+T6+T4	120,960	V506 = T37+T17+T10+T7+T5	120,960
V472 = T37+T22+T9+T6+T4	40,320	V507 = T37+T17+T11+T6+T4	40,320
V473 = T37+T22+T10+T7+T5	241,920	V508 = T37+T17+T11+T7+T4	161,280
V474 = T37+T22+T10+T7+T4	161,280	V509 = T37+T17+T11+T7+T5	241,920
V475 = T37+T22+T10+T6+T4	161,280	V510 = T37+T18+T11+T6+T4	20,160
V476 = T37+T22+T13+T7+T5	1,088,640	V511 = T37+T18+T11+T7+T4	80,640
V477 = T37+T22+T13+T7+T4	725,760	V512 = T37+T18+T11+T7+T5	120,960
V478 = T37+T22+T13+T6+T4	241,920	V513 = T38+T23+T12+T7+T5	1,088,640
V479 = T37+T22+T13+T8+T5	483,840	V514 = T38+T23+T12+T7+T4	725,760
V480 = T37+T22+T13+T8+T4	483,840	V515 = T38+T23+T12+T6+T4	362,880
V481 = T37+T22+T12+T7+T5	483,840	V516 = T38+T23+T12+T8+T5	1,088,640
V482 = T37+T22+T12+T7+T4	322,560	V517 = T38+T23+T12+T8+T4	1,088,640
V483 = T37+T22+T12+T6+T4	161,280	V518 = T38+T23+T10+T6+T4	725,760

Table A.16: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V519 = T38+T23+T10+T7+T4	725,760	V554 = T38+T19+T12+T8+T5	241,920
V520 = T38+T23+T10+T7+T5	1,088,640	V555 = T38+T19+T12+T8+T4	241,920
V521 = T38+T23+T9+T6+T4	362,880	V556 = T38+T16+T9+T6+T4	120,960
V522 = T38+T23+T13+T8+T5	2,903,040	V557 = T38+T16+T10+T6+T4	120,960
V523 = T38+T23+T13+T8+T4	2,903,040	V558 = T38+T16+T10+T7+T4	120,960
V524 = T38+T23+T13+T7+T4	4,354,560	V559 = T38+T16+T10+T7+T5	181,440
V525 = T38+T23+T13+T7+T5	6,531,840	V560 = T38+T20+T11+T7+T4	645,120
V526 = T38+T23+T13+T6+T4	1,451,520	V561 = T38+T20+T11+T7+T5	967,680
V527 = T38+T23+T11+T7+T4	1,451,520	V562 = T38+T20+T11+T6+T4	161,280
V528 = T38+T23+T11+T7+T5	2,177,280	V563 = T38+T20+T9+T6+T4	161,280
V529 = T38+T23+T11+T6+T4	362,880	V564 = T38+T20+T10+T7+T5	725,760
V530 = T38+T23+T14+T8+T5	1,814,400	V565 = T38+T20+T10+T7+T4	483,840
V531 = T38+T23+T14+T8+T4	1,814,400	V566 = T38+T20+T10+T6+T4	483,840
V532 = T38+T23+T14+T7+T5	5,443,200	V567 = T38+T20+T13+T7+T4	1,451,520
V533 = T38+T23+T14+T7+T4	3,628,800	V568 = T38+T20+T13+T7+T5	2,177,280
V534 = T38+T23+T14+T6+T4	1,814,400	V569 = T38+T20+T13+T6+T4	483,840
V535 = T38+T21+T10+T6+T4	241,920	V570 = T38+T20+T13+T8+T5	967,680
V536 = T38+T21+T10+T7+T4	241,920	V571 = T38+T20+T13+T8+T4	967,680
V537 = T38+T21+T10+T7+T5	362,880	V572 = T38+T20+T12+T7+T5	967,680
V538 = T38+T21+T9+T6+T4	241,920	V573 = T38+T20+T12+T7+T4	645,120
V539 = T38+T21+T11+T7+T4	967,680	V574 = T38+T20+T12+T6+T4	322,560
V540 = T38+T21+T11+T7+T5	1,451,520	V575 = T38+T20+T12+T8+T5	967,680
V541 = T38+T21+T11+T6+T4	241,920	V576 = T38+T20+T12+T8+T4	967,680
V542 = T38+T21+T13+T7+T4	725,760	V577 = T38+T17+T9+T6+T4	120,960
V543 = T38+T21+T13+T7+T5	1,088,640	V578 = T38+T17+T10+T6+T4	241,920
V544 = T38+T21+T13+T6+T4	241,920	V579 = T38+T17+T10+T7+T4	241,920
V545 = T38+T21+T13+T8+T5	483,840	V580 = T38+T17+T10+T7+T5	362,880
V546 = T38+T21+T13+T8+T4	483,840	V581 = T38+T17+T11+T6+T4	120,960
V547 = T38+T19+T10+T7+T5	241,920	V582 = T38+T17+T11+T7+T4	483,840
V548 = T38+T19+T10+T7+T4	161,280	V583 = T38+T17+T11+T7+T5	725,760
V549 = T38+T19+T10+T6+T4	161,280	V584 = T38+T15+T9+T6+T4	40,320
V550 = T38+T19+T9+T6+T4	80,640	V585 = T38+T22+T13+T7+T5	2,177,280
V551 = T38+T19+T12+T7+T5	241,920	V586 = T38+T22+T13+T7+T4	1,451,520
V552 = T38+T19+T12+T7+T4	161,280	V587 = T38+T22+T13+T6+T4	483,840
V553 = T38+T19+T12+T6+T4	80,640	V588 = T38+T22+T13+T8+T5	967,680

Table A.17: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V589 = T38+T22+T13+T8+T4	967,680	V624 = T38+T24+T11+T7+T4	1,451,520
V590 = T38+T22+T12+T7+T5	967,680	V625 = T38+T24+T11+T7+T5	2,177,280
V591 = T38+T22+T12+T7+T4	645,120	V626 = T38+T24+T9+T6+T4	241,920
V592 = T38+T22+T12+T6+T4	322,560	V627 = T39+T21+T13+T8+T5	80,640
V593 = T38+T22+T12+T8+T5	967,680	V628 = T39+T21+T13+T8+T4	80,640
V594 = T38+T22+T12+T8+T4	967,680	V629 = T39+T21+T13+T7+T4	120,960
V595 = T38+T22+T11+T6+T4	241,920	V630 = T39+T21+T13+T7+T5	181,440
V596 = T38+T22+T11+T7+T4	967,680	V631 = T39+T21+T13+T6+T4	40,320
V597 = T38+T22+T11+T7+T5	1,451,520	V632 = T39+T21+T11+T7+T4	161,280
V598 = T38+T22+T10+T6+T4	322,560	V633 = T39+T21+T11+T7+T5	241,920
V599 = T38+T22+T10+T7+T4	322,560	V634 = T39+T21+T11+T6+T4	40,320
V600 = T38+T22+T10+T7+T5	483,840	V635 = T39+T21+T10+T7+T5	60,480
V601 = T38+T22+T9+T6+T4	80,640	V636 = T39+T21+T10+T7+T4	40,320
V602 = T38+T18+T11+T6+T4	120,960	V637 = T39+T21+T10+T6+T4	40,320
V603 = T38+T18+T11+T7+T4	483,840	V638 = T39+T21+T9+T6+T4	40,320
V604 = T38+T18+T11+T7+T5	725,760	V639 = T39+T17+T11+T7+T4	161,280
V605 = T38+T24+T14+T8+T5	604,800	V640 = T39+T17+T11+T7+T5	241,920
V606 = T38+T24+T14+T8+T4	604,800	V641 = T39+T17+T11+T6+T4	40,320
V607 = T38+T24+T14+T7+T5	1,814,400	V642 = T39+T17+T10+T7+T5	120,960
V608 = T38+T24+T14+T7+T4	1,209,600	V643 = T39+T17+T10+T7+T4	80,640
V609 = T38+T24+T14+T6+T4	604,800	V644 = T39+T17+T10+T6+T4	80,640
V610 = T38+T24+T13+T7+T5	4,354,560	V645 = T39+T17+T9+T6+T4	40,320
V611 = T38+T24+T13+T7+T4	2,903,040	V646 = T39+T16+T10+T7+T5	60,480
V612 = T38+T24+T13+T6+T4	967,680	V647 = T39+T16+T10+T7+T4	40,320
V613 = T38+T24+T13+T8+T5	1,935,360	V648 = T39+T16+T10+T6+T4	40,320
V614 = T38+T24+T13+T8+T4	1,935,360	V649 = T39+T16+T9+T6+T4	40,320
V615 = T38+T24+T12+T7+T5	1,451,520	V650 = T39+T15+T9+T6+T4	40,320
V616 = T38+T24+T12+T7+T4	967,680	V651 = T40+T22+T13+T7+T5	1,088,640
V617 = T38+T24+T12+T6+T4	483,840	V652 = T40+T22+T13+T7+T4	725,760
V618 = T38+T24+T12+T8+T5	1,451,520	V653 = T40+T22+T13+T6+T4	241,920
V619 = T38+T24+T12+T8+T4	1,451,520	V654 = T40+T22+T13+T8+T5	483,840
V620 = T38+T24+T10+T6+T4	725,760	V655 = T40+T22+T13+T8+T4	483,840
V621 = T38+T24+T10+T7+T4	725,760	V656 = T40+T22+T12+T7+T5	483,840
V622 = T38+T24+T10+T7+T5	1,088,640	V657 = T40+T22+T12+T7+T4	322,560
V623 = T38+T24+T11+T6+T4	362,880	V658 = T40+T22+T12+T6+T4	161,280

Table A.18: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V659 = T40+T22+T12+T8+T5	483,840	V694 = T40+T21+T13+T7+T4	362,880
V660 = T40+T22+T12+T8+T4	483,840	V695 = T40+T21+T13+T7+T5	544,320
V661 = T40+T22+T10+T6+T4	161,280	V696 = T40+T21+T13+T6+T4	120,960
V662 = T40+T22+T10+T7+T4	161,280	V697 = T40+T21+T11+T7+T4	483,840
V663 = T40+T22+T10+T7+T5	241,920	V698 = T40+T21+T11+T7+T5	725,760
V664 = T40+T22+T11+T7+T4	483,840	V699 = T40+T21+T11+T6+T4	120,960
V665 = T40+T22+T11+T7+T5	725,760	V700 = T40+T21+T10+T7+T5	181,440
V666 = T40+T22+T11+T6+T4	120,960	V701 = T40+T21+T10+T7+T4	120,960
V667 = T40+T22+T9+T6+T4	40,320	V702 = T40+T21+T10+T6+T4	120,960
V668 = T40+T20+T12+T7+T5	483,840	V703 = T40+T21+T9+T6+T4	120,960
V669 = T40+T20+T12+T7+T4	322,560	V704 = T40+T18+T11+T6+T4	120,960
V670 = T40+T20+T12+T6+T4	161,280	V705 = T40+T18+T11+T7+T4	483,840
V671 = T40+T20+T12+T8+T5	483,840	V706 = T40+T18+T11+T7+T5	725,760
V672 = T40+T20+T12+T8+T4	483,840	V707 = T40+T16+T10+T7+T5	120,960
V673 = T40+T20+T10+T6+T4	241,920	V708 = T40+T16+T10+T7+T4	80,640
V674 = T40+T20+T10+T7+T4	241,920	V709 = T40+T16+T10+T6+T4	80,640
V675 = T40+T20+T10+T7+T5	362,880	V710 = T40+T16+T9+T6+T4	80,640
V676 = T40+T20+T13+T8+T5	483,840	V711 = T40+T15+T9+T6+T4	40,320
V677 = T40+T20+T13+T8+T4	483,840	V712 = T41+T23+T13+T8+T5	322,560
V678 = T40+T20+T13+T7+T4	725,760	V713 = T41+T23+T13+T8+T4	322,560
V679 = T40+T20+T13+T7+T5	1,088,640	V714 = T41+T23+T13+T7+T4	483,840
V680 = T40+T20+T13+T6+T4	241,920	V715 = T41+T23+T13+T7+T5	725,760
V681 = T40+T20+T11+T7+T4	322,560	V716 = T41+T23+T13+T6+T4	161,280
V682 = T40+T20+T11+T7+T5	483,840	V717 = T41+T23+T11+T7+T4	161,280
V683 = T40+T20+T11+T6+T4	80,640	V718 = T41+T23+T11+T7+T5	241,920
V684 = T40+T20+T9+T6+T4	80,640	V719 = T41+T23+T11+T6+T4	40,320
V685 = T40+T17+T10+T6+T4	241,920	V720 = T41+T23+T10+T7+T5	120,960
V686 = T40+T17+T10+T7+T4	241,920	V721 = T41+T23+T10+T7+T4	80,640
V687 = T40+T17+T10+T7+T5	362,880	V722 = T41+T23+T10+T6+T4	80,640
V688 = T40+T17+T11+T7+T4	483,840	V723 = T41+T23+T9+T6+T4	40,320
V689 = T40+T17+T11+T7+T5	725,760	V724 = T41+T23+T14+T8+T5	201,600
V690 = T40+T17+T11+T6+T4	120,960	V725 = T41+T23+T14+T8+T4	201,600
V691 = T40+T17+T9+T6+T4	120,960	V726 = T41+T23+T14+T7+T5	604,800
V692 = T40+T21+T13+T8+T5	241,920	V727 = T41+T23+T14+T7+T4	403,200
V693 = T40+T21+T13+T8+T4	241,920	V728 = T41+T23+T14+T6+T4	201,600

Table A.19: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V729 = T41+T23+T12+T7+T5	120,960	V764 = T41+T21+T13+T8+T5	80,640
V730 = T41+T23+T12+T7+T4	80,640	V765 = T41+T21+T13+T8+T4	80,640
V731 = T41+T23+T12+T6+T4	40,320	V766 = T41+T21+T13+T7+T4	120,960
V732 = T41+T23+T12+T8+T5	120,960	V767 = T41+T21+T13+T7+T5	181,440
V733 = T41+T23+T12+T8+T4	120,960	V768 = T41+T21+T13+T6+T4	40,320
V734 = T41+T20+T11+T7+T4	161,280	V769 = T41+T21+T11+T7+T4	161,280
V735 = T41+T20+T11+T7+T5	241,920	V770 = T41+T21+T11+T7+T5	241,920
V736 = T41+T20+T11+T6+T4	40,320	V771 = T41+T21+T11+T6+T4	40,320
V737 = T41+T20+T10+T7+T5	181,440	V772 = T41+T21+T10+T7+T5	60,480
V738 = T41+T20+T10+T7+T4	120,960	V773 = T41+T21+T10+T7+T4	40,320
V739 = T41+T20+T10+T6+T4	120,960	V774 = T41+T21+T10+T6+T4	40,320
V740 = T41+T20+T9+T6+T4	40,320	V775 = T41+T21+T9+T6+T4	40,320
V741 = T41+T20+T13+T7+T4	362,880	V776 = T41+T17+T11+T7+T4	161,280
V742 = T41+T20+T13+T7+T5	544,320	V777 = T41+T17+T11+T7+T5	241,920
V743 = T41+T20+T13+T6+T4	120,960	V778 = T41+T17+T11+T6+T4	40,320
V744 = T41+T20+T13+T8+T5	241,920	V779 = T41+T17+T10+T7+T5	120,960
V745 = T41+T20+T13+T8+T4	241,920	V780 = T41+T17+T10+T7+T4	80,640
V746 = T41+T20+T12+T7+T5	241,920	V781 = T41+T17+T10+T6+T4	80,640
V747 = T41+T20+T12+T7+T4	161,280	V782 = T41+T17+T9+T6+T4	40,320
V748 = T41+T20+T12+T6+T4	80,640	V783 = T41+T15+T9+T6+T4	40,320
V749 = T41+T20+T12+T8+T5	241,920	V784 = T42+T23+T13+T8+T5	645,120
V750 = T41+T20+T12+T8+T4	241,920	V785 = T42+T23+T13+T8+T4	645,120
V751 = T41+T19+T10+T7+T5	120,960	V786 = T42+T23+T13+T7+T4	967,680
V752 = T41+T19+T10+T7+T4	80,640	V787 = T42+T23+T13+T7+T5	1,451,520
V753 = T41+T19+T10+T6+T4	80,640	V788 = T42+T23+T13+T6+T4	322,560
V754 = T41+T19+T9+T6+T4	40,320	V789 = T42+T23+T11+T7+T4	322,560
V755 = T41+T19+T12+T7+T5	120,960	V790 = T42+T23+T11+T7+T5	483,840
V756 = T41+T19+T12+T7+T4	80,640	V791 = T42+T23+T11+T6+T4	80,640
V757 = T41+T19+T12+T6+T4	40,320	V792 = T42+T23+T10+T7+T5	241,920
V758 = T41+T19+T12+T8+T5	120,960	V793 = T42+T23+T10+T7+T4	161,280
V759 = T41+T19+T12+T8+T4	120,960	V794 = T42+T23+T10+T6+T4	161,280
V760 = T41+T16+T9+T6+T4	80,640	V795 = T42+T23+T9+T6+T4	80,640
V761 = T41+T16+T10+T6+T4	80,640	V796 = T42+T23+T14+T8+T5	403,200
V762 = T41+T16+T10+T7+T4	80,640	V797 = T42+T23+T14+T8+T4	403,200
V763 = T41+T16+T10+T7+T5	120,960	V798 = T42+T23+T14+T7+T5	1,209,600

Table A.20: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V799 = T42+T23+T14+T7+T4	806,400	V834 = T42+T16+T10+T7+T4	120,960
V800 = T42+T23+T14+T6+T4	403,200	V835 = T42+T16+T10+T7+T5	181,440
V801 = T42+T23+T12+T7+T5	241,920	V836 = T42+T24+T14+T8+T5	403,200
V802 = T42+T23+T12+T7+T4	161,280	V837 = T42+T24+T14+T8+T4	403,200
V803 = T42+T23+T12+T6+T4	80,640	V838 = T42+T24+T14+T7+T5	1,209,600
V804 = T42+T23+T12+T8+T5	241,920	V839 = T42+T24+T14+T7+T4	806,400
V805 = T42+T23+T12+T8+T4	241,920	V840 = T42+T24+T14+T6+T4	403,200
V806 = T42+T20+T11+T7+T4	483,840	V841 = T42+T24+T13+T7+T5	2,903,040
V807 = T42+T20+T11+T7+T5	725,760	V842 = T42+T24+T13+T7+T4	1,935,360
V808 = T42+T20+T11+T6+T4	120,960	V843 = T42+T24+T13+T6+T4	645,120
V809 = T42+T20+T10+T7+T5	544,320	V844 = T42+T24+T13+T8+T5	1,290,240
V810 = T42+T20+T10+T7+T4	362,880	V845 = T42+T24+T13+T8+T4	1,290,240
V811 = T42+T20+T10+T6+T4	362,880	V846 = T42+T24+T12+T7+T5	967,680
V812 = T42+T20+T9+T6+T4	120,960	V847 = T42+T24+T12+T7+T4	645,120
V813 = T42+T20+T13+T7+T4	1,088,640	V848 = T42+T24+T12+T6+T4	322,560
V814 = T42+T20+T13+T7+T5	1,632,960	V849 = T42+T24+T12+T8+T5	967,680
V815 = T42+T20+T13+T6+T4	362,880	V850 = T42+T24+T12+T8+T4	967,680
V816 = T42+T20+T13+T8+T5	725,760	V851 = T42+T24+T10+T6+T4	483,840
V817 = T42+T20+T13+T8+T4	725,760	V852 = T42+T24+T10+T7+T4	483,840
V818 = T42+T20+T12+T7+T5	725,760	V853 = T42+T24+T10+T7+T5	725,760
V819 = T42+T20+T12+T7+T4	483,840	V854 = T42+T24+T11+T6+T4	241,920
V820 = T42+T20+T12+T6+T4	241,920	V855 = T42+T24+T11+T7+T4	967,680
V821 = T42+T20+T12+T8+T5	725,760	V856 = T42+T24+T11+T7+T5	1,451,520
V822 = T42+T20+T12+T8+T4	725,760	V857 = T42+T24+T9+T6+T4	161,280
V823 = T42+T19+T10+T7+T5	241,920	V858 = T42+T22+T13+T7+T4	1,451,520
V824 = T42+T19+T10+T7+T4	161,280	V859 = T42+T22+T13+T7+T5	2,177,280
V825 = T42+T19+T10+T6+T4	161,280	V860 = T42+T22+T13+T6+T4	483,840
V826 = T42+T19+T9+T6+T4	80,640	V861 = T42+T22+T13+T8+T5	967,680
V827 = T42+T19+T12+T7+T5	241,920	V862 = T42+T22+T13+T8+T4	967,680
V828 = T42+T19+T12+T7+T4	161,280	V863 = T42+T22+T10+T6+T4	322,560
V829 = T42+T19+T12+T6+T4	80,640	V864 = T42+T22+T10+T7+T4	322,560
V830 = T42+T19+T12+T8+T5	241,920	V865 = T42+T22+T10+T7+T5	483,840
V831 = T42+T19+T12+T8+T4	241,920	V866 = T42+T22+T12+T7+T5	967,680
V832 = T42+T16+T9+T6+T4	120,960	V867 = T42+T22+T12+T7+T4	645,120
V833 = T42+T16+T10+T6+T4	120,960	V868 = T42+T22+T12+T6+T4	322,560

Table A.21: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V869 = T42+T22+T12+T8+T5	967,680	V904 = T43+T25+T13+T6+T4	403,200
V870 = T42+T22+T12+T8+T4	967,680	V905 = T43+T25+T12+T8+T5	604,800
V871 = T42+T22+T11+T6+T4	241,920	V906 = T43+T25+T12+T8+T4	604,800
V872 = T42+T22+T11+T7+T4	967,680	V907 = T43+T25+T12+T7+T4	403,200
V873 = T42+T22+T11+T7+T5	1,451,520	V908 = T43+T25+T12+T7+T5	604,800
V874 = T42+T22+T9+T6+T4	80,640	V909 = T43+T25+T12+T6+T4	201,600
V875 = T42+T17+T10+T6+T4	161,280	V910 = T43+T25+T10+T7+T4	403,200
V876 = T42+T17+T10+T7+T4	161,280	V911 = T43+T25+T10+T7+T5	604,800
V877 = T42+T17+T10+T7+T5	241,920	V912 = T43+T25+T10+T6+T4	403,200
V878 = T42+T17+T9+T6+T4	80,640	V913 = T43+T25+T11+T7+T5	1,209,600
V879 = T42+T17+T11+T7+T4	322,560	V914 = T43+T25+T11+T7+T4	806,400
V880 = T42+T17+T11+T7+T5	483,840	V915 = T43+T25+T11+T6+T4	201,600
V881 = T42+T17+T11+T6+T4	80,640	V916 = T43+T25+T9+T6+T4	201,600
V882 = T42+T21+T13+T8+T5	80,640	V917 = T43+T24+T13+T7+T4	1,935,360
V883 = T42+T21+T13+T8+T4	80,640	V918 = T43+T24+T13+T7+T5	2,903,040
V884 = T42+T21+T13+T7+T4	120,960	V919 = T43+T24+T13+T6+T4	645,120
V885 = T42+T21+T13+T7+T5	181,440	V920 = T43+T24+T13+T8+T5	1,290,240
V886 = T42+T21+T13+T6+T4	40,320	V921 = T43+T24+T13+T8+T4	1,290,240
V887 = T42+T21+T11+T7+T4	161,280	V922 = T43+T24+T10+T7+T4	483,840
V888 = T42+T21+T11+T7+T5	241,920	V923 = T43+T24+T10+T7+T5	725,760
V889 = T42+T21+T11+T6+T4	40,320	V924 = T43+T24+T10+T6+T4	483,840
V890 = T42+T21+T10+T7+T5	60,480	V925 = T43+T24+T12+T7+T5	967,680
V891 = T42+T21+T10+T7+T4	40,320	V926 = T43+T24+T12+T7+T4	645,120
V892 = T42+T21+T10+T6+T4	40,320	V927 = T43+T24+T12+T6+T4	322,560
V893 = T42+T21+T9+T6+T4	40,320	V928 = T43+T24+T12+T8+T5	967,680
V894 = T42+T15+T9+T6+T4	40,320	V929 = T43+T24+T12+T8+T4	967,680
V895 = T43+T25+T14+T8+T5	201,600	V930 = T43+T24+T11+T7+T4	967,680
V896 = T43+T25+T14+T8+T4	201,600	V931 = T43+T24+T11+T7+T5	1,451,520
V897 = T43+T25+T14+T7+T5	604,800	V932 = T43+T24+T11+T6+T4	241,920
V898 = T43+T25+T14+T7+T4	403,200	V933 = T43+T24+T9+T6+T4	161,280
V899 = T43+T25+T14+T6+T4	201,600	V934 = T43+T24+T14+T8+T5	403,200
V900 = T43+T25+T13+T8+T5	806,400	V935 = T43+T24+T14+T8+T4	403,200
V901 = T43+T25+T13+T8+T4	806,400	V936 = T43+T24+T14+T7+T5	1,209,600
V902 = T43+T25+T13+T7+T5	1,814,400	V937 = T43+T24+T14+T7+T4	806,400
V903 = T43+T25+T13+T7+T4	1,209,600	V938 = T43+T24+T14+T6+T4	403,200

Table A.22: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V939 = T43+T23+T12+T7+T5	120,960	V974 = T43+T22+T13+T7+T4	725,760
V940 = T43+T23+T12+T7+T4	80,640	V975 = T43+T22+T13+T6+T4	241,920
V941 = T43+T23+T12+T6+T4	40,320	V976 = T43+T22+T13+T8+T5	483,840
V942 = T43+T23+T12+T8+T5	120,960	V977 = T43+T22+T13+T8+T4	483,840
V943 = T43+T23+T12+T8+T4	120,960	V978 = T43+T22+T10+T7+T4	161,280
V944 = T43+T23+T10+T6+T4	80,640	V979 = T43+T22+T10+T7+T5	241,920
V945 = T43+T23+T10+T7+T4	80,640	V980 = T43+T22+T10+T6+T4	161,280
V946 = T43+T23+T10+T7+T5	120,960	V981 = T43+T22+T9+T6+T4	40,320
V947 = T43+T23+T9+T6+T4	40,320	V982 = T43+T22+T12+T7+T5	483,840
V948 = T43+T23+T13+T8+T5	322,560	V983 = T43+T22+T12+T7+T4	322,560
V949 = T43+T23+T13+T8+T4	322,560	V984 = T43+T22+T12+T6+T4	161,280
V950 = T43+T23+T13+T7+T4	483,840	V985 = T43+T22+T12+T8+T5	483,840
V951 = T43+T23+T13+T7+T5	725,760	V986 = T43+T22+T12+T8+T4	483,840
V952 = T43+T23+T13+T6+T4	161,280	V987 = T43+T22+T11+T7+T5	725,760
V953 = T43+T23+T11+T7+T4	161,280	V988 = T43+T22+T11+T7+T4	483,840
V954 = T43+T23+T11+T7+T5	241,920	V989 = T43+T22+T11+T6+T4	120,960
V955 = T43+T23+T11+T6+T4	40,320	V990 = T43+T20+T12+T7+T5	241,920
V956 = T43+T23+T14+T8+T5	201,600	V991 = T43+T20+T12+T7+T4	161,280
V957 = T43+T23+T14+T8+T4	201,600	V992 = T43+T20+T12+T6+T4	80,640
V958 = T43+T23+T14+T7+T5	604,800	V993 = T43+T20+T12+T8+T5	241,920
V959 = T43+T23+T14+T7+T4	403,200	V994 = T43+T20+T12+T8+T4	241,920
V960 = T43+T23+T14+T6+T4	201,600	V995 = T43+T20+T10+T6+T4	120,960
V961 = T43+T21+T10+T6+T4	40,320	V996 = T43+T20+T10+T7+T4	120,960
V962 = T43+T21+T10+T7+T4	40,320	V997 = T43+T20+T10+T7+T5	181,440
V963 = T43+T21+T10+T7+T5	60,480	V998 = T43+T20+T9+T6+T4	40,320
V964 = T43+T21+T9+T6+T4	40,320	V999 = T43+T20+T13+T8+T5	241,920
V965 = T43+T21+T11+T7+T4	161,280	V1000 = T43+T20+T13+T8+T4	241,920
V966 = T43+T21+T11+T7+T5	241,920	V1001 = T43+T20+T13+T7+T4	362,880
V967 = T43+T21+T11+T6+T4	40,320	V1002 = T43+T20+T13+T7+T5	544,320
V968 = T43+T21+T13+T7+T4	120,960	V1003 = T43+T20+T13+T6+T4	120,960
V969 = T43+T21+T13+T7+T5	181,440	V1004 = T43+T20+T11+T7+T4	161,280
V970 = T43+T21+T13+T6+T4	40,320	V1005 = T43+T20+T11+T7+T5	241,920
V971 = T43+T21+T13+T8+T5	80,640	V1006 = T43+T20+T11+T6+T4	40,320
V972 = T43+T21+T13+T8+T4	80,640	V1007 = T43+T17+T10+T6+T4	80,640
V973 = T43+T22+T13+T7+T5	1,088,640	V1008 = T43+T17+T10+T7+T4	80,640

Table A.23: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1009 = T43+T17+T10+T7+T5	120,960	V1044 = T44+T24+T9+T6+T4	161,280
V1010 = T43+T17+T9+T6+T4	40,320	V1045 = T44+T24+T14+T8+T5	403,200
V1011 = T43+T17+T11+T7+T4	161,280	V1046 = T44+T24+T14+T8+T4	403,200
V1012 = T43+T17+T11+T7+T5	241,920	V1047 = T44+T24+T14+T7+T5	1,209,600
V1013 = T43+T17+T11+T6+T4	40,320	V1048 = T44+T24+T14+T7+T4	806,400
V1014 = T43+T19+T12+T8+T5	60,480	V1049 = T44+T24+T14+T6+T4	403,200
V1015 = T43+T19+T12+T8+T4	60,480	V1050 = T44+T23+T12+T8+T5	725,760
V1016 = T43+T19+T12+T7+T4	40,320	V1051 = T44+T23+T12+T8+T4	725,760
V1017 = T43+T19+T12+T7+T5	60,480	V1052 = T44+T23+T12+T7+T4	483,840
V1018 = T43+T19+T12+T6+T4	20,160	V1053 = T44+T23+T12+T7+T5	725,760
V1019 = T43+T19+T10+T7+T4	40,320	V1054 = T44+T23+T12+T6+T4	241,920
V1020 = T43+T19+T10+T7+T5	60,480	V1055 = T44+T23+T10+T7+T4	483,840
V1021 = T43+T19+T10+T6+T4	40,320	V1056 = T44+T23+T10+T7+T5	725,760
V1022 = T43+T19+T9+T6+T4	20,160	V1057 = T44+T23+T10+T6+T4	483,840
V1023 = T43+T16+T10+T7+T4	40,320	V1058 = T44+T23+T13+T8+T5	1,935,360
V1024 = T43+T16+T10+T7+T5	60,480	V1059 = T44+T23+T13+T8+T4	1,935,360
V1025 = T43+T16+T10+T6+T4	40,320	V1060 = T44+T23+T13+T7+T4	2,903,040
V1026 = T43+T16+T9+T6+T4	40,320	V1061 = T44+T23+T13+T7+T5	4,354,560
V1027 = T43+T15+T9+T6+T4	20,160	V1062 = T44+T23+T13+T6+T4	967,680
V1028 = T44+T24+T13+T8+T5	1,290,240	V1063 = T44+T23+T11+T7+T4	967,680
V1029 = T44+T24+T13+T8+T4	1,290,240	V1064 = T44+T23+T11+T7+T5	1,451,520
V1030 = T44+T24+T13+T7+T5	2,903,040	V1065 = T44+T23+T11+T6+T4	241,920
V1031 = T44+T24+T13+T7+T4	1,935,360	V1066 = T44+T23+T9+T6+T4	241,920
V1032 = T44+T24+T13+T6+T4	645,120	V1067 = T44+T23+T14+T8+T5	1,209,600
V1033 = T44+T24+T12+T8+T5	967,680	V1068 = T44+T23+T14+T8+T4	1,209,600
V1034 = T44+T24+T12+T8+T4	967,680	V1069 = T44+T23+T14+T7+T5	3,628,800
V1035 = T44+T24+T12+T7+T4	645,120	V1070 = T44+T23+T14+T7+T4	2,419,200
V1036 = T44+T24+T12+T7+T5	967,680	V1071 = T44+T23+T14+T6+T4	1,209,600
V1037 = T44+T24+T12+T6+T4	322,560	V1072 = T44+T21+T10+T7+T4	120,960
V1038 = T44+T24+T10+T7+T4	483,840	V1073 = T44+T21+T10+T7+T5	181,440
V1039 = T44+T24+T10+T7+T5	725,760	V1074 = T44+T21+T10+T6+T4	120,960
V1040 = T44+T24+T10+T6+T4	483,840	V1075 = T44+T21+T11+T7+T4	483,840
V1041 = T44+T24+T11+T7+T4	967,680	V1076 = T44+T21+T11+T7+T5	725,760
V1042 = T44+T24+T11+T7+T5	1,451,520	V1077 = T44+T21+T11+T6+T4	120,960
V1043 = T44+T24+T11+T6+T4	241,920	V1078 = T44+T21+T9+T6+T4	120,960

Table A.24: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1079 = T44+T21+T13+T7+T4	362,880	V1114 = T44+T22+T12+T7+T4	806,400
V1080 = T44+T21+T13+T7+T5	544,320	V1115 = T44+T22+T12+T6+T4	403,200
V1081 = T44+T21+T13+T6+T4	120,960	V1116 = T44+T22+T12+T8+T5	1,209,600
V1082 = T44+T21+T13+T8+T5	241,920	V1117 = T44+T22+T12+T8+T4	1,209,600
V1083 = T44+T21+T13+T8+T4	241,920	V1118 = T44+T19+T10+T7+T5	241,920
V1084 = T44+T20+T11+T7+T4	806,400	V1119 = T44+T19+T10+T7+T4	161,280
V1085 = T44+T20+T11+T7+T5	1,209,600	V1120 = T44+T19+T10+T6+T4	161,280
V1086 = T44+T20+T11+T6+T4	201,600	V1121 = T44+T19+T9+T6+T4	80,640
V1087 = T44+T20+T10+T7+T5	907,200	V1122 = T44+T19+T12+T7+T5	241,920
V1088 = T44+T20+T10+T7+T4	604,800	V1123 = T44+T19+T12+T7+T4	161,280
V1089 = T44+T20+T10+T6+T4	604,800	V1124 = T44+T19+T12+T6+T4	80,640
V1090 = T44+T20+T9+T6+T4	201,600	V1125 = T44+T19+T12+T8+T5	241,920
V1091 = T44+T20+T13+T7+T4	1,814,400	V1126 = T44+T19+T12+T8+T4	241,920
V1092 = T44+T20+T13+T7+T5	2,721,600	V1127 = T44+T16+T9+T6+T4	80,640
V1093 = T44+T20+T13+T6+T4	604,800	V1128 = T44+T16+T10+T6+T4	80,640
V1094 = T44+T20+T13+T8+T5	1,209,600	V1129 = T44+T16+T10+T7+T4	80,640
V1095 = T44+T20+T13+T8+T4	1,209,600	V1130 = T44+T16+T10+T7+T5	120,960
V1096 = T44+T20+T12+T7+T5	1,209,600	V1131 = T44+T17+T10+T6+T4	241,920
V1097 = T44+T20+T12+T7+T4	806,400	V1132 = T44+T17+T10+T7+T4	241,920
V1098 = T44+T20+T12+T6+T4	403,200	V1133 = T44+T17+T10+T7+T5	362,880
V1099 = T44+T20+T12+T8+T5	1,209,600	V1134 = T44+T17+T11+T7+T4	483,840
V1100 = T44+T20+T12+T8+T4	1,209,600	V1135 = T44+T17+T11+T7+T5	725,760
V1101 = T44+T22+T11+T7+T5	1,814,400	V1136 = T44+T17+T11+T6+T4	120,960
V1102 = T44+T22+T11+T7+T4	1,209,600	V1137 = T44+T17+T9+T6+T4	120,960
V1103 = T44+T22+T11+T6+T4	302,400	V1138 = T44+T18+T11+T6+T4	60,480
V1104 = T44+T22+T9+T6+T4	100,800	V1139 = T44+T18+T11+T7+T4	241,920
V1105 = T44+T22+T10+T7+T5	604,800	V1140 = T44+T18+T11+T7+T5	362,880
V1106 = T44+T22+T10+T7+T4	403,200	V1141 = T44+T15+T9+T6+T4	20,160
V1107 = T44+T22+T10+T6+T4	403,200	V1142 = T45+T24+T14+T8+T5	100,800
V1108 = T44+T22+T13+T7+T5	2,721,600	V1143 = T45+T24+T14+T8+T4	100,800
V1109 = T44+T22+T13+T7+T4	1,814,400	V1144 = T45+T24+T14+T7+T5	302,400
V1110 = T44+T22+T13+T6+T4	604,800	V1145 = T45+T24+T14+T7+T4	201,600
V1111 = T44+T22+T13+T8+T5	1,209,600	V1146 = T45+T24+T14+T6+T4	100,800
V1112 = T44+T22+T13+T8+T4	1,209,600	V1147 = T45+T24+T13+T7+T5	725,760
V1113 = T44+T22+T12+T7+T5	1,209,600	V1148 = T45+T24+T13+T7+T4	483,840

Table A.25: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1149 = T45+T24+T13+T6+T4	161,280	V1184 = T45+T20+T12+T8+T5	241,920
V1150 = T45+T24+T13+T8+T5	322,560	V1185 = T45+T20+T12+T8+T4	241,920
V1151 = T45+T24+T13+T8+T4	322,560	V1186 = T45+T20+T10+T6+T4	120,960
V1152 = T45+T24+T12+T7+T5	241,920	V1187 = T45+T20+T10+T7+T4	120,960
V1153 = T45+T24+T12+T7+T4	161,280	V1188 = T45+T20+T10+T7+T5	181,440
V1154 = T45+T24+T12+T6+T4	80,640	V1189 = T45+T20+T13+T8+T5	241,920
V1155 = T45+T24+T12+T8+T5	241,920	V1190 = T45+T20+T13+T8+T4	241,920
V1156 = T45+T24+T12+T8+T4	241,920	V1191 = T45+T20+T13+T7+T4	362,880
V1157 = T45+T24+T10+T6+T4	120,960	V1192 = T45+T20+T13+T7+T5	544,320
V1158 = T45+T24+T10+T7+T4	120,960	V1193 = T45+T20+T13+T6+T4	120,960
V1159 = T45+T24+T10+T7+T5	181,440	V1194 = T45+T20+T11+T7+T4	161,280
V1160 = T45+T24+T11+T6+T4	60,480	V1195 = T45+T20+T11+T7+T5	241,920
V1161 = T45+T24+T11+T7+T4	241,920	V1196 = T45+T20+T11+T6+T4	40,320
V1162 = T45+T24+T11+T7+T5	362,880	V1197 = T45+T20+T9+T6+T4	40,320
V1163 = T45+T24+T9+T6+T4	40,320	V1198 = T45+T17+T10+T6+T4	161,280
V1164 = T45+T22+T13+T7+T5	1,088,640	V1199 = T45+T17+T10+T7+T4	161,280
V1165 = T45+T22+T13+T7+T4	725,760	V1200 = T45+T17+T10+T7+T5	241,920
V1166 = T45+T22+T13+T6+T4	241,920	V1201 = T45+T17+T11+T7+T4	322,560
V1167 = T45+T22+T13+T8+T5	483,840	V1202 = T45+T17+T11+T7+T5	483,840
V1168 = T45+T22+T13+T8+T4	483,840	V1203 = T45+T17+T11+T6+T4	80,640
V1169 = T45+T22+T12+T7+T5	483,840	V1204 = T45+T17+T9+T6+T4	80,640
V1170 = T45+T22+T12+T7+T4	322,560	V1205 = T45+T18+T11+T6+T4	60,480
V1171 = T45+T22+T12+T6+T4	161,280	V1206 = T45+T18+T11+T7+T4	241,920
V1172 = T45+T22+T12+T8+T5	483,840	V1207 = T45+T18+T11+T7+T5	362,880
V1173 = T45+T22+T12+T8+T4	483,840	V1208 = T45+T21+T13+T8+T5	80,640
V1174 = T45+T22+T10+T6+T4	161,280	V1209 = T45+T21+T13+T8+T4	80,640
V1175 = T45+T22+T10+T7+T4	161,280	V1210 = T45+T21+T13+T7+T4	120,960
V1176 = T45+T22+T10+T7+T5	241,920	V1211 = T45+T21+T13+T7+T5	181,440
V1177 = T45+T22+T11+T7+T4	483,840	V1212 = T45+T21+T13+T6+T4	40,320
V1178 = T45+T22+T11+T7+T5	725,760	V1213 = T45+T21+T11+T7+T4	161,280
V1179 = T45+T22+T11+T6+T4	120,960	V1214 = T45+T21+T11+T7+T5	241,920
V1180 = T45+T22+T9+T6+T4	40,320	V1215 = T45+T21+T11+T6+T4	40,320
V1181 = T45+T20+T12+T7+T5	241,920	V1216 = T45+T21+T10+T7+T5	60,480
V1182 = T45+T20+T12+T7+T4	161,280	V1217 = T45+T21+T10+T7+T4	40,320
V1183 = T45+T20+T12+T6+T4	80,640	V1218 = T45+T21+T10+T6+T4	40,320

Table A.26: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1219 = T45+T21+T9+T6+T4	40,320	V1254 = T46+T24+T10+T6+T4	846,720
V1220 = T45+T16+T10+T7+T5	60,480	V1255 = T46+T24+T12+T7+T5	1,693,440
V1221 = T45+T16+T10+T7+T4	40,320	V1256 = T46+T24+T12+T7+T4	1,128,960
V1222 = T45+T16+T10+T6+T4	40,320	V1257 = T46+T24+T12+T6+T4	564,480
V1223 = T45+T16+T9+T6+T4	40,320	V1258 = T46+T24+T12+T8+T5	1,693,440
V1224 = T45+T15+T9+T6+T4	40,320	V1259 = T46+T24+T12+T8+T4	1,693,440
V1225 = T46+T25+T14+T8+T5	302,400	V1260 = T46+T24+T11+T7+T5	2,540,160
V1226 = T46+T25+T14+T8+T4	302,400	V1261 = T46+T24+T11+T7+T4	1,693,440
V1227 = T46+T25+T14+T7+T5	907,200	V1262 = T46+T24+T11+T6+T4	423,360
V1228 = T46+T25+T14+T7+T4	604,800	V1263 = T46+T24+T9+T6+T4	282,240
V1229 = T46+T25+T14+T6+T4	302,400	V1264 = T46+T24+T14+T8+T5	705,600
V1230 = T46+T25+T13+T8+T5	1,209,600	V1265 = T46+T24+T14+T8+T4	705,600
V1231 = T46+T25+T13+T8+T4	1,209,600	V1266 = T46+T24+T14+T7+T5	2,116,800
V1232 = T46+T25+T13+T7+T5	2,721,600	V1267 = T46+T24+T14+T7+T4	1,411,200
V1233 = T46+T25+T13+T7+T4	1,814,400	V1268 = T46+T24+T14+T6+T4	705,600
V1234 = T46+T25+T13+T6+T4	604,800	V1269 = T46+T23+T12+T7+T5	483,840
V1235 = T46+T25+T12+T8+T5	907,200	V1270 = T46+T23+T12+T7+T4	322,560
V1236 = T46+T25+T12+T8+T4	907,200	V1271 = T46+T23+T12+T6+T4	161,280
V1237 = T46+T25+T12+T7+T4	604,800	V1272 = T46+T23+T12+T8+T5	483,840
V1238 = T46+T25+T12+T7+T5	907,200	V1273 = T46+T23+T12+T8+T4	483,840
V1239 = T46+T25+T12+T6+T4	302,400	V1274 = T46+T23+T10+T6+T4	322,560
V1240 = T46+T25+T10+T7+T4	604,800	V1275 = T46+T23+T10+T7+T4	322,560
V1241 = T46+T25+T10+T7+T5	907,200	V1276 = T46+T23+T10+T7+T5	483,840
V1242 = T46+T25+T10+T6+T4	604,800	V1277 = T46+T23+T9+T6+T4	161,280
V1243 = T46+T25+T11+T7+T5	1,814,400	V1278 = T46+T23+T13+T8+T5	1,290,240
V1244 = T46+T25+T11+T7+T4	1,209,600	V1279 = T46+T23+T13+T8+T4	1,290,240
V1245 = T46+T25+T11+T6+T4	302,400	V1280 = T46+T23+T13+T7+T4	1,935,360
V1246 = T46+T25+T9+T6+T4	302,400	V1281 = T46+T23+T13+T7+T5	2,903,040
V1247 = T46+T24+T13+T7+T5	5,080,320	V1282 = T46+T23+T13+T6+T4	645,120
V1248 = T46+T24+T13+T7+T4	3,386,880	V1283 = T46+T23+T11+T7+T4	645,120
V1249 = T46+T24+T13+T6+T4	1,128,960	V1284 = T46+T23+T11+T7+T5	967,680
V1250 = T46+T24+T13+T8+T5	2,257,920	V1285 = T46+T23+T11+T6+T4	161,280
V1251 = T46+T24+T13+T8+T4	2,257,920	V1286 = T46+T23+T14+T8+T5	806,400
V1252 = T46+T24+T10+T7+T4	846,720	V1287 = T46+T23+T14+T8+T4	806,400
V1253 = T46+T24+T10+T7+T5	1,270,080	V1288 = T46+T23+T14+T7+T5	2,419,200

Table A.27: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1289 = T46+T23+T14+T7+T4	1,612,800	V1324 = T46+T20+T10+T7+T4	241,920
V1290 = T46+T23+T14+T6+T4	806,400	V1325 = T46+T20+T10+T6+T4	241,920
V1291 = T46+T21+T10+T6+T4	161,280	V1326 = T46+T20+T9+T6+T4	80,640
V1292 = T46+T21+T10+T7+T4	161,280	V1327 = T46+T20+T13+T7+T4	725,760
V1293 = T46+T21+T10+T7+T5	241,920	V1328 = T46+T20+T13+T7+T5	1,088,640
V1294 = T46+T21+T9+T6+T4	161,280	V1329 = T46+T20+T13+T6+T4	241,920
V1295 = T46+T21+T11+T7+T4	645,120	V1330 = T46+T20+T13+T8+T5	483,840
V1296 = T46+T21+T11+T7+T5	967,680	V1331 = T46+T20+T13+T8+T4	483,840
V1297 = T46+T21+T11+T6+T4	161,280	V1332 = T46+T20+T12+T7+T5	483,840
V1298 = T46+T21+T13+T7+T4	483,840	V1333 = T46+T20+T12+T7+T4	322,560
V1299 = T46+T21+T13+T7+T5	725,760	V1334 = T46+T20+T12+T6+T4	161,280
V1300 = T46+T21+T13+T6+T4	161,280	V1335 = T46+T20+T12+T8+T5	483,840
V1301 = T46+T21+T13+T8+T5	322,560	V1336 = T46+T20+T12+T8+T4	483,840
V1302 = T46+T21+T13+T8+T4	322,560	V1337 = T46+T19+T10+T7+T5	120,960
V1303 = T46+T22+T11+T6+T4	181,440	V1338 = T46+T19+T10+T7+T4	80,640
V1304 = T46+T22+T11+T7+T4	725,760	V1339 = T46+T19+T10+T6+T4	80,640
V1305 = T46+T22+T11+T7+T5	1,088,640	V1340 = T46+T19+T9+T6+T4	40,320
V1306 = T46+T22+T9+T6+T4	60,480	V1341 = T46+T19+T12+T7+T5	120,960
V1307 = T46+T22+T10+T7+T5	362,880	V1342 = T46+T19+T12+T7+T4	80,640
V1308 = T46+T22+T10+T7+T4	241,920	V1343 = T46+T19+T12+T6+T4	40,320
V1309 = T46+T22+T10+T6+T4	241,920	V1344 = T46+T19+T12+T8+T5	120,960
V1310 = T46+T22+T13+T7+T4	1,088,640	V1345 = T46+T19+T12+T8+T4	120,960
V1311 = T46+T22+T13+T7+T5	1,632,960	V1346 = T46+T16+T9+T6+T4	80,640
V1312 = T46+T22+T13+T6+T4	362,880	V1347 = T46+T16+T10+T6+T4	80,640
V1313 = T46+T22+T13+T8+T5	725,760	V1348 = T46+T16+T10+T7+T4	80,640
V1314 = T46+T22+T13+T8+T4	725,760	V1349 = T46+T16+T10+T7+T5	120,960
V1315 = T46+T22+T12+T6+T4	241,920	V1350 = T46+T17+T10+T7+T4	161,280
V1316 = T46+T22+T12+T7+T5	725,760	V1351 = T46+T17+T10+T7+T5	241,920
V1317 = T46+T22+T12+T7+T4	483,840	V1352 = T46+T17+T10+T6+T4	161,280
V1318 = T46+T22+T12+T8+T5	725,760	V1353 = T46+T17+T9+T6+T4	80,640
V1319 = T46+T22+T12+T8+T4	725,760	V1354 = T46+T17+T11+T7+T4	322,560
V1320 = T46+T20+T11+T7+T4	322,560	V1355 = T46+T17+T11+T7+T5	483,840
V1321 = T46+T20+T11+T7+T5	483,840	V1356 = T46+T17+T11+T6+T4	80,640
V1322 = T46+T20+T11+T6+T4	80,640	V1357 = T46+T18+T11+T7+T5	362,880
V1323 = T46+T20+T10+T7+T5	362,880	V1358 = T46+T18+T11+T7+T4	241,920

Table A.28: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1359 = T46+T18+T11+T6+T4	60,480	V1394 = T47+T24+T10+T7+T5	907,200
V1360 = T46+T15+T9+T6+T4	40,320	V1395 = T47+T24+T10+T6+T4	604,800
V1361 = T47+T25+T14+T8+T5	120,960	V1396 = T47+T24+T11+T7+T4	1,209,600
V1362 = T47+T25+T14+T8+T4	120,960	V1397 = T47+T24+T11+T7+T5	1,814,400
V1363 = T47+T25+T14+T7+T5	362,880	V1398 = T47+T24+T11+T6+T4	302,400
V1364 = T47+T25+T14+T7+T4	241,920	V1399 = T47+T24+T9+T6+T4	201,600
V1365 = T47+T25+T14+T6+T4	120,960	V1400 = T47+T24+T14+T8+T5	504,000
V1366 = T47+T25+T13+T8+T5	483,840	V1401 = T47+T24+T14+T8+T4	504,000
V1367 = T47+T25+T13+T8+T4	483,840	V1402 = T47+T24+T14+T7+T5	1,512,000
V1368 = T47+T25+T13+T7+T5	1,088,640	V1403 = T47+T24+T14+T7+T4	1,008,000
V1369 = T47+T25+T13+T7+T4	725,760	V1404 = T47+T24+T14+T6+T4	504,000
V1370 = T47+T25+T13+T6+T4	241,920	V1405 = T47+T23+T12+T8+T5	604,800
V1371 = T47+T25+T12+T8+T5	362,880	V1406 = T47+T23+T12+T8+T4	604,800
V1372 = T47+T25+T12+T8+T4	362,880	V1407 = T47+T23+T12+T7+T4	403,200
V1373 = T47+T25+T12+T7+T4	241,920	V1408 = T47+T23+T12+T7+T5	604,800
V1374 = T47+T25+T12+T7+T5	362,880	V1409 = T47+T23+T12+T6+T4	201,600
V1375 = T47+T25+T12+T6+T4	120,960	V1410 = T47+T23+T10+T7+T4	403,200
V1376 = T47+T25+T10+T7+T4	241,920	V1411 = T47+T23+T10+T7+T5	604,800
V1377 = T47+T25+T10+T7+T5	362,880	V1412 = T47+T23+T10+T6+T4	403,200
V1378 = T47+T25+T10+T6+T4	241,920	V1413 = T47+T23+T13+T8+T5	1,612,800
V1379 = T47+T25+T11+T7+T5	725,760	V1414 = T47+T23+T13+T8+T4	1,612,800
V1380 = T47+T25+T11+T7+T4	483,840	V1415 = T47+T23+T13+T7+T4	2,419,200
V1381 = T47+T25+T11+T6+T4	120,960	V1416 = T47+T23+T13+T7+T5	3,628,800
V1382 = T47+T25+T9+T6+T4	120,960	V1417 = T47+T23+T13+T6+T4	806,400
V1383 = T47+T24+T13+T8+T5	1,612,800	V1418 = T47+T23+T11+T7+T4	806,400
V1384 = T47+T24+T13+T8+T4	1,612,800	V1419 = T47+T23+T11+T7+T5	1,209,600
V1385 = T47+T24+T13+T7+T5	3,628,800	V1420 = T47+T23+T11+T6+T4	201,600
V1386 = T47+T24+T13+T7+T4	2,419,200	V1421 = T47+T23+T9+T6+T4	201,600
V1387 = T47+T24+T13+T6+T4	806,400	V1422 = T47+T23+T14+T8+T5	1,008,000
V1388 = T47+T24+T12+T8+T5	1,209,600	V1423 = T47+T23+T14+T8+T4	1,008,000
V1389 = T47+T24+T12+T8+T4	1,209,600	V1424 = T47+T23+T14+T7+T5	3,024,000
V1390 = T47+T24+T12+T7+T4	806,400	V1425 = T47+T23+T14+T7+T4	2,016,000
V1391 = T47+T24+T12+T7+T5	1,209,600	V1426 = T47+T23+T14+T6+T4	1,008,000
V1392 = T47+T24+T12+T6+T4	403,200	V1427 = T47+T21+T10+T7+T4	161,280
V1393 = T47+T24+T10+T7+T4	604,800	V1428 = T47+T21+T10+T7+T5	241,920

Table A.29: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1429 = T47+T21+T10+T6+T4	161,280	V1464 = T47+T20+T13+T7+T5	2,177,280
V1430 = T47+T21+T11+T7+T4	645,120	V1465 = T47+T20+T13+T6+T4	483,840
V1431 = T47+T21+T11+T7+T5	967,680	V1466 = T47+T20+T13+T8+T5	967,680
V1432 = T47+T21+T11+T6+T4	161,280	V1467 = T47+T20+T13+T8+T4	967,680
V1433 = T47+T21+T9+T6+T4	161,280	V1468 = T47+T20+T12+T7+T5	967,680
V1434 = T47+T21+T13+T7+T4	483,840	V1469 = T47+T20+T12+T7+T4	645,120
V1435 = T47+T21+T13+T7+T5	725,760	V1470 = T47+T20+T12+T6+T4	322,560
V1436 = T47+T21+T13+T6+T4	161,280	V1471 = T47+T20+T12+T8+T5	967,680
V1437 = T47+T21+T13+T8+T5	322,560	V1472 = T47+T20+T12+T8+T4	967,680
V1438 = T47+T21+T13+T8+T4	322,560	V1473 = T47+T19+T10+T7+T5	241,920
V1439 = T47+T22+T11+T7+T5	1,451,520	V1474 = T47+T19+T10+T7+T4	161,280
V1440 = T47+T22+T11+T7+T4	967,680	V1475 = T47+T19+T10+T6+T4	161,280
V1441 = T47+T22+T11+T6+T4	241,920	V1476 = T47+T19+T9+T6+T4	80,640
V1442 = T47+T22+T9+T6+T4	80,640	V1477 = T47+T19+T12+T7+T5	241,920
V1443 = T47+T22+T10+T7+T5	483,840	V1478 = T47+T19+T12+T7+T4	161,280
V1444 = T47+T22+T10+T7+T4	322,560	V1479 = T47+T19+T12+T6+T4	80,640
V1445 = T47+T22+T10+T6+T4	322,560	V1480 = T47+T19+T12+T8+T5	241,920
V1446 = T47+T22+T13+T7+T5	2,177,280	V1481 = T47+T19+T12+T8+T4	241,920
V1447 = T47+T22+T13+T7+T4	1,451,520	V1482 = T47+T16+T9+T6+T4	120,960
V1448 = T47+T22+T13+T6+T4	483,840	V1483 = T47+T16+T10+T6+T4	120,960
V1449 = T47+T22+T13+T8+T5	967,680	V1484 = T47+T16+T10+T7+T4	120,960
V1450 = T47+T22+T13+T8+T4	967,680	V1485 = T47+T16+T10+T7+T5	181,440
V1451 = T47+T22+T12+T7+T5	967,680	V1486 = T47+T17+T9+T6+T4	120,960
V1452 = T47+T22+T12+T7+T4	645,120	V1487 = T47+T17+T10+T6+T4	241,920
V1453 = T47+T22+T12+T6+T4	322,560	V1488 = T47+T17+T10+T7+T4	241,920
V1454 = T47+T22+T12+T8+T5	967,680	V1489 = T47+T17+T10+T7+T5	362,880
V1455 = T47+T22+T12+T8+T4	967,680	V1490 = T47+T17+T11+T6+T4	120,960
V1456 = T47+T20+T11+T7+T4	645,120	V1491 = T47+T17+T11+T7+T4	483,840
V1457 = T47+T20+T11+T7+T5	967,680	V1492 = T47+T17+T11+T7+T5	725,760
V1458 = T47+T20+T11+T6+T4	161,280	V1493 = T47+T18+T11+T6+T4	60,480
V1459 = T47+T20+T10+T7+T5	725,760	V1494 = T47+T18+T11+T7+T4	241,920
V1460 = T47+T20+T10+T7+T4	483,840	V1495 = T47+T18+T11+T7+T5	362,880
V1461 = T47+T20+T10+T6+T4	483,840	V1496 = T47+T15+T9+T6+T4	40,320
V1462 = T47+T20+T9+T6+T4	161,280	V1497 = T48+T25+T14+T8+T5	20,160
V1463 = T47+T20+T13+T7+T4	1,451,520	V1498 = T48+T25+T14+T8+T4	20,160

Table A.30: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1499 = T48+T25+T14+T7+T5	60,480	V1534 = T48+T24+T11+T6+T4	60,480
V1500 = T48+T25+T14+T7+T4	40,320	V1535 = T48+T24+T9+T6+T4	40,320
V1501 = T48+T25+T14+T6+T4	20,160	V1536 = T48+T24+T14+T8+T5	100,800
V1502 = T48+T25+T13+T8+T5	80,640	V1537 = T48+T24+T14+T8+T4	100,800
V1503 = T48+T25+T13+T8+T4	80,640	V1538 = T48+T24+T14+T7+T5	302,400
V1504 = T48+T25+T13+T7+T5	181,440	V1539 = T48+T24+T14+T7+T4	201,600
V1505 = T48+T25+T13+T7+T4	120,960	V1540 = T48+T24+T14+T6+T4	100,800
V1506 = T48+T25+T13+T6+T4	40,320	V1541 = T48+T23+T12+T8+T5	120,960
V1507 = T48+T25+T12+T8+T5	60,480	V1542 = T48+T23+T12+T8+T4	120,960
V1508 = T48+T25+T12+T8+T4	60,480	V1543 = T48+T23+T12+T7+T4	80,640
V1509 = T48+T25+T12+T7+T4	40,320	V1544 = T48+T23+T12+T7+T5	120,960
V1510 = T48+T25+T12+T7+T5	60,480	V1545 = T48+T23+T12+T6+T4	40,320
V1511 = T48+T25+T12+T6+T4	20,160	V1546 = T48+T23+T10+T7+T4	80,640
V1512 = T48+T25+T10+T7+T4	40,320	V1547 = T48+T23+T10+T7+T5	120,960
V1513 = T48+T25+T10+T7+T5	60,480	V1548 = T48+T23+T10+T6+T4	80,640
V1514 = T48+T25+T10+T6+T4	40,320	V1549 = T48+T23+T13+T8+T5	322,560
V1515 = T48+T25+T11+T7+T5	120,960	V1550 = T48+T23+T13+T8+T4	322,560
V1516 = T48+T25+T11+T7+T4	80,640	V1551 = T48+T23+T13+T7+T4	483,840
V1517 = T48+T25+T11+T6+T4	20,160	V1552 = T48+T23+T13+T7+T5	725,760
V1518 = T48+T25+T9+T6+T4	20,160	V1553 = T48+T23+T13+T6+T4	161,280
V1519 = T48+T24+T13+T8+T5	322,560	V1554 = T48+T23+T11+T7+T4	161,280
V1520 = T48+T24+T13+T8+T4	322,560	V1555 = T48+T23+T11+T7+T5	241,920
V1521 = T48+T24+T13+T7+T5	725,760	V1556 = T48+T23+T11+T6+T4	40,320
V1522 = T48+T24+T13+T7+T4	483,840	V1557 = T48+T23+T9+T6+T4	40,320
V1523 = T48+T24+T13+T6+T4	161,280	V1558 = T48+T23+T14+T8+T5	201,600
V1524 = T48+T24+T12+T8+T5	241,920	V1559 = T48+T23+T14+T8+T4	201,600
V1525 = T48+T24+T12+T8+T4	241,920	V1560 = T48+T23+T14+T7+T5	604,800
V1526 = T48+T24+T12+T7+T4	161,280	V1561 = T48+T23+T14+T7+T4	403,200
V1527 = T48+T24+T12+T7+T5	241,920	V1562 = T48+T23+T14+T6+T4	201,600
V1528 = T48+T24+T12+T6+T4	80,640	V1563 = T48+T21+T10+T7+T4	40,320
V1529 = T48+T24+T10+T7+T4	120,960	V1564 = T48+T21+T10+T7+T5	60,480
V1530 = T48+T24+T10+T7+T5	181,440	V1565 = T48+T21+T10+T6+T4	40,320
V1531 = T48+T24+T10+T6+T4	120,960	V1566 = T48+T21+T11+T7+T4	161,280
V1532 = T48+T24+T11+T7+T4	241,920	V1567 = T48+T21+T11+T7+T5	241,920
V1533 = T48+T24+T11+T7+T5	362,880	V1568 = T48+T21+T11+T6+T4	40,320

Table A.31: Tree-equivalent regular vines with 8 nodes (Continuation).

Tree sequence & # Tree-equivalent Labeled Regular Vines		Tree sequence & # Tree-equivalent Labeled Regular Vines	
V1569 = T48+T21+T9+T6+T4	40,320	V1604 = T48+T20+T12+T7+T5	241,920
V1570 = T48+T21+T13+T7+T4	120,960	V1605 = T48+T20+T12+T7+T4	161,280
V1571 = T48+T21+T13+T7+T5	181,440	V1606 = T48+T20+T12+T6+T4	80,640
V1572 = T48+T21+T13+T6+T4	40,320	V1607 = T48+T20+T12+T8+T5	241,920
V1573 = T48+T21+T13+T8+T5	80,640	V1608 = T48+T20+T12+T8+T4	241,920
V1574 = T48+T21+T13+T8+T4	80,640	V1609 = T48+T19+T10+T7+T5	60,480
V1575 = T48+T22+T11+T7+T5	362,880	V1610 = T48+T19+T10+T7+T4	40,320
V1576 = T48+T22+T11+T7+T4	241,920	V1611 = T48+T19+T10+T6+T4	40,320
V1577 = T48+T22+T11+T6+T4	60,480	V1612 = T48+T19+T9+T6+T4	20,160
V1578 = T48+T22+T9+T6+T4	20,160	V1613 = T48+T19+T12+T7+T5	60,480
V1579 = T48+T22+T10+T7+T5	120,960	V1614 = T48+T19+T12+T7+T4	40,320
V1580 = T48+T22+T10+T7+T4	80,640	V1615 = T48+T19+T12+T6+T4	20,160
V1581 = T48+T22+T10+T6+T4	80,640	V1616 = T48+T19+T12+T8+T5	60,480
V1582 = T48+T22+T13+T7+T5	544,320	V1617 = T48+T19+T12+T8+T4	60,480
V1583 = T48+T22+T13+T7+T4	362,880	V1618 = T48+T16+T9+T6+T4	40,320
V1584 = T48+T22+T13+T6+T4	120,960	V1619 = T48+T16+T10+T6+T4	40,320
V1585 = T48+T22+T13+T8+T5	241,920	V1620 = T48+T16+T10+T7+T4	40,320
V1586 = T48+T22+T13+T8+T4	241,920	V1621 = T48+T16+T10+T7+T5	60,480
V1587 = T48+T22+T12+T7+T5	241,920	V1622 = T48+T17+T9+T6+T4	40,320
V1588 = T48+T22+T12+T7+T4	161,280	V1623 = T48+T17+T10+T6+T4	80,640
V1589 = T48+T22+T12+T6+T4	80,640	V1624 = T48+T17+T10+T7+T4	80,640
V1590 = T48+T22+T12+T8+T5	241,920	V1625 = T48+T17+T10+T7+T5	120,960
V1591 = T48+T22+T12+T8+T4	241,920	V1626 = T48+T17+T11+T6+T4	40,320
V1592 = T48+T20+T11+T7+T4	161,280	V1627 = T48+T17+T11+T7+T4	161,280
V1593 = T48+T20+T11+T7+T5	241,920	V1628 = T48+T17+T11+T7+T5	241,920
V1594 = T48+T20+T11+T6+T4	40,320	V1629 = T48+T18+T11+T6+T4	20,160
V1595 = T48+T20+T10+T7+T5	181,440	V1630 = T48+T18+T11+T7+T4	80,640
V1596 = T48+T20+T10+T7+T4	120,960	V1631 = T48+T18+T11+T7+T5	120,960
V1597 = T48+T20+T10+T6+T4	120,960	V1632 = T48+T15+T9+T6+T4	20,160
V1598 = T48+T20+T9+T6+T4	40,320		
V1599 = T48+T20+T13+T7+T4	362,880		
V1600 = T48+T20+T13+T7+T5	544,320		
V1601 = T48+T20+T13+T6+T4	120,960		
V1602 = T48+T20+T13+T8+T5	241,920		
V1603 = T48+T20+T13+T8+T4	241,920		

Table A.32: *Tree-equivalent regular vines with 8 nodes (Continuation).*

Summary

Bayesian Belief Nets and Vines *in aviation safety and other applications*

Oswaldo Morales Nápoles.

The relationship between probability theory and graph theory has enabled great improvements for both. In particular, probability theory has benefitted from graph theory for the representation of multivariate distributions. Two such possible representations are Bayesian Belief Nets (BBNs) and Vines. The main interest of this thesis is in the description of applications of BBNs and vines to problems in which quantifying uncertainty is of prime importance. Two such problems studied in this thesis are in the identification and measurement of risks in the aviation industry and earth dams.

Vines and BBNs are closely related. Despite of this fact, graphical properties of BBNs have been more studied than vines. Before presenting applications of BBNs and vines in aviation and earth dam safety, this thesis presents a study of vines as graphs. This study (chapter 2) represents a first step towards a more systematic approach to studying vines as graphs.

The largest part of the thesis is concerned with aviation safety. The aviation sector is generally recognized for its high levels of safety. In fact the fatal and non-fatal accident rate worldwide has not reached the levels of 1996 in any year in the period from 1997 to 2007. However, the total number of flights for the same period has grown from about 31 million in 1996 to about 48 in 2007. If this growth trend continues, the accident rate must decrease further in order to maintain low the total number of accidents.

Different studies have shown that accidents in the aviation industry have humans as their main contributing factor. A model that aspires to improve safety in the aviation industry should include human error and all its other components. The Netherlands Ministry of Transport and Water Management commissioned the construction of a model for comparing alternatives, strengthening safety measures

and finding causes of incidents and accidents and for quantifying the probability of adverse events in the aviation system. This model is known as the Causal Model for Air Transport Safety or CATS.

The CATS model is in fact a BBN that consists of 1,504 nodes and 4,976 arcs. The construction of such a model was a great challenge that was undertaken with the effort of many people. The major focus of this thesis is in the description of the quantification of such a BBN. Emphasis is placed in the techniques used for the quantification of dependence for Non Parametric Continuous-Discrete BBNs. For the CATS model this was done mainly through the use of structured expert judgment in human error models for flight crew, air traffic control and maintenance crew.

BBNs are tools flexible enough to be used in different fields. This is shown in the 6th chapter of this thesis where a BBN for earth dams safety in the State of Mexico is presented. From the end of October and up to the end of November 2007 flooding was observed in about 70% of the Tabasco flatland affecting more than 1 million people. The model described in chapter 6 may be of assistance for earth dam engineers in the State of Mexico in preventing situations such as the one observed in Tabasco.

The main conclusions of these thesis summarized in chapter 7 concern vines, BBNs and the applications discussed. For vines the need to study them in a systematic way as graphical structures is discussed. Taking this step could help increase its range of applications. BBNs have proved in this thesis to be a powerful tool for risk and uncertainty analysis. Methods successfully used in practice to quantify them from experts have been advanced. Further research in combination of individual experts' dependence estimates is suggested.

With regards to applications, according to the CATS BBN, experienced crews and newer aircrafts lead to a greater reduction of the accident rate than measures concerning maintenance technicians or air traffic controllers. This does not entail that investments in air traffic control and maintenance crew are discouraged. It does suggest however that investments pointing towards experienced crew and aircraft fleet renewal might deserve priority. Finally with respect to earth dam safety in the State of Mexico this thesis shows the experts' belief that given a dam failure, the economic consequences are approximately constant regardless of the size of a flooding. Thus, the occurrence of a failure should be avoided if costs are to remain minimum.

Samenvatting

Bayesiaanse Netwerken en Vines in veiligheid van de luchtvaart en andere toepassingen

Oswaldo Morales Nápoles.

De relatie tussen de kansrekening en grafentheorie heeft grote verbeteringen voor beide mogelijk gemaakt. In het bijzonder, heeft kansrekening profiteerd van de grafentheorie voor de vertegenwoordiging van hoog dimensionale verdelingen. Twee van zulke mogelijke voorstellingen zijn Bayesiaanse Netwerken (BBN) en Vines. Het voornaamste belang van dit proefschrift is de beschrijving van toepassingen van BBNs en Vines in problemen waarin het kwantificeren van onzekerheid van het grootste betekenis is. Twee van dergelijke problemen onderzocht in dit proefschrift zijn de identificatie en meting van risico's in de luchtvaartindustrie en bij dijken.

Vines en BBN's hebben een nauwe relatie. Ondanks van dit feit, zijn grafische eigenschappen van BBN's meer onderzocht dan Vines. Voor het behandelen van BBNs en Vines in de luchtvaart en dijk veiligheid, presenteert dit proefschrift een studie van Vines als grafieken. Deze studie (hoofdstuk 2) vormt een eerste stap naar een meer systematische aanpak van het bestuderen van vines als grafieken.

Het grootste deel van het proefschrift houdt zich bezig met veiligheid in de luchtvaart. De luchtvaartsector is bekend oom zijn hoge mate van veiligheid. In feite, heeft in de periode van 1997 tot 2007 het aantal fatale en niet-fatale ongevallen cijfer per jaar wereldwijd niet de hoogte bereikt van 1996. Echter, het totale aantal vluchten is in dezelfde periode gegroeid van ongeveer 31 miljoen in 1996 tot ongeveer 48 in 2007. Als deze groei trend doorzet, moet het ongevallen cijfer verder afnemen om het totale aantal ongevallen laag te houden.

Verschillende studies hebben aangetoond dat ongevallen in de luchtvaartindustrie menselijk factoren als belangrijkste oorzak hebben. Een model dat streeft naar verbetering van de veiligheid in de luchtvaartsector moet ook menselijke fouten en alle andere componenten bevatten opnemen. De Nederlandse Ministerie

van Verkeer en Waterstaat heeft opdracht gegeven voor de constructie van een model voor het vergelijken van alternatieven, aanscherping van de veiligheidsmaatregelen, en het vinden van oorzaken van incidenten en ongevallen en voor het kwantificeren van de kans op ongewenste voorvallen in de luchtvaart. Dit model staat bekend als het Causal Model for Air Transport Safety of CATS.

Het CATS-model is in feite een BBN dat bestaat uit 1.504 knooppunten en 4.976 bogen. De bouw van een dergelijk model is een grote uitdaging die werd gerealiseerd door de inspanning van velen. De belangrijkste focus van dit proefschrift ligt op de beschrijving van de kwantificering van een dergelijke BBN. De nadruk wordt gelegd op de technieken gebruikt voor de kwantificering van de afhankelijkheid voor niet-parametrische continue-discrete BBN's. Voor het CATS-model was dit vooral gedaan door het gebruik van gestructureerde expert mening in menselijke fout modellen voor de cockpitpersoneel, de luchtverkeersleiding en onderhoud technici.

BBNs zijn tools die flexibel genoeg zijn om gebruikt te worden in verschillende gebieden. Dit wordt weergegeven in het 6de hoofdstuk van dit proefschrift waar een BBN voor dijken veiligheid in de Staat van Mexico wordt gepresenteerd. Vanaf eind oktober tot eind november 2007 werden er overstromingen waargenomen in ongeveer 70% van de lage gebied van Tabasco. Dit heb meer dan 1 miljoen mensen getroffen. Het model beschreven in hoofdstuk 6 kan worden van een hulpmiddel voor civile technische ingenieurs in de Staat van Mexco bij het voorkomen van situaties zoals die zijn waargenomen in Tabasco.

De belangrijkste conclusies van dit proefschrift samengevat in hoofdstuk 7 betreffen Vines, BBN's en de besproken toepassingen. De noodzaak om Vines te bestuderen op een systematische manier als grafische structuren wordt besproken. Het nemen van deze stap zou de mogelijke toepassing kunnen helpen vergroten. BBNs hebben bewezen in dit proefschrift een krachtig instrument voor risico en onzekerheidsanalyse te zijn. Methoden die met succes in de praktijk zijn gebruikt om ze te kwantificeren van experts zijn geavanceerd. Verder onderzoek over de combinatie van de individueel afhankelijkheid beoordelen van experts is gesuggereerd.

Met betrekking tot de toepassingen, zouden volgens de CATS BBN, ervaren cockpitpersoneel en nieuwere vliegtuigen het ongevallen cijfer meer verminderen dan maatregelen met betrekking tot de onderhoud technici of luchtverkeersleiders. Dit houdt niet in dat investeringen in de luchtverkeersleiding en onderhoudbemanning worden afgeraden. Er wordt echter gesuggereerd dat de investeringen die gericht zij op ervaren bemanning en vliegtuig vernieuwing van de vloot prioriteit zou kunnen verdienen. Tenslotte toont dit proefschrift, met betrekking tot de veiligheid van de dijken in de Staat van Mexico aan, dat expert geloven dat de economische gevolgen ongeveer constant zijn, ongeacht de grootte van een overstroming. Daarom moet er overstromingen worden voorkomen zodat de kosten minimaal blijven.

Acknowledgments

I have lived now in the Netherlands for more than seven years. These have been among the most edifying in my life. For giving me the opportunity to come to the Netherlands initially as a master student, when in Mexico many doors were being closed, I must thank prof. dr. Roger M. Cooke. Since very young, I set for myself the goal of pursuing a PhD degree. For giving me the opportunity to go ahead in this quest I must thank again Roger. His advise, not only with regards to this thesis has meant for me a lot more than supervision. His example has influenced greatly and will continue to do so in the future different aspects of my life. I send my most warm and sincere thanks to Roger for my seven years at the TU Delft.

I must also acknowledge dr. Dorota Kurowicka for the long and fruitful discussions that we held during approximately five years with the purpose of improving some of the work leading to this thesis. I am very grateful to all members of the PhD committee for their suggestions, comments and ideas that have greatly contributed to this thesis. In particular I want to thank Prof. Harry Joe for his many ideas patiently communicated to me regarding equivalence classes of regular vines. This is a subject of great interest to me and I hope to have the opportunity to follow his and others contributions in this topic in the future. Prof. Ben Ale showed great leadership in the development of the CATS model. Contributions from professors Mosleh and van Tooren as part of the expert evaluation group of the CATS model were not only beneficial for the CATS consortium but also for this and surly other theses. Dr. De Leon made important remarks for the chapter on earth dam safety in the State of Mexico. I can only hope for future collaboration with his group in Toluca.

I thank Cindy Bosman and Carl Schneider for making it easier for me during seven years to deal with different needs at the TU Delft. I also thank CICAT for handling my four year term as a PhD student at the TU Delft. In particular Theda Olsder and Veronique van der Varst provided great help in dealing with visum and insurance related issues.

During the development of the CATS model I had the pleasure to work with master students for their graduation project. Collaboration with Tina Singuran for the coupling of the flight crew and air traffic control models resulted in important changes in the model. Iwona Jagielska contributed to the development of the air traffic control model and of a software tool for the elicitation of rank and

conditional rank correlations from experts. This tool was also improved further by Kasia Jóźwiak. This tool was not used in this thesis, however their contributions are important for the future development of a stand alone tool for the elicitation of rank and conditional rank correlations from experts. Kasia Krugła is responsible for about half of the massive and tedious job of putting the the CATS BBN into UNINET. Her contributions in the air traffic control model are also very valuable.

Collaboration with colleagues for the applications presented in this thesis is invaluable. For the CATS model the ideas shared with Alfred Roelen (NLR), Gerben van Baren (NLR) and John Spouge (DNV) is greatly appreciated. My appreciation also for the rest of the members in the CATS consortium. Work done together with David Joaquín Delgado Hernández has been fruitful and I hope to keep sharing professional challenges with him. Daniel Lewandowski took over my duties with the CATS model by the end of the project. That made it easier for me to finish writing this thesis and for that I am grateful. My gratitude goes also to all my fellow PhD students, master students and colleagues at the Delf Institute of Applied Mathematics with whom I shared great moments during my studies.

I have been lucky enough as to make a good number of good friends in Delft. To all of them thanks for sharing the better and the worse moments during this process with me. Despite the distance I know that my friends in Mexico and the US had me in their thoughts and that was of great help for finalizing the thesis.

My family has continuously been attempt to my personal and professional development and for that I am deeply grateful. Particular thanks go to dr. Samuel Morales Sales for helping me develop an interest in research and science. Special thanks to my mother Lilia Nápoles Cruz, my father Osvaldo Morales Sales and brother Luis Antonio Morales Nápoles. Without them I could have never reached the stage of defending my thesis. I have learned to enjoy life in a different way thanks to Camila. She is also a very important source of motivation for me. This thesis is as much mine as Sandra's I cannot express with words enough appreciation for all her contributions and support in all ingredients required to complete this effort.

Curriculum Vitae

Oswaldo Morales Nápoles was born in Toluca Estado de México, México on February the 21st 1977. He graduated with a Bachelor's degree on Economics from the Autonomous University of the State of Mexico on 2000 with the thesis: *Economic value of the environmental quality in the Alto Lerma*.

In 2002 he moved to the Netherlands to continue his studies in the Master's program Risk and Environmental Modelling offered at the Delft University of Technology. He graduated in 2004 with an M.Sc. degree on Applied Mathematics with the thesis: *Mathematical Models for Air Pollution Health Effects*.

From 2004 to September 2009 he conducted research at the Delft Institute of Applied Mathematics that resulted in this Ph.D. thesis. Oswaldo's research has been supervised by Prof. Dr. Roger M. Cooke.

Since September 14 2009 Oswaldo is part of the Group of Risk and Reliability Analyzes in Civil Infrastructure at the Netherlands Organization for Applied Scientific Research (TNO) in Delft.

