University of Strathclyde Glasgow

Strathprints Institutional Repository

Morales-Nápoles, O and Kurowicka, D and Cooke, Roger (2008) *EEMCS final report for the causal modeling for air transport safety (CATS) project.* [Report]

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (http://strathprints.strath.ac.uk/) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: mailto:strathprints@strath.ac.uk

http://strathprints.strath.ac.uk/

EEMCS Final Report for the Causal Modeling for Air Transport Safety (CATS) Project.

O. Morales-Nápoles, D. Kurowicka, R.M. Cooke

EEMCS-TU Delft

Delft Institute of Applied Mathematics

The Netherlands, December 08

This page is left blank intentionally.

Table of contents.

1 Introduction: the Risk Perspective

World wide aviation data shows an increase in volume in all types of air transportation. A representative graph from the Dutch Civil Aviation Safety Data 1989-2003 [34] shows the number of flights doubling between 1980 and 2003.

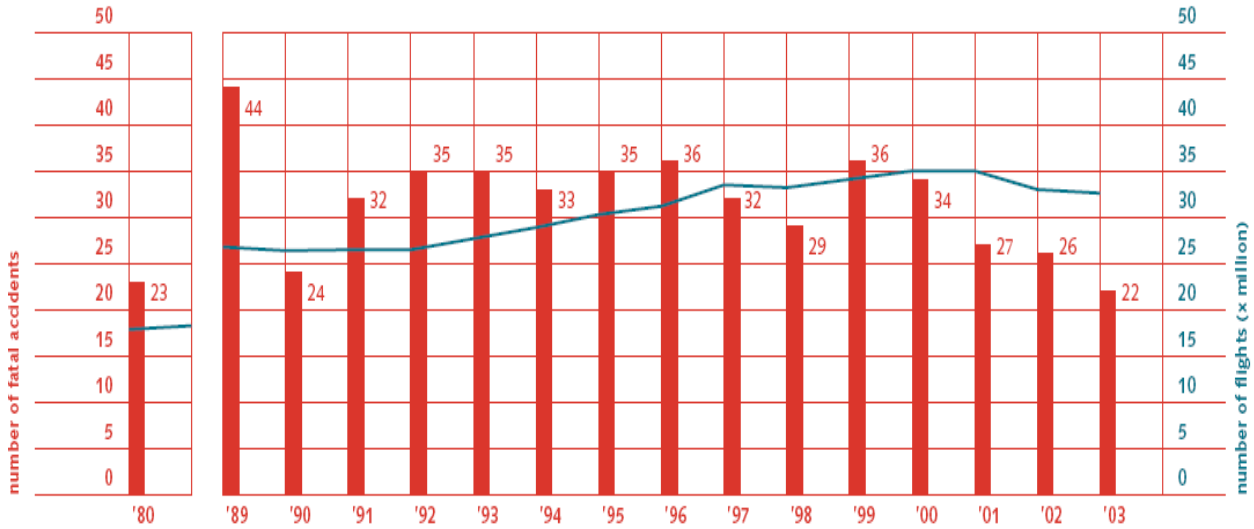


Figure 1.1: Number of accidents involving a least one fatality within 30 days (red bars) and number of flights (x million) (blue line) from 1980 to 2003, for world wide commercial aircraft with take-off weight of 5700 kg or heavier.

The world wide frequency of accidents (number per year) shows no marked trend. However, the number of accidents per flight is decreasing, both world wide and for European Air Safety Agency countries (see Figure 2). Whereas world wide, the accident rate has been deceasing by 2.8% per year, for EASA countries, the fatal accident rate is deceasing by 4.9% per year.

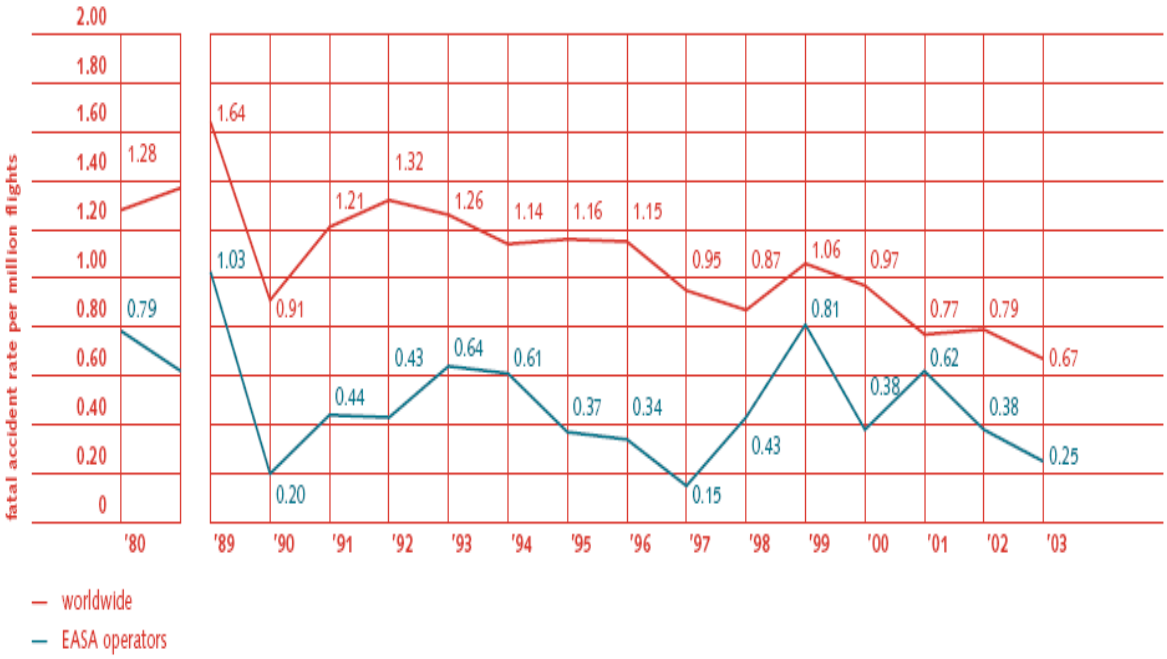


Figure 1.2: Fatal accidents (at least on fatality within 30 days) with commercial aviation (take off weight 5700 kg or heavier), per million flights, world wide (red) and for EASA (blue).

The breakdown in types of accidents shows that controlled flight into terrain (CFIT) is the dominant accident type. The main causal contributor is "cockpit crew".

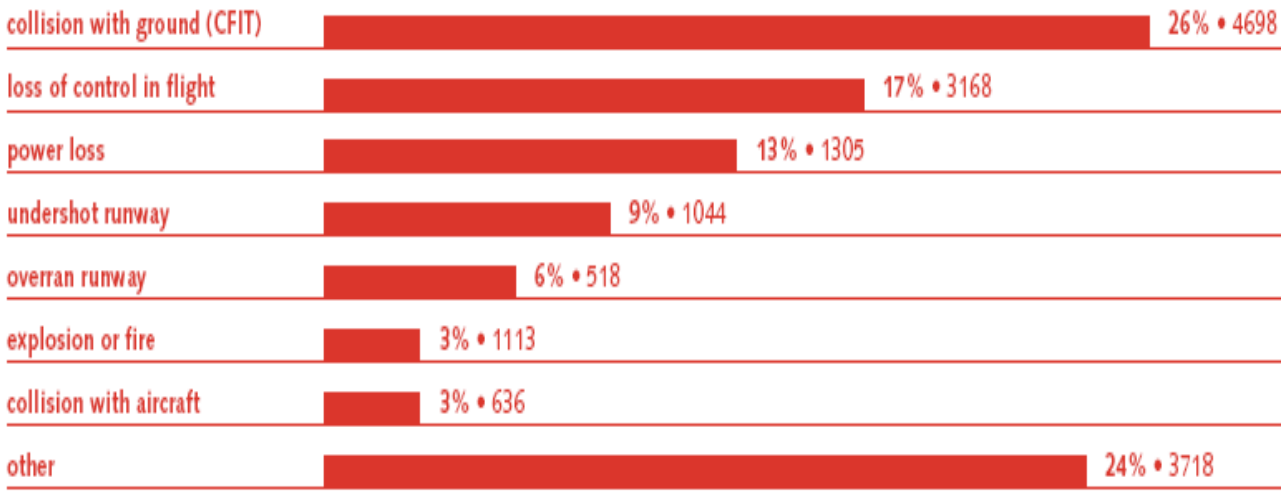


Figure 1.3: Principal causes of fatal accidents for commercial aviation (take off weight > 5700kg), categorized accordint to the International Civil Aviation Organization, 1989-2003.

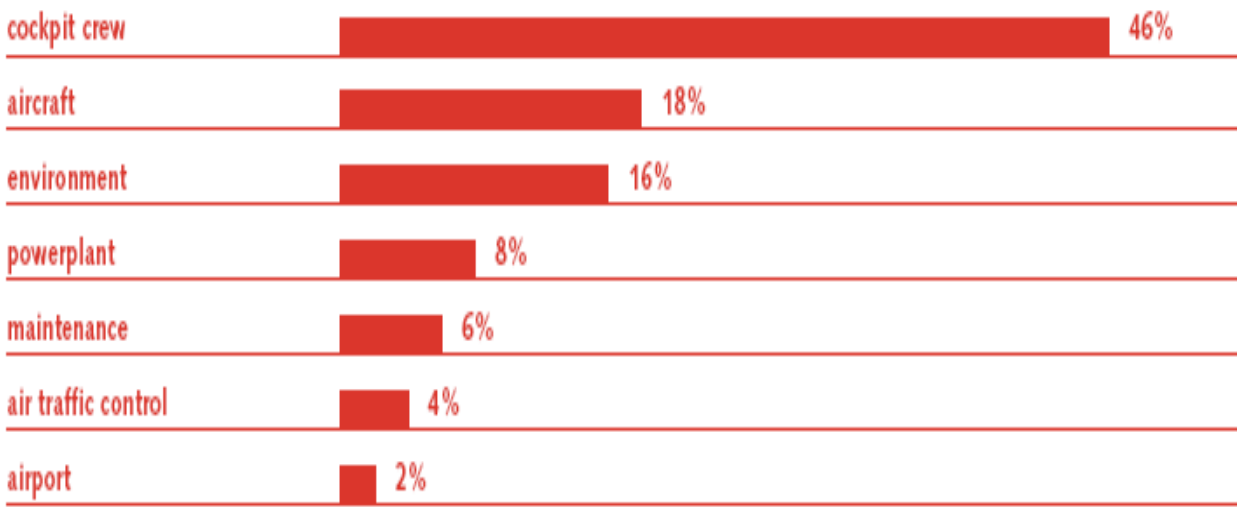


Figure 1.4: Relative importance of contributing factors to fatal accidents with commercial aviation (take off weight > 5700 kg) where a contributing factor is "a decisive factor in the causal chain of events leading to a fatal accident", 1980-2003.

The FAA in 2006 [33] forecast a 4.4 percent growth rate in civil air transportation volume, implying a doubling in 16 years. If historical trends continue, this growth in volume must be accompanied with a *decrease* in the accident rate per flight which is at least as great, preferably greater. Hence, designing for increased volume must be coupled with designing for decreasing risk.

Given the already very low accident rate for commercial aviation, together with the very high complexity the total civil aviation system, 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 has 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 ([3],[5]). 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 Causal Model for Air Transport Safety (CATS) combines Event Sequence Diagrams (ESDs) [18], Fault Trees (FTs) [19] and Bayesian Belief Nets (BBNs) [8] into a single continuous-discrete nonparametric BBN ([6], [7]). The ESD methodology is used for representing accident scenarios. An ESD is a representation of an event tree which distinguishes different types of events. An ESD consists of an initiating event, pivotal event(s), and an end state. Where necessary, the initiating and pivotal events are detailed in a sub-model which can be a Fault Tree or a Bayesian Belief Net. A schematic representation of the CATS model is presented in Figure 1.5.

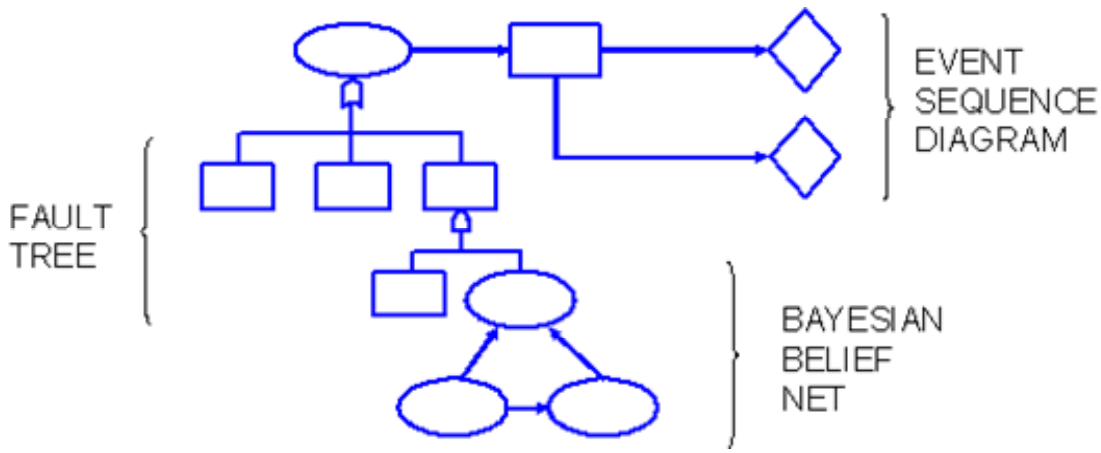


Figure 1.5 Schematic representation of the CATS model

Figure 1. shows the BBN representing the CATS model. It is evident that the simple idea represented in Figure realizes a very complicated graphical structure once all the elements of the model are finally quantified and integrated into a single BBN.

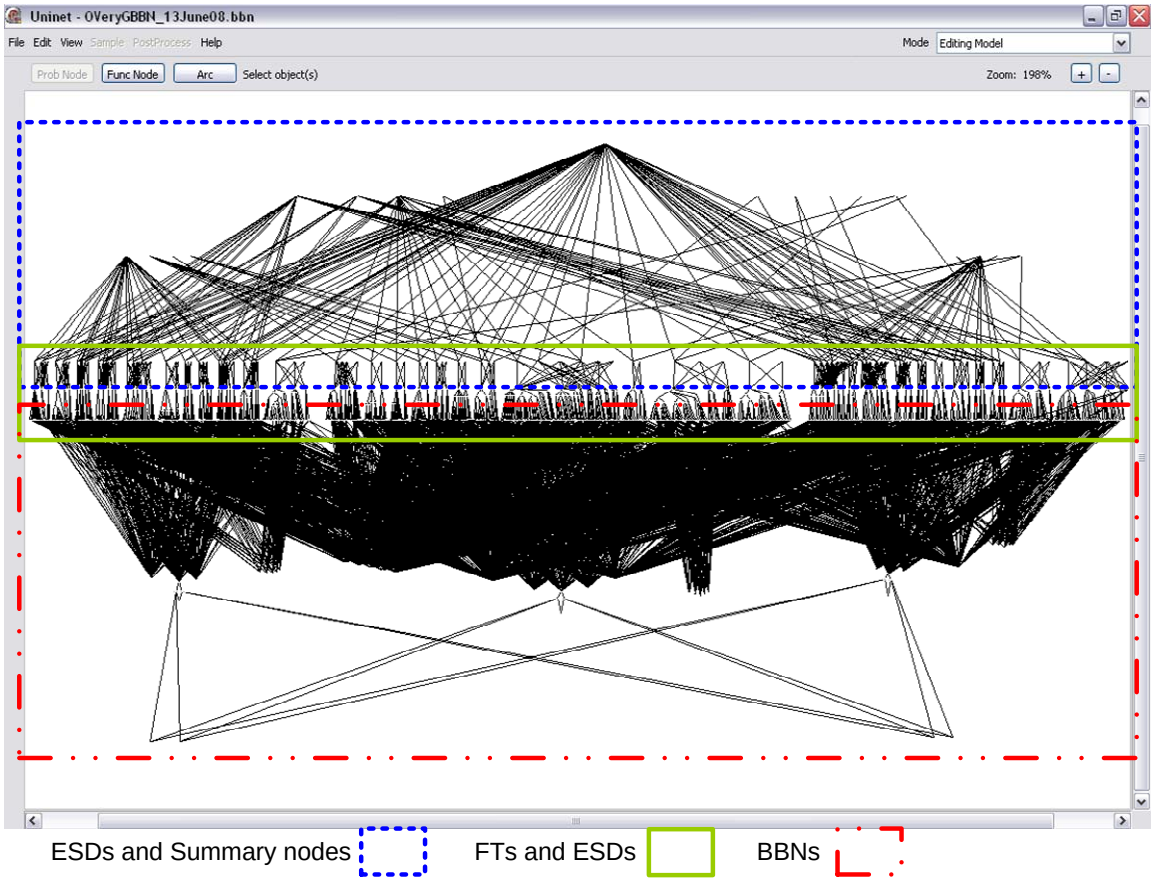


Figure 1.6 The BBN representing the CATS model

This document reports on the work realized by the DIAM in relation to the completion of the CATS model as presented in Figure 1.6 and tries to explain some of the steps taken for its completion. The project spans over a period of time of three years. Intermediate reports have been presented throughout the project's progress. These are presented in Appendix 1.

In this report the continuous-discrete distribution-free BBNs are briefly discussed. The human reliability models developed for dealing with dependence in the model variables are described and the software application UniNet is presented.

2 Graphical Models: Event Trees, Fault Trees and Bayesian Belief Nets

This chapter compares and contrasts three common types of graphical models used in risk and reliability theory. Event Trees, Fault Trees and Bayesian Belief Nets (BBNs). All are used in the CATS model. The discussion is aimed at non-technical readers, and is designed to show how these types of models work together. .Fault trees and event trees have been used together in risk analysis since its inception; therefore, the discussion is economized by focusing on BBNs and Fault trees.

We follow customary parlance that "arc" denotes a directed arrow between two nodes, whereas "edge" is an undirected link. Event Trees and Fault trees both have implicit directionality associated with their links, and we therefore term them arcs, even though they are usually drawn without arrowheads.

2.1 Event trees

Event trees consist of two basic elements: Nodes and arcs. Nodes are also called events even if there is not really anything that "happens". The arcs connect the events. They are usually represented by arrows, indicating the logical or temporal progression through the tree. In most cases the logic of the tree runs from cause to consequence. The cause is also called the parent; the consequence is called the child. Nodes generally have only two states: *Y*es or True and No or False. We restrict attention to two-valued events. If we exit the parent node via the 'true' arc, then all subsequent events on this path are conditional upon the Parent being True; similarly for False.

Since the tree has a direction: it belongs to the class of Directed Graphs. Since a child cannot be the parent of its ancestors, a cycle is not possible. So an event tree belongs to the class of Directed Acyclic Graphs or DAGs. A distinctive feature of event trees among DAGs is that they have exactly one source node, that is, one nod without a parent. On the other hand they have many sink nodes – nodes without children. These correspond to all possible end states of the system being modelled.

Although we tend to think that time progresses if we go from parent to child to grandchild etc., this is actually not what the tree represents. The tree is strictly LOGICAL. This means that the states set themselves instantaneously.

Therefore in event trees, events can be strung together that do not have a causal

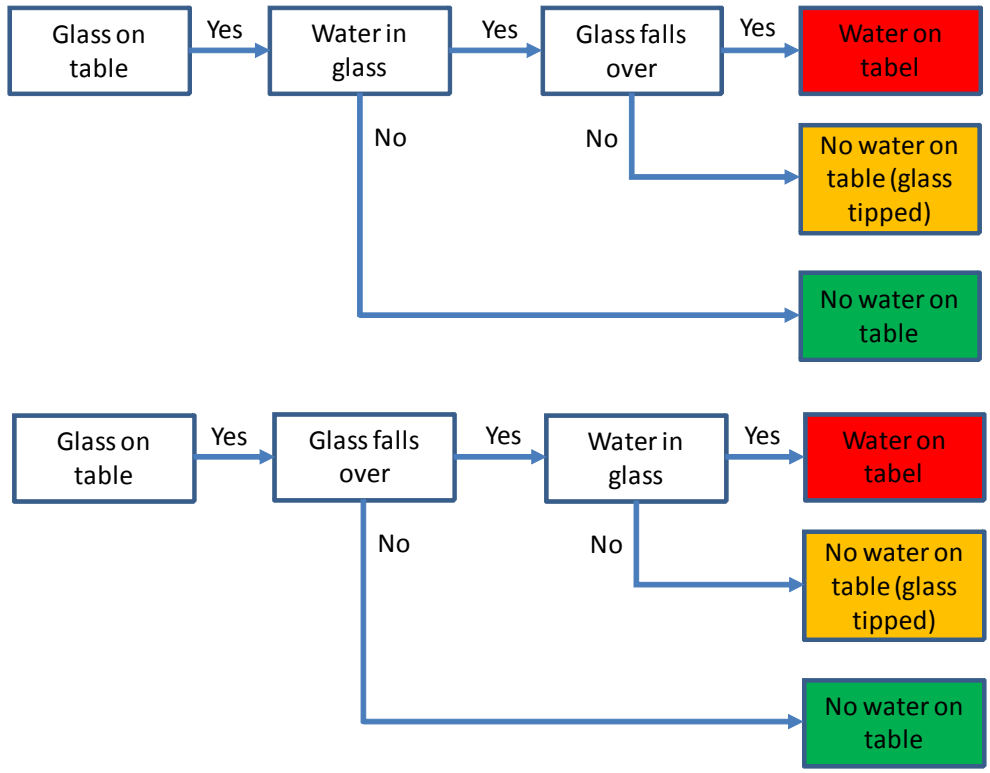


Figure 2.1 Two versions of the event tree for the same accident: water spilled on table.

relationship. For instance to have water spilled form a glass (Figure 2.1), there has to be water in the glass and the glass has to fall over. The water in the glass is not the cause of it falling over, nor the other way around.

The two states of each node are associated with probabilities which must sum to one. The probabilities are the chances that the state of the node is True or False. These are conditional probabilities, conditional on all predecessors. This means that the total probability of the true state of a node is the probability that the node is reached in the tree multiplied by the probability of going into the true exit.

2.2 Fault trees

Fault trees have three basic elements: nodes and arcs, as in event trees, and gates. For an overview on FTs see ([19]). In event trees the tree branches out at every node. Fault-trees combine arcs in gates. The logical flow is from basic events to a “Top Event”. Cycles are forbidden, so fault trees are also DAGs. In contrast to event trees, there are many source nodes – these are the basic events of the fault tree. There is typically one sink node – a node

without children, namely the Top Event. Whereas Event Trees model the possible states of a system, Fault Trees model the possible ways in which a given event can occur.

A node can have only two states: true or false. The state of a node with parents is completely determined by the state of the parents and the type of gate in which they combine. If they combine in an AND gate, the node is True if and only if all parents are true. If only one parent is False, the child is False. (Figure 2.2).

If the parents combine in an OR gate, the node is True if and only if one or more of the parents is True. There are more types of gates, but they occur less frequently.

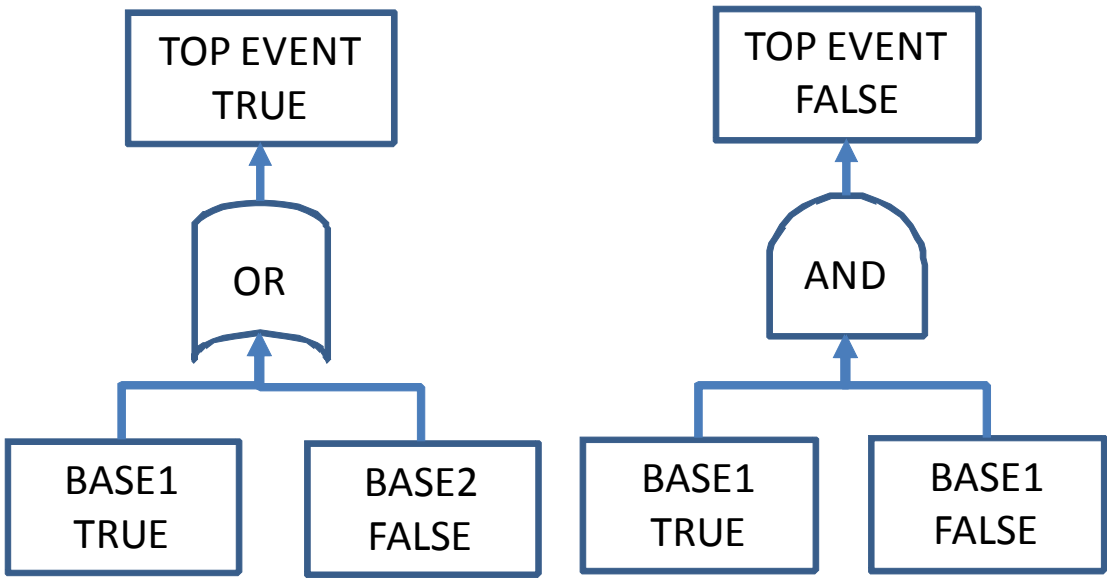


Figure 2.2 Examples of OR and AND gates.

Fault trees have one final sink node, that is, a node without children. This event is called the top event of the tree. The state of the top event is completely determined when the states of the basic events are specified.

The states of the base events can also be associated with probabilities, as will be discussed later.

2.3 Interdependencies in fault and event trees

In many real systems the logic is not as straightforward as described above; what goes on in one branch of the tree actually influences what goes on in another branch. For example a system might have two redundant pumps for coolant. For the cooling system to fail it is sufficient that both pumps fail. Under normal circumstances the probability of both pumps failing independently would be the product of their individual failure probabilities. However,

they both depend on the electrical system; a power failure would take out both pumps together. In this case the pumps are said to fail due to a common cause. Whenever possible, common cause dependencies should be modelled in the fault tree. This is easily said, but often less easily done. In scramming a nuclear reactor, an accident may result if three or more neighbouring control rods fail to insert properly. These rods may fail independently. However, of greater concern is the possibility that they fail due to some common cause. In this case the set of possible common causes is more difficult to identify and more difficult to model in the fault tree. Mathematical models must be posited. Detecting the effect of common causes from data is often very challenging .

Whereas hardware common causes can in principle be identified and modelled with engineering knowledge of the system, with other common causes this is more problematic. A few examples illustrate this point:

1. A fire near a cable tray can disable multiple components simultaneously,
2. A flaw in the training of maintenance personnel can cause maintenance errors to occur simultaneously in physically separated sub-systems,
3. An error in the building design may force evacuation of the control room in certain otherwise non-critical situations, thus increasing the chance of operator error.
4. A poor safety culture at top management levels can degrade operator alertness, maintenance performance, training, acquisition and testing of spare parts, quality of emergency procedures, etc, etc.
5. A flaw in the regulatory regime may allow unsafe situations and practices to arise simultaneously in an entire population of systems.

All of these features can increase the probability of multiple simultaneous failures in a system. Hence if we simply gathered failure data from system components, and assume that these systems can only fail independently, we may produce an unrealistically optimistic prediction of system performance. With an appropriate model, we can predict the effects of different levels of training, inspection, maintenance etc.

Dependence modelling in fault trees works well when dependences can be associated with failures of support system hardware components. Such support systems might include the electrical system, the sprinkler system, the lubrication system, and the software control system. In such cases, the failure of a support system *causes* the failures, or unavailabilities, of multiple components.

Other dependences, such as numbers (2) – (5) above, do not express themselves directly by causing simultaneous component failures. Rather, they simultaneously influence the *probability* of failure of multiple components. Examples include poor maintenance, poor operator training, poor incident reporting, poor safety culture, etc. Fault tree modelling cannot readily capture dependences that influence the probabilities of failure. Influences acting on the probability of failure, rather than on failure itself, must be captured in the uncertainty analysis of fault trees. Bayesian Belief Nets are a modelling tool specifically designed to capture probabilistic influence. Before we can fully understand uncertainty

analysis of fault trees, we must understand the difference between Boolean and ordinary arithmetic.

2.4 Boolean Arithmetic for Fault Trees

A Fault Tree is just a picture of a Boolean formula. In Boolean arithmetic, variables take only the values 0 or 1. Boolean arithmetic operates in Mod 2. Suppose X and Y are Boolean variables, in Boolean arithmetic, $X +_b Y$, and $X \times_b Y$ are also Boolean variables, and hence take values 0 or 1. This means that $X + Y$ must take only values 0 or 1, which is arranged by defining $X +_b Y = X + Y - XY$; $X \times_b Y = XY$. In Boolean arithmetic, addition and multiplication correspond to the operators AND and OR in propositional logic. Thus, $X +_b Y$ = "either X or Y or both", $X \times_b Y$ = "X and Y". In particular, in Boolean arithmetic, $X^2 = X$; this corresponds to saying that the event X AND X is the same as the event X.

A fault tree is a Boolean formula, when we fill in 0's and 1's for the basic events, and apply the AND and OR operators, we always obtain a 0 or 1 for the top event. In most cases, we don't know whether a given basic event occurs, we know only its probability of occurrence. The probability of event X occurring is the expectation of the random variable X; since

$$E(X) = Prob\{X=1\} \times 1 + Prob\{X=0\} \times 0 = Prob\{X=1\}.$$

If the events X and Y are independent, then it is easy to check that

$$E(X \times_b Y) = E(X)E(Y);$$

$$E(X +_b Y) = E(X) + E(Y) - E(XY) =$$
$$Prob\{X=1\} + Prob\{Y=1\} - Prob\{X \text{ AND } Y = 1\} = Prob\{X \text{ OR } Y\}.$$

This 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. A simple example illustrates this feature. Suppose the Top Event occurs when either X AND Y or X AND Z occurs. We could represent this with two simple, logically equivalent, fault trees:

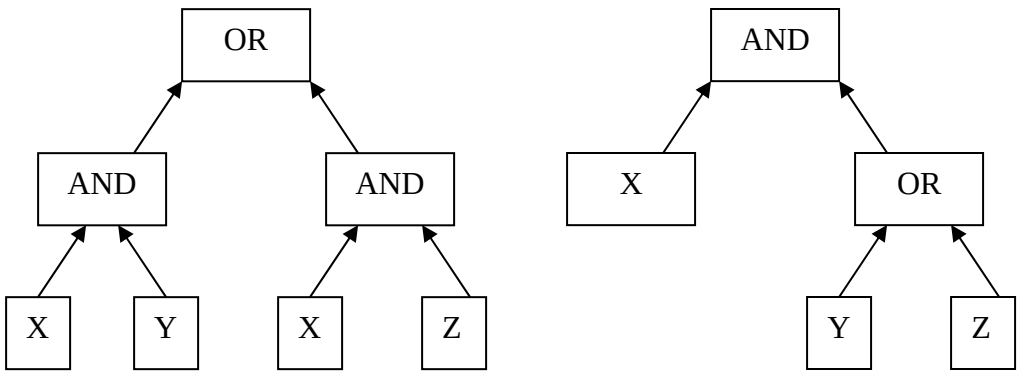


Figure 2.3Two logically equivalent FTs.

The Boolean formula from the left tree is

$$X \times_b Y \;\; +_b \;\; X \times_b Z \;\; = \;\; XY + XZ - XYXZ$$

The Boolean formula from the right tree is

$$X \times_b (Y +_b Z) = X(Y + Y - YZ) = XY + XY - XYZ.$$

If we apply Boolean reduction to the first formula: $XYXZ = XYZ$, we see that these two formulas are equivalent. However, if we replace the variables by their expectations and apply ordinary arithmetic, we will get different answers. From the first tree we would get the incorrect formula:

$$\begin{gathered} Prob(Top\ Event) = \\ Prob\{X=1\}Prob\{Y=1\} + Prob\{X=1\}Prob\{Z=1\} - \\ Prob\{X=1\}Prob\{Y=1\}Prob\{X=1\}\ Prob\{Z=1\} \end{gathered} \qquad (1)$$

The correct calculation would be obtained from the right tree:

$$\begin{gathered} Prob(Top\ Event) = \\ Prob\{X=1\}Prob\{Y=1\}+Prob\{X=1\}Prob\{Z=1\}- Prob\{X=1\}Prob\{Y=1\}Prob\{Z=1\} \end{gathered} \qquad (2)$$

The problem is that when we compute (*X* AND *Y*) AND (*X* AND *Z*) with expectations in the left tree, we would include the term "P{*X*=1}" term 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.

When repeated events are present, we can estimate the probability of the top event by various methods. One method is to reduce the fault tree to a *minimal cut set* representation: we write the top event as a disjunction of conjunctions of several basic events, where each combination is a minimal set of events whose joint occurrence is sufficient to fail the system. A prose reading of the minimal cut set equation would be something like:

Top event happens if and only if EITHER:

BE1 And BE5 And BE23
Or
BE23 And BE8 And BE31 And BE 26
Or
Etc.

We can estimate the probability of occurrence using the inclusion-exclusion principle. However, if the combinations which fail the system have low probability, then we may approximate the probability of the disjunction as the sum of the probabilities of the combinations. We do not explain this further here, as this material is available in any standard text, and the fault trees in the CATS model do not involve repeated events. In the CATS model, we *can* simply replace basic events with expectations and compute with ordinary arithmetic.

2.5 Bayesian Belief Net

A Bayesian Belief Net is a special kind of directed acyclic graph. In a BBN nodes represent random variables and arcs represent probabilistic or functional influence. Since Boolean gates in a fault tree are simply functional dependences of a particular sort, and Event trees are probabilistic dependences of a particular sort, it is evident that BBNs are a more general structure than either event of fault trees. Thus it is possible to represent both as BBNs. In the CATS model there are event trees and fault trees, which are represented as parts of an over-arching BBN. The point of using a BBN, however, is not simply to replicate modelling of event trees and fault trees; rather, the point resides in the fact that BBNs can capture probabilistic influences between random variables – not just two valued events - which cannot be modelled functionally. This enables us, in principle, to capture factors which influence the *probability* of failure for basic events in a fault tree: the factors (2) – (5) in section 1.1.3.

A simple example helps. Suppose we are interested in a particular human error event. In a fault tree this is a Boolean variable, but since our fault tree has no repeated events, we may replace this Boolean variable by its expectation. We wish to model those factors which influence this human error probability. Suppose we identify Training and Fatigue as relevant influences. We construct the simple BBN shown below

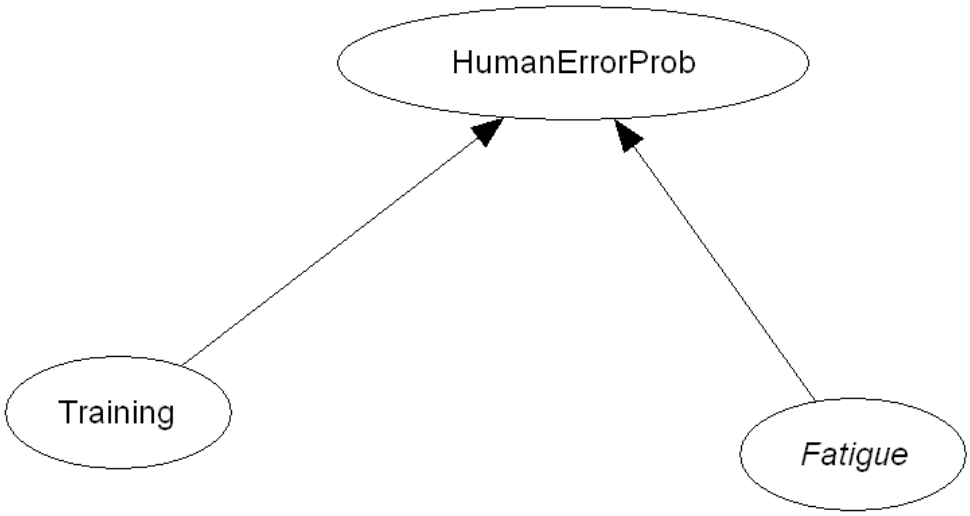


Figure 2.4 Simple graph representing a BBN for human error probability.

Human error probability is operationalized as the relative frequency of error per demand. Training is operationalized as 'hours spent in refresher courses over last 3 years' and Fatigue is operationalized on the Stanford sleepiness scale. From data we can recover the distribution of operators training and the distribution over possible fatigue states. We also have human error data which, with some appropriate (Bayesian) procedure will enable us to form a distribution over the probability of human error. Thus we have the marginal distributions for all the above variables. However, marginal data does not tell us how Training and Fatigue influence Human Error Probability.

Functional influence

As good engineers, we might try to capture the relation between these vaiables with some mathematical model. We could always start with a Taylor expansion around certain nominal values (note the inclusion of the first interaction term):

$$HEP = HEP_0 + (TR - TR_0)\partial HEP/\partial TR_0 + (FA - FA_0)\partial HEP/\partial FA_0 + (TR - TR_0)(FA - FA_0)\partial^2 HEP/(\partial TA_0\, \partial FA_0), \quad (3)$$

This of course is a possible approach, but it requires an appropriate choice of nominal values, a truncation of higher order terms at an appropriate point, and an estimation of partial derivatives. If none of these choices can be supported by data, it might make more sense to capture the influence as probabilistic influence.

Probabilistic influence

In opting for probabilistic influence, we are acknowledging data do not support a functional dependence relation. Instead, we opt for coarse modelling of how high / low values tend to occur together. This is measured by rank correlation.

Suppose we have some multivariate data of simultaneous observations of all three variables. That is, we have observations of equivalent systems s = 1,…K in which we observe *(HEP_s, TR_s, FA_s)* for each s = 1…K. We could then capture the influence as probabilistic influence in various ways, we might ask, for example:

In how many systems s are HEP_s AND *TR both above their median values?*

If that number is larger than one quarter of all observed systems, then high values of HEP and TR tend to occur together, if it is lower than one quarter, then high values of HEP tend to occur with low values of TR, and conversely. The tendency of high values to occur together is measured by rank correlation, and from the answer to this question, we could estimate the rank correlation between HEP and TR. The reader is spared the mathematical details for the time being. cite

To capture the influence of Fatigue in addition to TR, we might ask:

In how many of those systems in which both HEP_s and TR_s were above their medians, was also FA_s above its median?

With this information we could estimate the conditional rank correlation of FA and HEP given TR.

The above graph has no influence between FA and TR; this is a modelling assumption which might be checked against data. For example, we might ask whether systems characterized by high values of TR tend to show lower values of FA. For this example, we stick with the above graph. The reader may accept on faith that the information acquired, together with the independence implied by the graph, together with an assumption on the type of distributions realizing the rank correlations, completely determine the joint distribution of HEP, TR and FA (cite). The rank and conditional rank correlations obtained may be shown in the BBN as in Figure 2.5 below.

The correlation between TR and HEP is negative: high values of training tend to go with low values of HEP.

What if we do not have the sort of multivariate data needed to answer the above questions? In that case we ask the very same questions which we *would* ask to the data, if we had it, to knowledgeable experts. Many people balk at the introduction of expert judgment into quantitative risk studies. Of course, real data is always preferable. On the other hand, it

would be very unscientific to assume that influences for which we have no data, therefore do not exist.

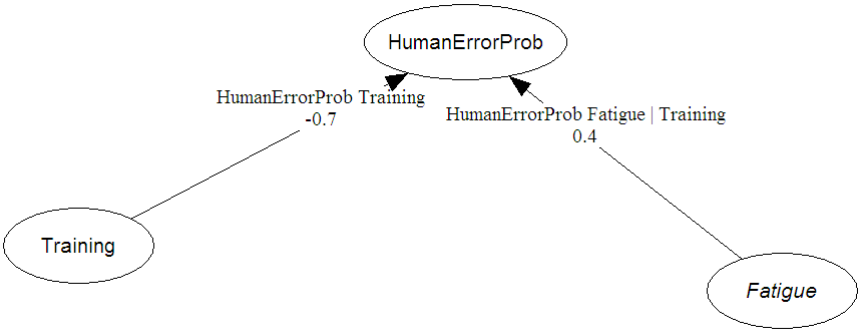


Figure 2.5 Simple graph representing a BBN for human error probability with rank and conditional rank correlations attached to its edges.

2.6 Using a BBN

In this section some features of BBNs will be discussed, next chapter deals in more detail with some of the concepts presented here. This section already uses UniNet (see next chapter) to some extent. The first thing we can do with a BBN is to compute joint and marginal distributions. Switching to the histogram view[1], we first show the unconditional (marginal) distributions. The mean and standard deviations for each variable are shown beneath the histogram in Figure 2.6.

Suppose we want to set a training level equal to 50 hrs per year; or 150 hours in three years. How might that affect the human error probability distribution? If our training program does not change the system in any other way, our answer to this question is obtained by conditionalizing on TR = 150. This means, that changing a given system to have 150 hours of training would produce the same effect that we would see in a large subset of the original systems for which TR = 150 (Figure 2.7)

The distribution of HEP is shifted to the left; the mean has dropped from 0.00902 to 0.00396.

Of course, a real data set might not contain enough systems with TR = 150 to assess this effect. For this reason we build a density function for the three variables, based on the information displayed in the BBN. Again, the details are suppressed in this discussion. A

[1] In UniNet

density may be conceived as a mathematical smoothing of the histograms displayed in the BBN's, but then applied to all three variables at once.

We might further ask whether training enables operators to perform better when fatigued. Conditionalizing on TR = 150 and FA = 6, we see that at fatigue level 6, the well trained operators still have a lower average human error probability than the overall average (Figure 2.8).

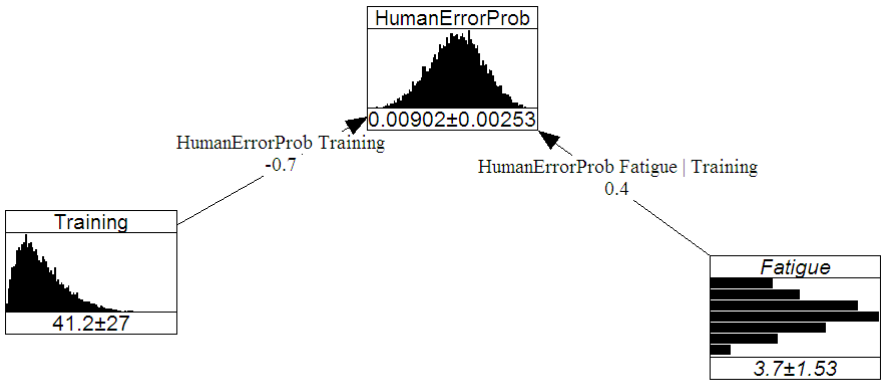


Figure 2.6 Simple graph representing a BBN for human error probability with rank and conditional rank correlations attached to its edges and nodes as marginal distributions.

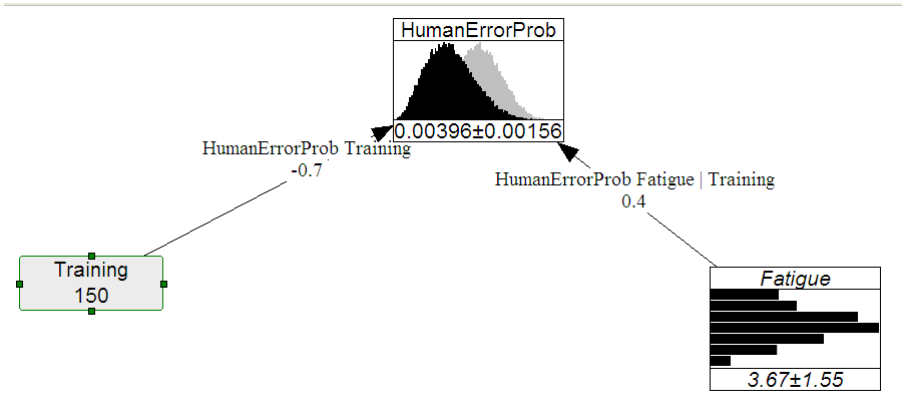


Figure 2.7 Conditional Distribution (TR = 150) of human error probability for the graph in figure 2.6.

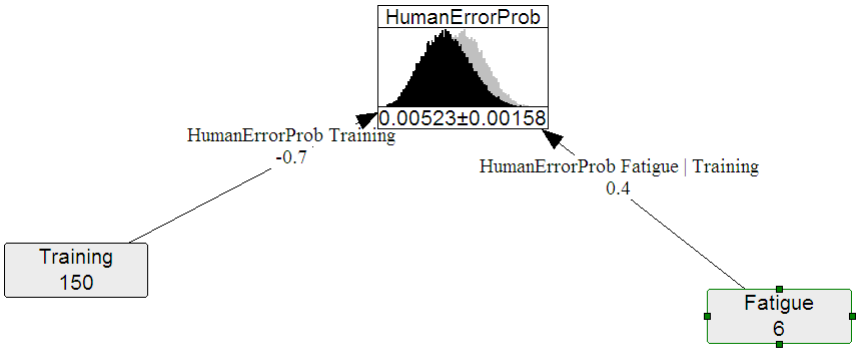


Figure 2.8 Conditional Distribution (TR = 150) and (FA = 6) of human error probability for the graph in figure 2.6.

2.7 Fault Trees in BBN

Our purpose of modelling HEP is to capture probabilistic influence on basic events of a fault tree. Let us assume that our Top Event, System Failure, can occur in only two ways, either failure scenario 1 occurs, or failure scenario 2 occurs (or both). In both these scenarios, failure occurs if a subsystem fails AND human operator fails to recover. The subsystems are different, and the operators of the two systems may be different. The occurrence of human error in subsystem 1 is not related to human error in subsystem 2. However, the *probability* of error in both cases is influenced by training and fatigue. We assume that training and fatigue are the same for both subsystems. The combined system fault tree with HEP BBN may be represented in one BBN, as shown below in figure 2.9.

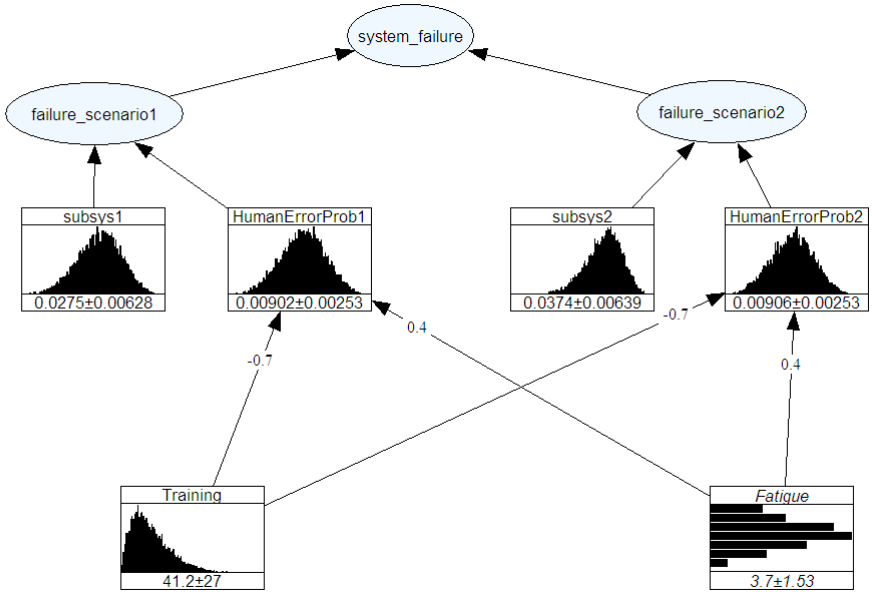


Figure 2.9 Conditional Distribution (TR = 150) and (FA = 6) of human error probability for the graph in figure 2.6.

Failure_scenario1, failure_scenario2 and system_failure are represented as functional nodes. Failure_scenario1 is a probability, which is the probability of subsys1 AND HEP1, similarly for failure_scenario2. All these probabilities are random variables. The system failure is a function of the two failure scenarios, as shown in figure 2.10

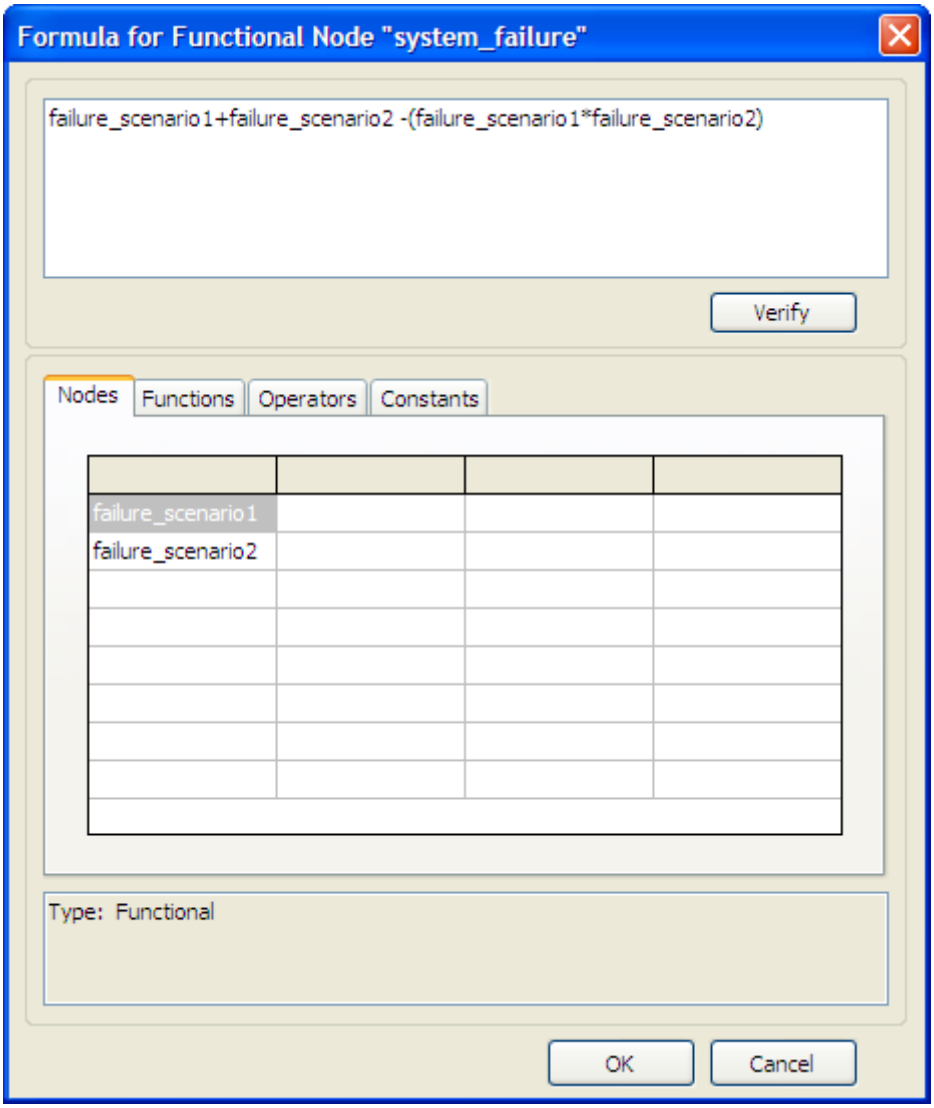


Figure 2.10 UniNet view of functional node representing system failure

If we shift to the histogram view in UniNet, we see the distributions of all variables (figure 2.11).

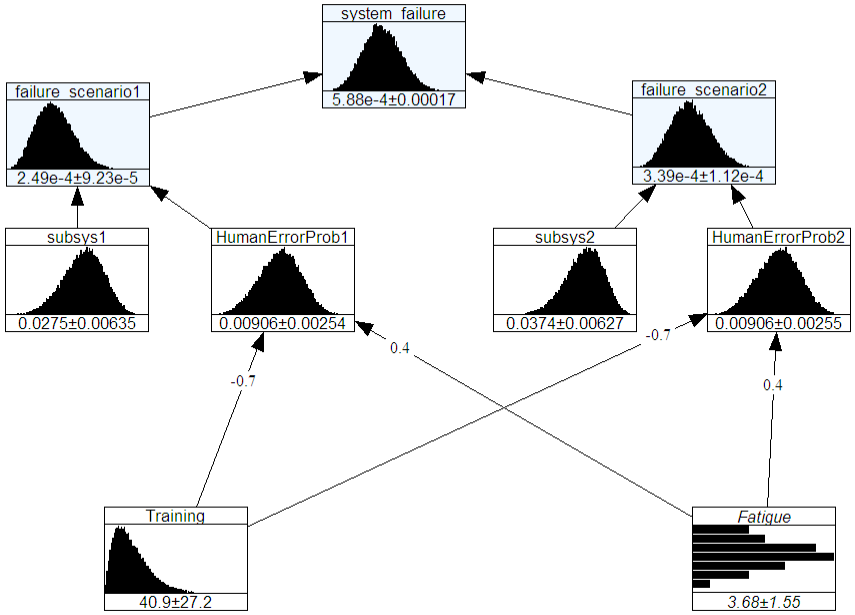


Figure 2.11 UniNet histogram view of the example BBN in figure 2.9

A detailed picture of system_failure is shown below in figure 2.12 (also from UniNet). The median probability is 5.8148 E-4, with 95%-tile 8.7752E-4.

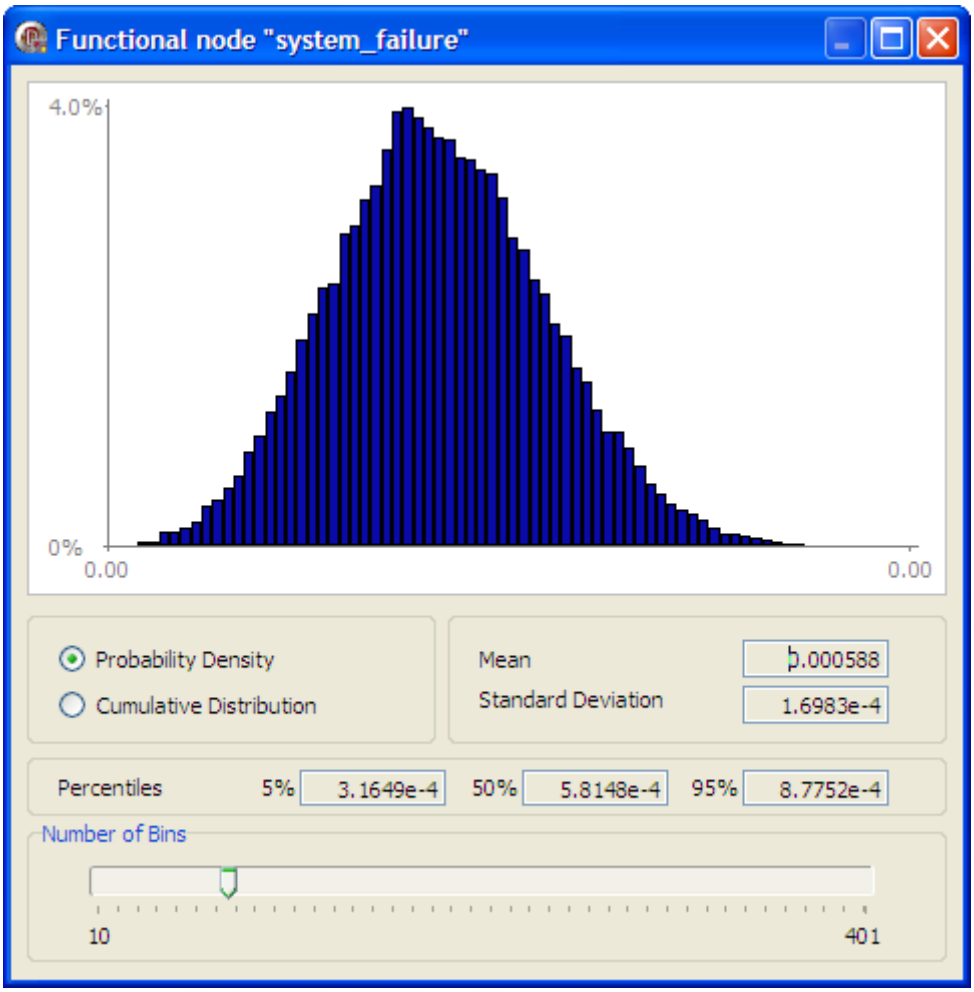


Figure 2.12 UniNet histogram view of the node system_failure from the example BBN in figure 2.9

The BBN for HEP has the effect of correlating the variables HEP1 and HEP2. In fact their correlation is 0.58. If we had not introduced the BBN for HEP, or equivalently, if we had made the influence of TR and FA on HEP equal to 0, then the distribution of the top event would change slightly. The result of this change is shown below in figure 2.13

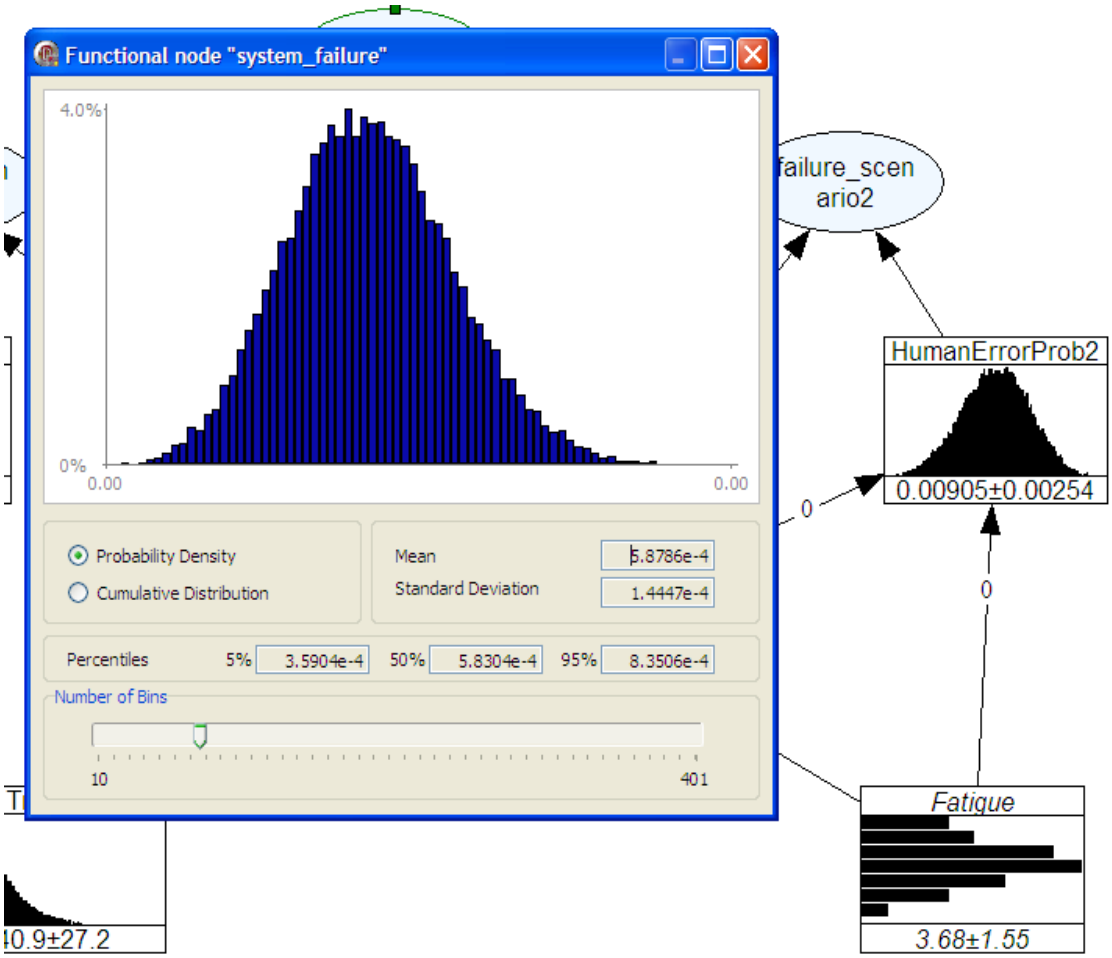


Figure 2.13 UniNet histogram view of the node system_failure from the example BBN in figure 2.9 making all correlations equal to zero

Without dependence on TR and FA, the 95%-tile of system failure shifts from 8.7752E-4 to 8.3506E-4.

The effect of conditionalizing TR at 150 hours in the BBN with the original dependence structure (figure 2.11) can now be propagated through the fault tree up to the system failure. The mean probability for system failure shifts from 5.88E-4 to 2.75E-4 (figure 2.14).

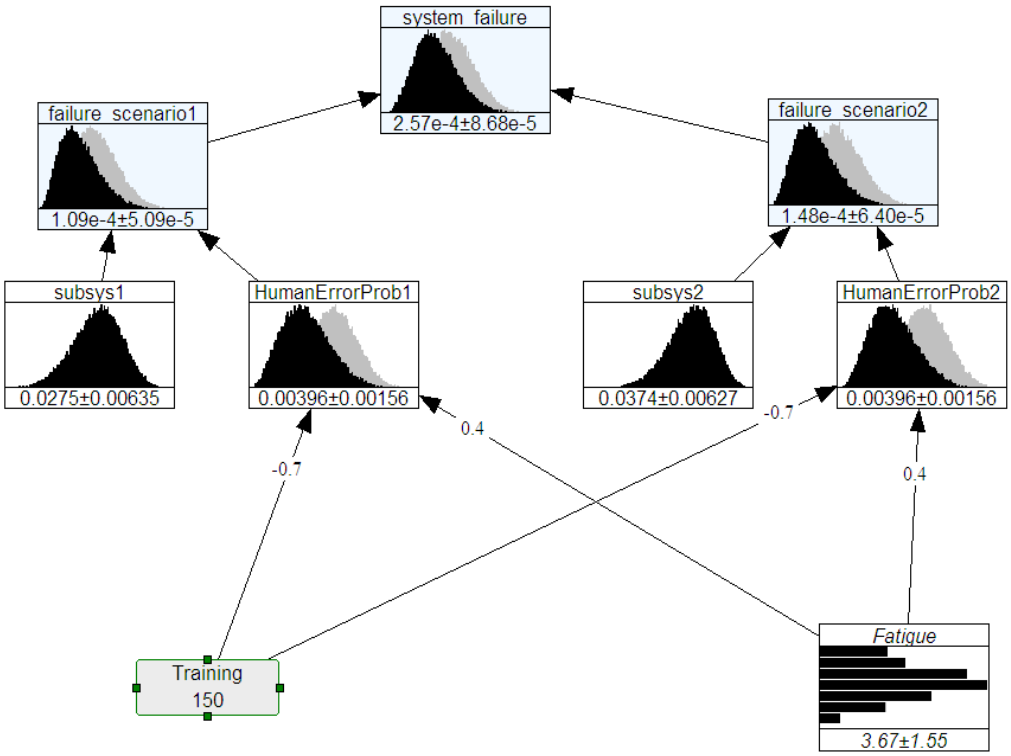


Figure 2.14 UniNet histogram view of the BBN in figure 2.11 conditionalized on (TR=150)

2.8 Calibration and apportionment

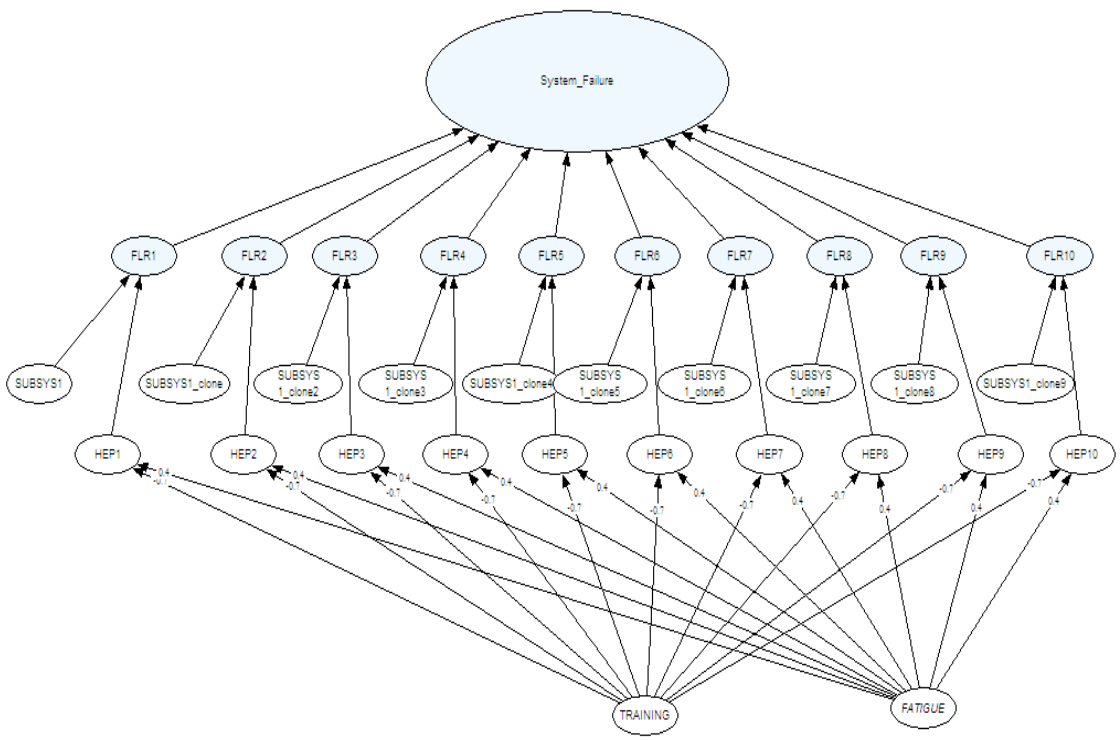
In the model discussed above we had data on the HEP's, on Training, Fatigue and on subsys1 and subsys2. With this information we predict the distribution of System_failure. In some cases, such as CATS, we have some data on System_failure, typically this will be in the form of expected values or generic probabilities for System_failure In such cases the modelling serves to apportion this generic probability over the underlying basic events and BBN nodes. The goal is to be able to predict how changes at a lower level will impact the probabilities at the higher levels.

In performing this apportionment, we may have some expected values at intermediate levels to guide us. The apportionment activity consists of:

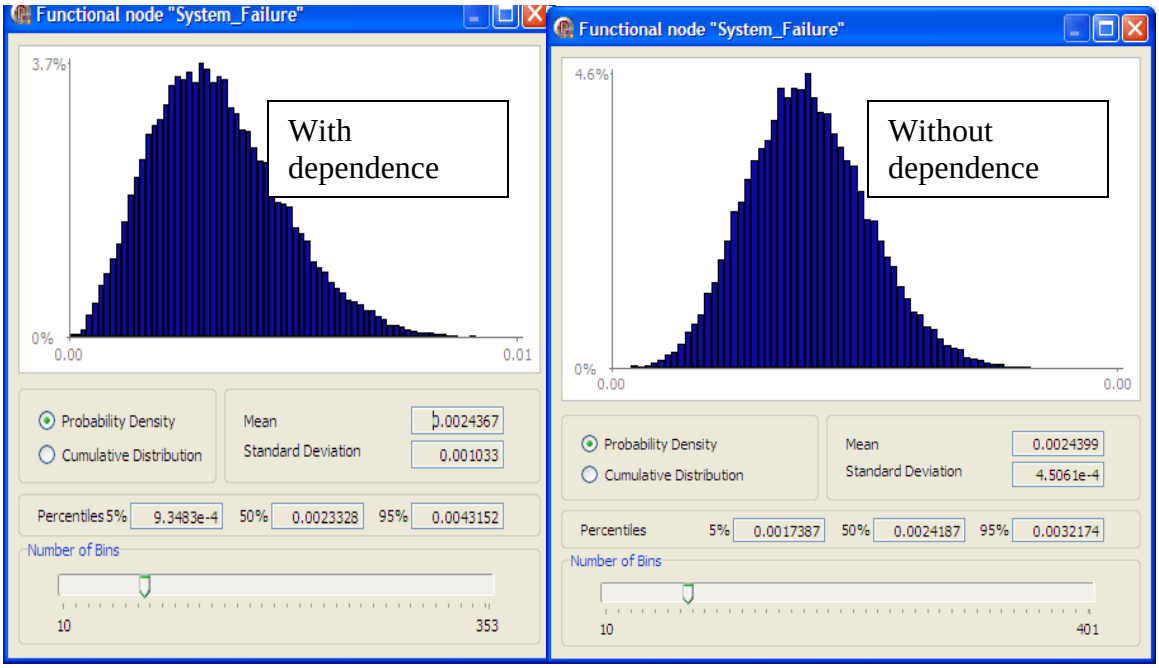
1. Developing expected values from data
2. Assessing the degree of variability consistent with the expected values, at those nodes for which distributional information is lacking.
3. Checking that with these variabilities, the expectations of the higher events are still calibrated on the data from step (1).

We saw in the example above that the dependence in basic events introduced by the BBN caused a slight shift in the distribution of System_failure. If the distributions are highly skewed, or if the dependences are very strong, the induced shift might be enough to disturb the calibration. In this case the above steps must be iterated.

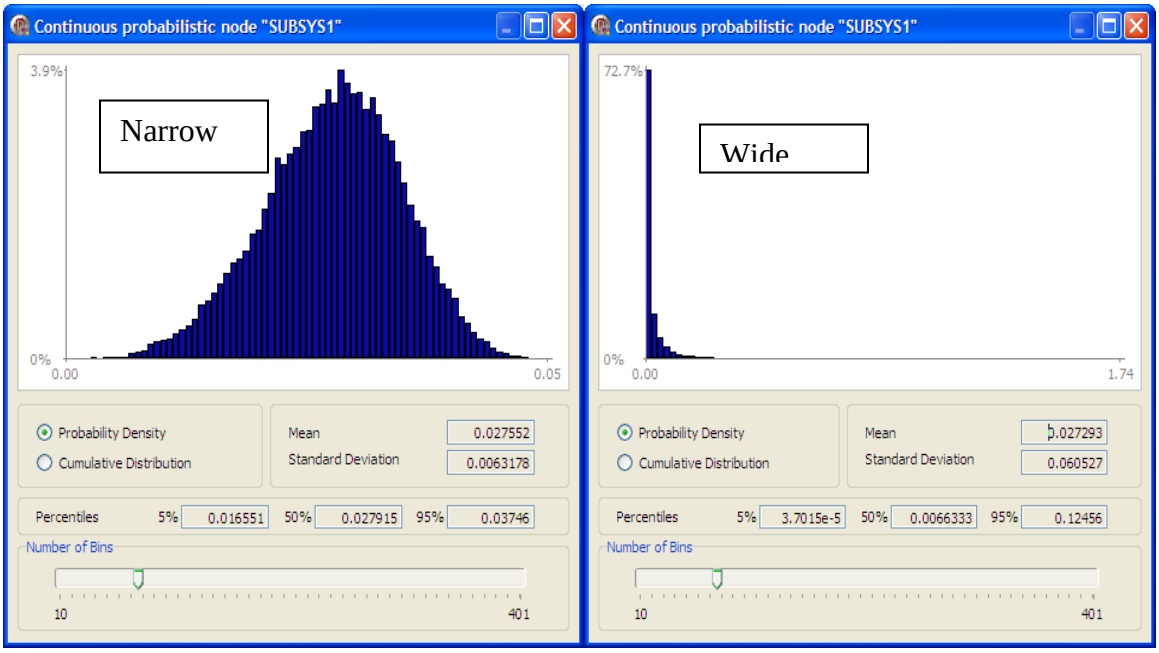
The simple example is extended to illustrate these points. We consider a system consisting of 10 failure scenarios, each one similar to the two failure scenarios discussed above. The BBN for this system is:



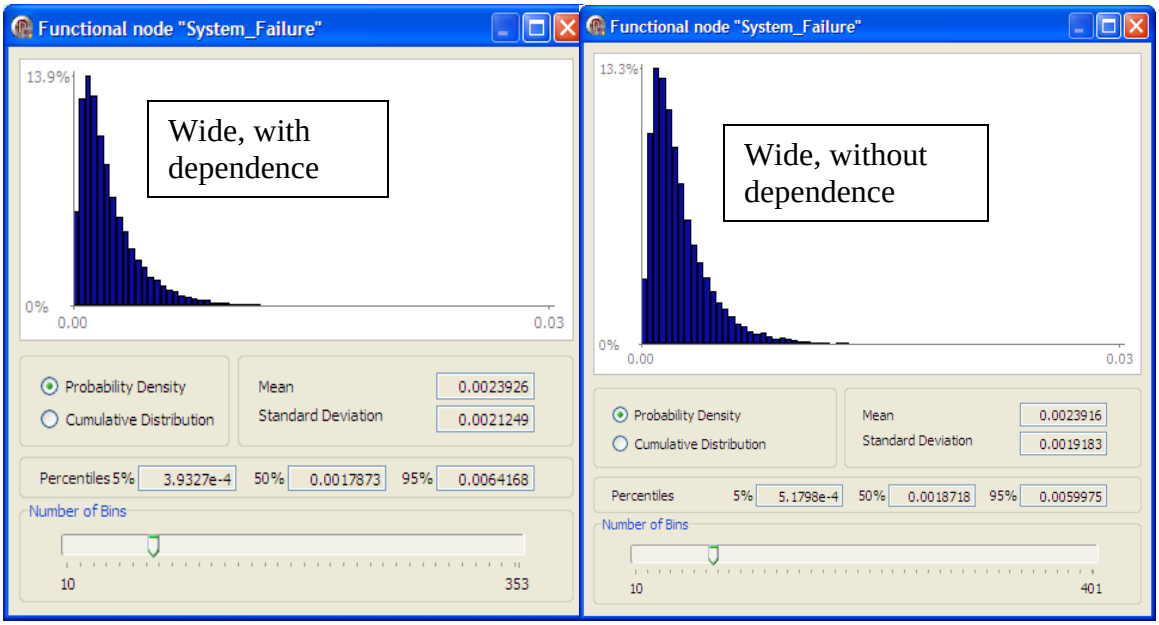
We compare the distribution of System _failure with and without dependence. Relative to the independent case, adding dependence causes the mean to shift up imperceptibly, but the standard deviation increases from 4.506×10^{-4} to 1.033×10^{-3}.



This effect could be altered if the distributions are more skewed to the right. To investigate, we change the distribution of all the subsys variables, keeping the mean fixed, but increasing the standard deviation by roughly a factor 10. The 95%-tile is increased by a factor 4:



The resulting effect on System_failure is shown below. Now the standard deviation and the mean remain effectively unchanged.



In these examples the effect of adding dependence is visible in the standard deviation of the top event, but not in the mean. Thus, in these examples the dependence induced by the BBN would not require iterating the calibration step. However, we cannot guarantee that this will remain true for very large systems like CATS, with very highly skewed distributions.

3 BBNs and UNINET

This chapter goes into more detail about the non-parametric continuous Bayesian Belief Nets, and explains essential features of the UNINET software system which implements this type of Belief Net. UNINET was developed for the CATS project, and its specific design features profited from previous projects in which commercial off the shelf BBN software platforms were tested.

To recall, a BBN is a graphical structure known as a *Directed Acyclic Graph* (DAG). A DAG is a set of nodes and arcs, or directed edges, between nodes, such that there is no directed cycle. In a BBN, nodes represent univariate random variables, and arcs represent influences. Node1 is a *parent* of Node2 if there is a directed arc going from Node1 toward Node2. Node1 is an *ancestor* of Node2 if there is a directed path from Node1 to Node2. A node with a non-empty set of parents is a *child node.* A node without parents is a *source node*, a node without children is a *sink node*.

The simple example in chapter 1 for Human Error Probability is used throughout this chapter to illustrate features.

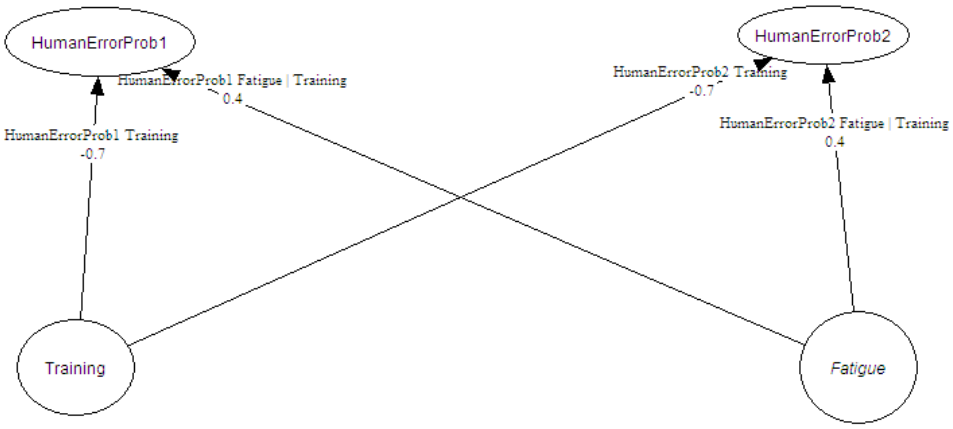


Figure 3.1 BBN example for chapter 3

3.1 Types of BBNs

Three types of BBNs may be distinguished: discrete, normal (or discrete-normal) and non parametric continuous-discrete. The first two are available in other commercial packages. All BBNs can have functional nodes (nodes which are functions of their parents); differences relate to the treatment of probabilistic nodes. The ensuing discussion refers to probabilistic nodes.

(i) Discrete

The random variables have discrete distributions. Source nodes are assigned (marginal) distributions, each child node is assigned a conditional probability table, giving the distribution over child's possible values, conditional on each possible combination of values of the parents. The number of probabilities that must be assessed and maintained for a child node is exponential in the number of parents. Adding a new parent node requires reassessing all children influences. If marginal distributions are known for child nodes, there is no way of entering this information directly; instead conditional probability tables must be tweaked to comply with all marginal data. Discrete BBNs have a heavy assessment/maintenance burden and are suitable for small problems only.

(ii) Discrete normal

In this case the set of variables is assumed to follow a joint normal distribution (certain source nodes may be discrete). Each child is associated with a mean value and conditional variance. Each influence is associated with a *partial regression coefficient*. Loosely the partial regression coefficient e.g. for 5 with respect to 1 says how much the conditional mean of 5 changes as a result of a unit change in the value of parent 1. Discrete normal BBNs work well if indeed the normality assumptions hold. If not, then:

- The individual variables must be transformed to normal (requiring of course the marginal distributions). If F is the cumulative distribution function of random variable X (continuous, invertible), and Φ is the standard normal cumulative distribution function, then $\Phi^{-1}(F(X))$ transforms X to normal (sometimes called the normal version of X),
- The conditional variance *in normal units* must be constant; this is the variance with respect to the units for which the distribution is normal. The partial regression coefficients apply to the normal units of the transformed variables, not to the original units. This places a heavy burden on any expert elicitation, and is dubious in most real applications, unless indeed the transformation to normal is almost linear.

If a parent node is added or deleted after quantification, then the previously assessed partial regression coefficients must be re-assessed.

(iii) Non-parametric continuous/discrete

UNINET supports non-parametric continuous/discrete BBNs with important distinguishing features. The main distinguishing features of UNINET are:

- UNINET supports non-parametric continuous/discrete BBNs, supplemented with functional nodes.

- Probabilistic nodes may be assigned arbitrary continuous or discrete distributions.

- Probabilistic influence is represented by *conditional rank correlation*, that is, by rank correlation in a conditional distribution, according to a protocol based on an indexing of the parents. Influence of the first parent on the child is unconditional rank correlation, influence of the second parent on the child is the conditional rank correlation given the first parent, etc. The conditional rank correlations are algebraically independent, that is, previously chosen values do not constrain future choices. Unspecified correlations may be accommodated without confronting the matrix completion problem ([20], [21])

- The rank correlation is realized using the joint normal copula. The indexing is user-editable, and parent nodes can be added without re-quantifying the previous influences.

- Analytic, real time conditioning is supported for thousands of probabilistic nodes. Analytic conditioning is conditioning on a single value, and applies only for probabilistic nodes.

- Sample based conditioning is supported. Sample based conditioning conditionalizes on intervals and applies both the probabilistic and functional nodes.

- UNINET can model an empirical multivariate distribution by building a joint density function, which can be analytically conditioned.

Graphics and sensitivity analysis are supported in satellite software packages UNIGRAPH and UNISENS.

3.2 BBNs for Uncertainty Analysis / Stochastic Modeling

In uncertainty analysis/stochastic modeling, the analyst:

- defines a set of random variables,
- defines a joint distribution for these random variables

- defines other variables which are functions of the random variables
- Monte Carlo samples the entire joint distributions (probabilistic and functional variables
- Interprets and communicates the results.

When using UNINET to perform uncertainty analysis or stochastic modeling, the random variables are assigned marginal distributions, either chosen from a list of parametric distributions, or from a user defined distribution file. The user also specifies a DAG to capture conditional relations. Probabilistic influence between parent and child is represented as conditional rank correlation. Functional influence is defined with the help of an extended functional parser. A joint probability density for the probabilistic nodes is built using the joint normal copula to realize the dependence relations. This density can be updated analytically, or sampled and analyzed with Monte Carlo tools.

3.3 Analytic conditioning

The distinctive feature of UNINET is its analytical conditioning capability. BBNs with thousands of probabilistic nodes can be conditionalized on arbitrary values of random variables, whereby the conditional distribution is computed in a few minutes. This is only possible with the joint normal copula. Random variables are treated as transformations of joint normal variables. If $Z = (Z_1,\ldots Z_n)$ denotes a joint normal distribution of normal variables Z_i with mean zero and unit variance, then random variables $X_i,\ldots X_n$ are written as

$$(X_1,\ldots X_n) = (F_1^{-1}\Phi(Z_1),\ldots F_n^{-1}\Phi(Z_n))$$

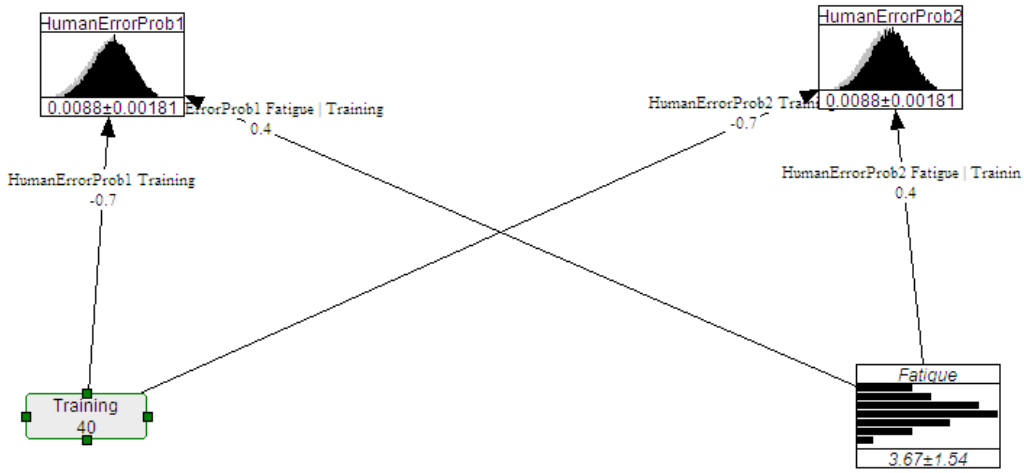
where F_i is the cumulative distribution function of random variable X_i, and Φ is the cumulative distribution function of the standard normal variable. If F_i is continuous invertible, this transformation is rank preserving, the rank correlations of X_i, X_j will be the same as that of Z_i,Z_j. Complications arise for discrete variables (see Hanea, [22]) The rank correlation of normal variables is computed from the product moment correlation with the Pearson transformation. For details see ([20])

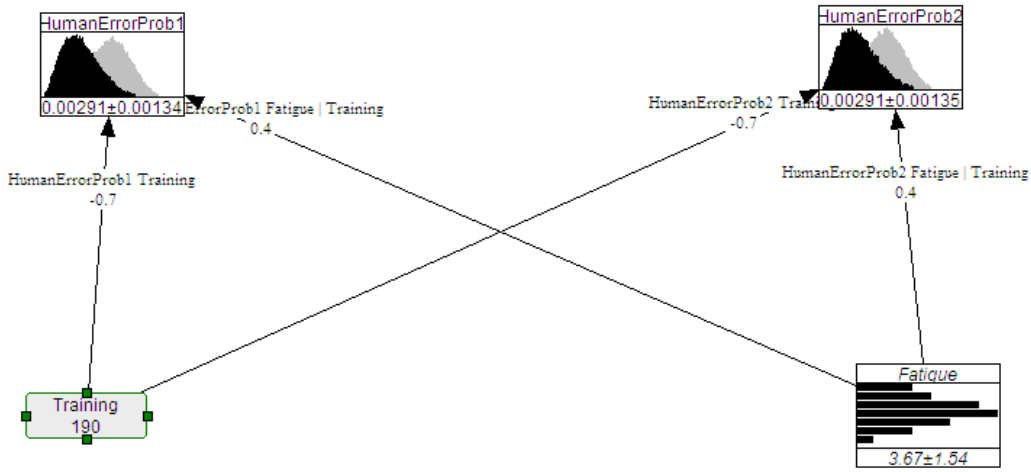
Conditioning on the value $X_i = x$ entails conditioning Z on $Z_i = \Phi^{-1} (F_i(x))$. Letting $Z(X_i=x)$ denote this conditional distribution, the *conditional distribution of* $X_1,\ldots X_n$ *, when* $X_1,\ldots X_n$ *are joined by the joint normal copula*, is given by

$$(X_1,\ldots X_n \mid X_i=x) = F_1^{-1}\Phi(Z_1 (X_i=x)),\ldots F_n^{-1}\Phi(Z_n (X_i=x)).$$

Note that Φ is the standard normal cumulative distribution function, but $Z_j(X_i=x)$ is not in general a standard normal variable, as its mean and variance are altered by the conditioning. For this reason, the *marginal distribution* of X_j given $X_i = x$ is no longer equal to the unconditional marginal $F_1^{-1}\Phi(Z_j)$, but has become $F_1^{-1}\Phi(Z_j (X_i=x))$.

The following two graphs illustrate by showing that the conditional standard deviation of HEP1 changes from 0.00181 to 0.00134 as we conditionalize on TR = 40 or TR = 190.





3.4 Ordinal Data Mining / Multivariate Density Modeling

In the previous chapter we discussed how a BBN might be quantified from data. In risk and reliability applications, the relevant multivariate data is often lacking. None the less, it is important to understand how a BBN would be built and quantified if data were available. This serves to underscore the point that a BBN model can be fully data driven if the data is available, and that structured expert judgment is used to fill in the data gaps. In most cases in CATS, the marginal distributions for the probabilistic nodes come from data, only the

dependence structure comes from experts. The features discussed here are unique to non-parametric continuous discrete BBNs, and to their implementation in UNINET, so the discussion focuses more on UNINET.

UNINET can be used to build a joint density for ordinal multivariate data. “Ordinal” means that the ordering of the values of the variables is meaningful. Ages, weights, IQ scores, income, body mass index, etc, are ordinal data. Street addresses, social security numbers, license plate numbers are not ordinal data. UNINET builds a joint distribution based on the empirical rank correlations and uses the joint normal copula. It assumes that the important dependences can be captured in dependences between the rankings of the variables.

The method will be illustraed on a very simple example, infact the Hunam Error Probability model of chapter 1 is used to generate multivariate data.

The multivariate data set must be in the form of an XL compliant comma separated file. An example of such a file, in XL, is shown below. The variable names are on the first row, each succeeding rows contains one sample from the multivariate distribution.

	A	B	C	D
1	HumanErrorProb1	HumanErrorProb2	Training	Fatigue
2	2.57E-04	4.64E-04	4.77E+02	6.00E+00
3	1.03E-02	2.01E-03	1.51E+02	5.00E+00
4	6.66E-04	1.09E-03	2.07E+02	5.00E+00
5	2.06E-03	2.13E-03	9.69E+01	6.00E+00
6	5.69E-06	5.19E-06	8.85E+02	1.00E+00

UNINET models an empirical multivariate distribution with a normal copula. The marginal distributions are taken directly from the data. The adequacy of such a model is evaluated statistically, by comparing an appropriate multivariate dependence measure for the empirical and the modeled distributions. The appropriate measure in this case is the determinant of the rank correlation matrix, which is closely related to the mutual information (Hanea et al 2008). We compare the determinant of the empirical rank correlation matrix (DER) with the determinant of the rank correlation matrix under the assumption that the rank dependencies are represented by the joint normal copula (DNR). If the hypothesis that the dependences come from a joint normal copula is not rejected at an appropriate significance level, then the UNINET model is appropriate. A BBN can be built by including the most important dependence relations. The adequacy of the BBN, relative to the ‘saturated’ graph, is again evaluated by comparing determinant of the rank correlation matrix based on the BBN (DBBNR) with the DNR. When a suitable BBN model is found, it can be analytically conditionalized from the empirical distribution, or alternatively the BBN density can be used to generate as many samples as desired, for further Monte Carlo post processing.

3.4.1 Rank correlation

The rank correlation of two continuous random variables *X,Y* is the product moment correlation of their respective quantile functions:

$\rho_r(X,Y) = \rho(F_X(X), F_Y(Y))$

Where F_X, F_Y are the cumulative distribution functions of *X, Y* respectively.

3.4.2 Empirical rank correlation

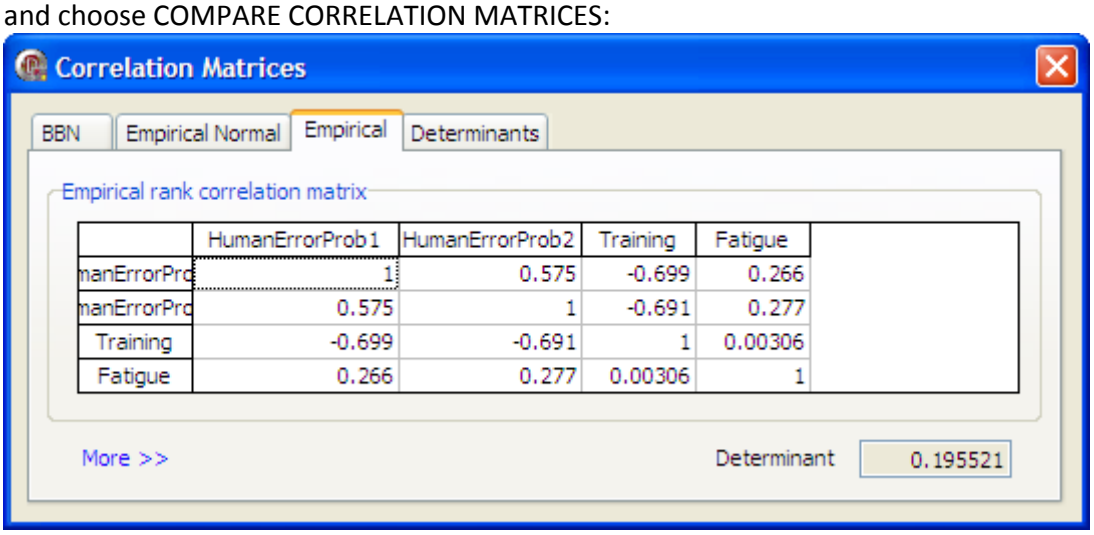
To determine the empirical rank correlation of a set $(x_1, y_1)\ldots (x_n, y_n)$ of n bivariate observations of *X* and *Y*, we first construct the *empirical distribution functions*:

$F_X(r) = \#\{x_i \leq r\}/ (n+1);\ \ F_Y(r) = \#\{y_i \leq r\}/ (n+1).$

The empirical rank correlation of *X* and *Y* is then the product moment correlation of $F_X(X)$ and $F_Y(Y)$:

$\rho_r(X,Y) = \rho(F_X(X), F_Y(Y))$

To view the empirical rank correlation matrix and the DER, select the DATA menu option and choose COMPARE CORRELATION MATRICES:



	HumanErrorProb1	HumanErrorProb2	Training	Fatigue
manErrorPro	1	0.575	-0.699	0.266
manErrorPro	0.575	1	-0.691	0.277
Training	-0.699	-0.691	1	0.00306
Fatigue	0.266	0.277	0.00306	1

Note tat HEP1 and HEP2 have empiical rank correlation 0.575; this is caused by their common anscesors in the BBN. Training and Fatigue are independent, the correlation 0.00306 is noise.

3.4.3 Joint normal copula

If $Z = (Z_1,\ldots Z_n)$ denotes a joint normal distribution of normal variables Z_i with mean zero and unit variance, then random variables $X_i,\ldots X_n$ are connected by the *joint normal copula* if

$$(X_1,\ldots X_n) = (F_1^{-1}\Phi(Z_1),\ldots F_n^{-1}\Phi(Z_n))$$

where F_i is the cumulative distribution function of random variable X_i, and Φ is the cumulative distribution function of the standard normal variable. If F_i, F_j are continuous invertible, the transformation $Z_i \rightarrow X_i$ is rank preserving, and the rank correlations of X_i, X_j will be the same as that of Z_i, Z_j. If X_i is discrete then rank correlation is not perfectly preserved, and the degree of discretization influences the lack of preservation (see Hanea 2008). The rank correlation of normal variables is computed from the product moment correlation with the Pearson transformation:

$$6 \arcsin(\rho(Z_i, Z_j) /2) / \pi = \rho(\Phi(Z_i), \Phi(Z_j)) = \rho_r(Z_i, Z_j)$$

Where ρ_r rank correlation and ρ denotes the product moment correlation:

$$\rho(Z_i, Z_j) = (E(Z_iZ_j) - E(Z_i) E(Z_j)) / (VAR(Z_i) VAR(Z_j))^{1/2}$$

Whereas every rank correlation between two variables can be realized as the rank correlation of a bivariate normal distribution, in higher dimensions this is NOT true. For details see (Kurowicka and Cooke, 2006-A).

3.4.4 Empirical normal rank correlation

If X_i denotes a random variable with empirical distribution function F_i, then

$$Z_i(X_i) = \Phi^{-1} (F_i(X_i))$$

is the *empirical normal version* of X_i . Applying the Pearson transformation, we find the *empirical normal rank correlation* of X_i, X_j as

$$6 \arcsin(\rho(Z_i(X_i), Z_j(X_j)) /2) / \pi.$$

The empirical normal rank correlation of X_i, X_j is not in equal to the empirical rank correlation of X_i, X_j unless X_i and X_j are joined by the normal copula.

3.4.5 BBN rank correlation

When a graphical structure is specified the rank correlation of all variables is computed assuming the conditional independence implied by the graph, and assuming that any two variables have the rank correlation induced by the joint normal copula.

3.4.6 Empirical rank determinant DER

Where F_i is the empirical cumulative distribution function of variable X_i , $i = 1…N$; the empirical rank determinant is

$$DER = | [\rho(F_i(X_i), F_j(X_j)]_{i,j = 1…N} |$$

where | | denotes the determinant. The determinant of a correlation matrix measures the "amount of linear dependence" between the variables. Its value is 1 if the variables are independent, and zero if one of the variables can be written as a linear combination of the others.

3.4.7 Empirical normal rank determinant DNR

The empirical normal rank determinant is the determinant of the rank correlation matrix of the empirical normal version of X_i, $i = 1,…N$:

$$DNR = | \{6 \arcsin(\rho(Z_i(X_i), Z_j(X_j)) /2) / \pi\}_{i,j = 1…N}$$

Where | | denotes the determinant. DNR is the determinant of the *saturated BBN*, that is, the BBN in which there is an arc between every pair of variables.

3.4.8 BBN rank determinant DBBNR

The BBN rank determinant is the determinant of the BBN rank correlation matrix. Unlike DNR, DBBNR reflects the conditional independence relations imposed by the BBN. If the BBN is not saturated, then DBBNR > DNR. DBBNR is approximated as:

$$DBBNR = \Pi_{\text{all conditional rank correlations } \rho r \text{ in BBN}} (1-\rho_r^2).$$

3.4.9 Pitfalls

The user should be aware of a number of pitfalls in using the DNR or DBBNR as a test statistic for assessing the adequacy of a BBN model.

i) Shrinking determinant

If the number of variables exceeds the number of samples in an empirical multivariate data set, then the DER will be zero. If some of the variables have repeated values, then this can occur even if the number of samples exceeds the number of variables. If the number of variables is large though not in excess of the number of samples, then DER can become very small.

ii) Discrete variables

When rank correlations are specified for discrete variables in stochastic modeling, the rank correlation realized is the rank correlation between two uniform variables. A discrete variable is realized by $F^{-1}(U)$, where F is the cumulative distribution function of the discrete variable. F^{-1} is thus a many-one transformation. This will cause empirical rank correlations involving this discrete variable to differ from the stipulated rank correlations (see Hanea 2008). In judging statistical adequacy, the user must take this into account by assuming that the discrete variable is a many-one transform of a normal variable. Otherwise, the lack of continuity will be treated as lack of fit of the joint normal copula.

iii) Directionality

It is impossible to infer directionality of influence from multivariate data. Insight into the causal processes generating the data should be used, whenever possible, in constructing a BBN model. Because of this fact, there are different BBNs that are wholly equivalent, and many non-equivalent BBNs may provide statistically acceptable models of a given multivariate ordinal data set.

3.5 Assessing model adequacy

The question of model adequacy is answered by defining a statistical measure for multivariate dependence and testing whether the hypothesis that the multivariate dependence could come from the joint normal copula should be rejected.

The UNINET approach to this question is based on the *mutual information* as a measure of multivariate dependence (Joe, 1989). The mutual information of a joint density $f(x_1, x_n)$ with marginal densities $f_1,...f_n$ is :

$$\int \ln(f(x_1,...x_n)/f_1(x_1)...f_n(x_n))\ f_1(x_1)...f_n(x_n)\ dx_1,...dx_n.$$

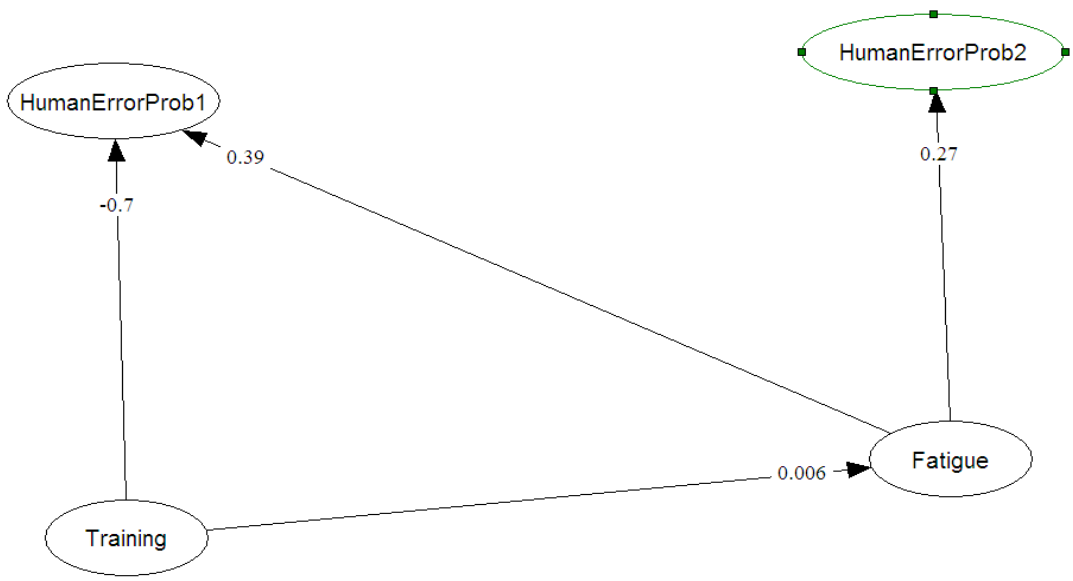
For a joint normal distribution the mutual information is given by exp(-2 ln |C|) where |C| is the determinant of the correlation matrix. Mutual information is invariant under monotone univariate transformations; therefore, this is also the mutual information of any distribution based on the joint normal copula. The determinant |C| is closely related to the determinant

|R| of the normal rank correlation matrix (DNR). To evaluate the suitability of a joint normal copula it is convenient to compare DNR with the determinant of the empirical rank correlation matrix DER. We could also compare |C| with the "empirical mutual information" of the multivariate data set, were it not that the computation of the "empirical mutual information" imposes obstacles which are as yet unsurmounted.

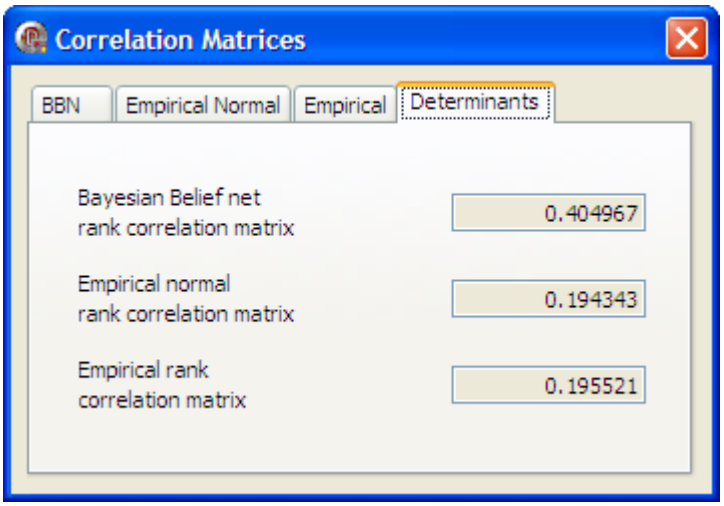
If the joint normal copula is a statistically acceptable model of the data, then the user can simplify the model by removing arcs whose conditional rank correlation is very close to zero, as these will not contribute significantly to the DNR. By considering the determinant of the rank correlation matrix based on the BBN (DBBNR) one can determine whether the BBN is a statistically acceptable model of the empirical normal rank correlation matrix.

The statistical tests are not currently supported in UNINET.

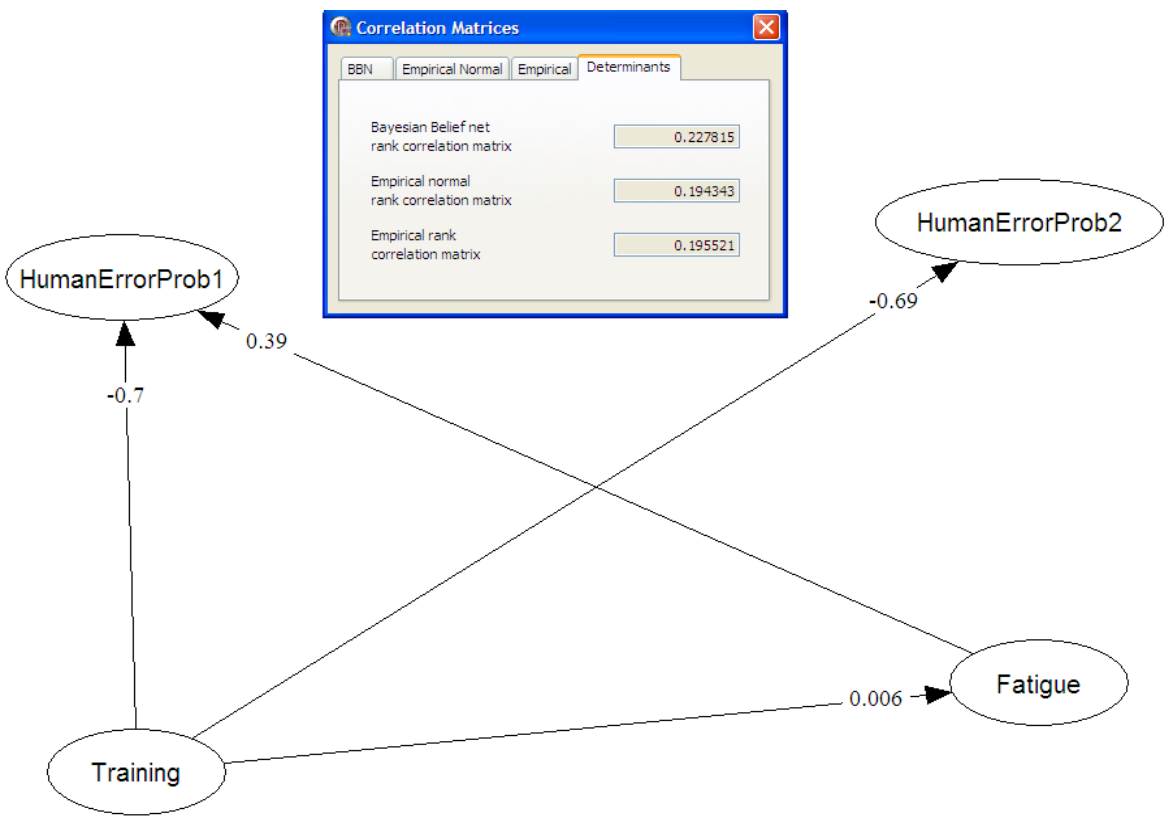
The ideas are illustrated with the BBN model created to explain the multivariate data. Suppose we believed that Training could influence HEP1 directly, but could influence HEP2 only by influencing the variable Fatigue. We would propose the following BBN.



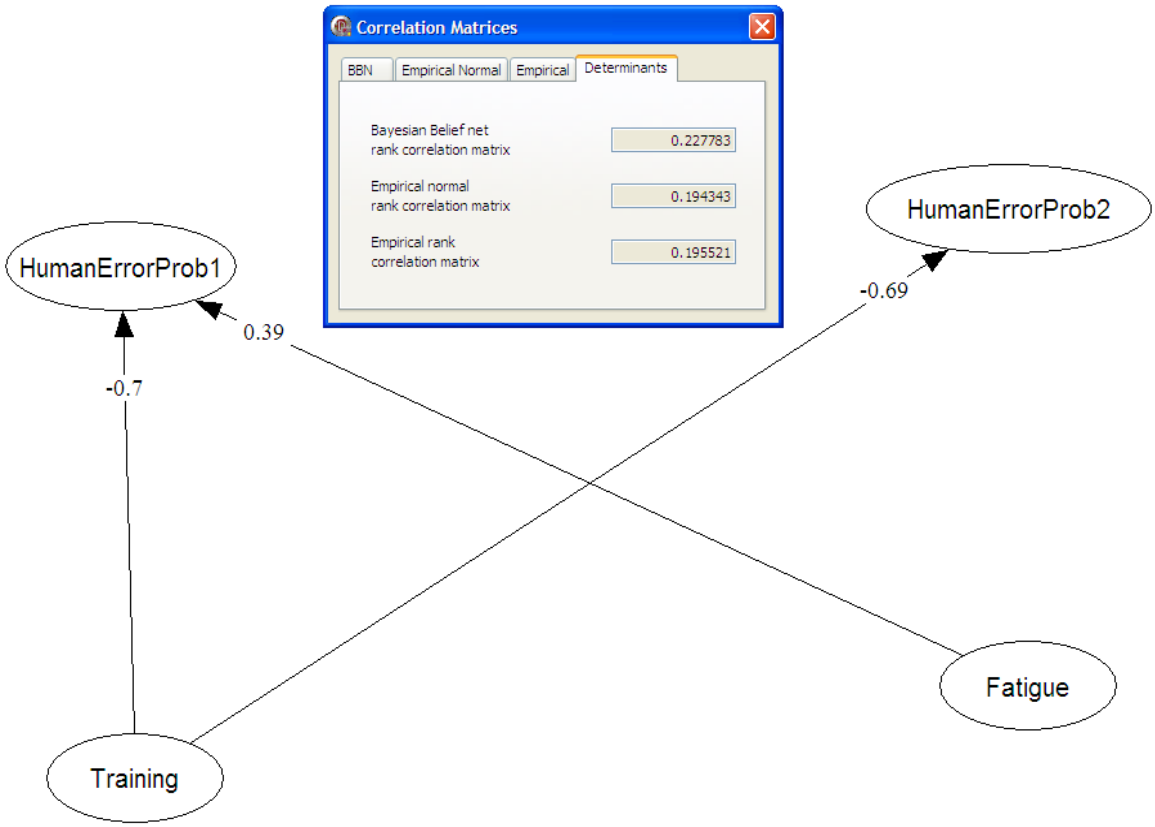
To check whether this represents a good model for our data, we compare the DBBN with DER:



The DBBN = 0.404967, while DER = 0.195521, In this case the difference is statistically too large. (Of course we know that the data was generated with the model from Chapter 1; this model used the normal copula, which explains why DER and DNR are equal.) We remove the arc from Fatigue to HEP2 and draw an arc from Training to HEP2:



The fit has improved, since more of the dependence in the data is captured in the BBN. The arc from Training to Fat captures very little dependence; its removal would hardly affect the DBBN:



In general, it will be impossible to recover exactly a model which generated the data, from the data alone. We have statistical test to determine whether a model is not rejected by the data, but there may be many non-rejected models. This situation is common in statistical inference.

4 Human reliability models.

Three Homan reliability models have been built and quantified in collaboration with NLR the models and results from the expert elicitation are presented briefly next. More information about the different models may be found in other sources (see appendix 1 and references)

4.1 The flight crew performance model.

The Flight Crew Performance model (FCP) was the first of the generic models that have been developed by the CATS consortium to represent dependence between base events in the FTs of the CATS model. This section is devoted to the description of the model, results from the expert elicitation and discussion.

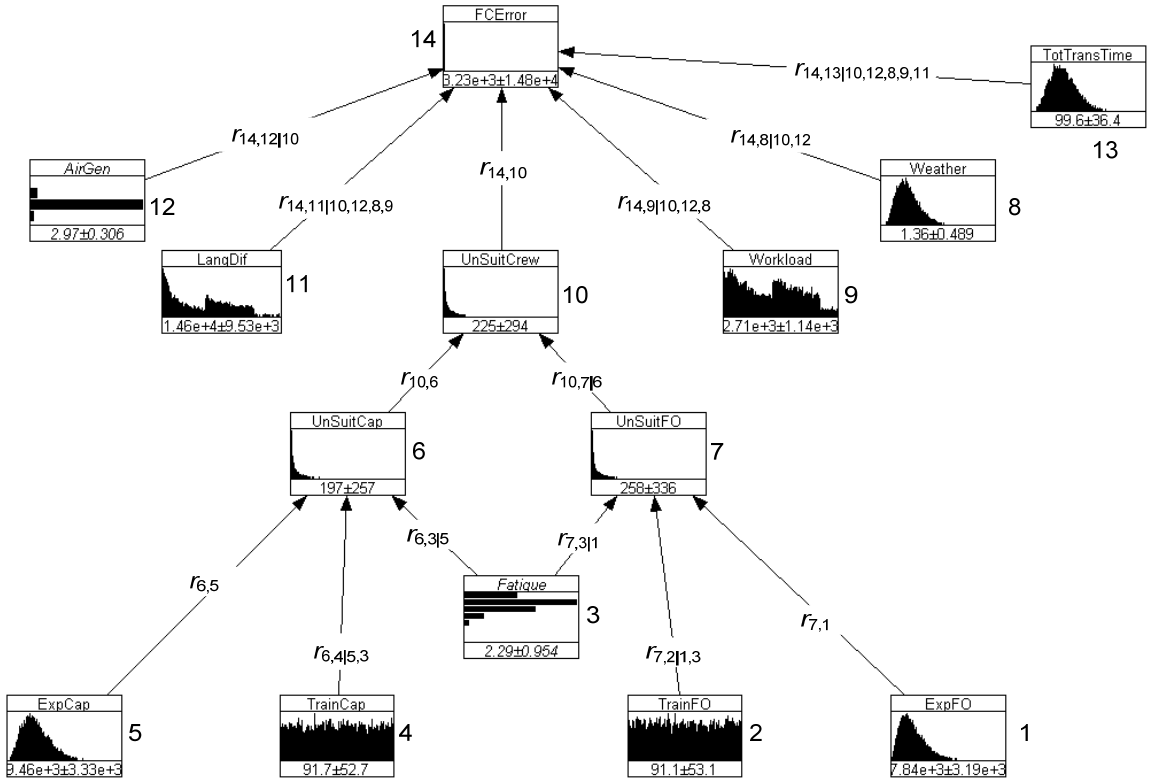


Figure 4.1 BBN representing the Flight Crew performance Model

The BBN representing the FCP model is shown in Figure 4.1. The model has been described in detail in [17] and hence only a brief description is included here.

Node #	Definition	Marginal distribution source
1	Total number of hours flown since the pilot's license obtaining by first officers.	Data
2	Number of days passed since last recurrence training for First Officers.	Data
3	Stanford Sleepiness Scale. 1 signifies "feeling active and vital; wide awake" and 7 stands for "almost in reverie; sleep onset soon; struggle to remain awake".	Data
4	Number of days passed since last recurrence training for Captains.	Data
5	Total number of hours flown since the pilot's license obtaining by captains.	Data
6	Number of Captains failing their proficiency check test per 10,000	SEJ[2]
7	Number of First Officers failing their proficiency check test per 10,000	SEJ
8	Rainfall rate (mm/hr) translated into airborne weather radar in a cockpit.	Data
9	Number of flights in which the pilot and first officer will have a different mother tongue per 100,000	SEJ
10	Number of Captains or/and First Officers failing their proficiency check test per 10,000	SEJ
11	Aircraft generation is a scale from 1 to 4 where 4 is the most recent generation of aircrafts.	Data
12	Number of times the crew members have to refer to the abnormal/emergency procedures section of the aircraft operation manual during flight per 100,000 flights.	Data
13	Total duration (in seconds) of the air/ground communications, per aircraft, for the approach and landing flight phase.	Data
14	Number of errors per 1e6 flights. A distribution is fitted to data provided by DNV on average number of human errors per million flights and some of the percentiles of the error distribution.	DNV FT

Table 4.1 Description of variables from the Model in

The variables represented as nodes in **Error! Reference source not found.** are briefly described in Table 4.1. The basis for the quantification of each marginal distribution is presented in column 3. Four variables come from data and the rest were elicited through structured expert judgment. Node 13 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 13 (corrected according to the own marginal distribution as computed in the FTs) will appear in the CATS model.

[2] Structured expert judgment.

An elicitation protocol was designed for obtaining the marginal distributions shown in Table 4.1 and the dependence information required by the model. A total of 4 marginal distributions, 11 questions for retrieving the dependence information and 8 calibration variables were asked to each expert[3]. Summary results from the classical method are presented in table. Calculations are performed with the EXCALBIUR software developed at the TU Delft.

Results of scoring experts and Relative Information to the DM
Bayesian Updates: no Weights: global DM Optimization: yes
Significance Level: 0.6638 Calibration Power: 1

Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight	Normaliz.we without DM	Normaliz.we with DM
C	0.001547	1.016	0.9689	8	0	0	0
A	0.02651	0.7119	0.4991	8	0	0	0
D	0.185	1.317	1.029	8	0	0	0
B	0.6638	0.95	0.574	8	0.381	1	0.5
E	5.115E-005	1.049	1.06	8	0	0	0
GWDM	0.6638	0.95	0.574	8	0.381		0.5

(c) 1989-2005 TU Delft

Table 4.2. Expert's and GW Decision Maker's Performance.

Table 4.2 shows the resulting scores for the five experts in this study. The first column gives the expert's id; the second column gives the calibration score. The ratio of highest to lowest score is about 1.30E+04. 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 it is 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 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).

Table 4.2 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. The recommended choice for the DM is the GWDM as it achieves better performance than the EWDM and the GWDM without optimization combinations.

[3] In total 14 rank correlations are required, 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. $r_{13,14|10,12,8,9,11}$ was elicited later from a single expert who is not a pilot but rather a risk analyst at NLR. Later versions of the model will incorporate more experts.

To elicit the rank correlations a total of 11 questions were asked to each expert. These were similar to those in relation (A2.1) in Appendix 2. Previously 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 4.3.

Rank Correlation	Value	Rank Correlation	Value
$r_{7,1}$	-0.95	$r_{10,7\|6}$	1
$r_{7,3\|1}$	0.86	$r_{14,10}$	0.3
$r_{7,2\|1,3}$	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	$r_{13,14\|10,11,8,12,9}$	0.16

Table 4.3. GWDM Dependence Information

4.2 The ATC performance model.

The Air Traffic Control Performance model (ATCP) is the second one of the generic models that has been developed by EWI-NLR to represent dependence between base events in the FTs of the CATS model. The model is presented more extensively in [31]. This section is devoted to the description of the model and results from the expert elicitation. The model is shown in figure 4.2 and the variables taken into account are briefly described below (table 4.4) according to their labeling in figure 4.2.

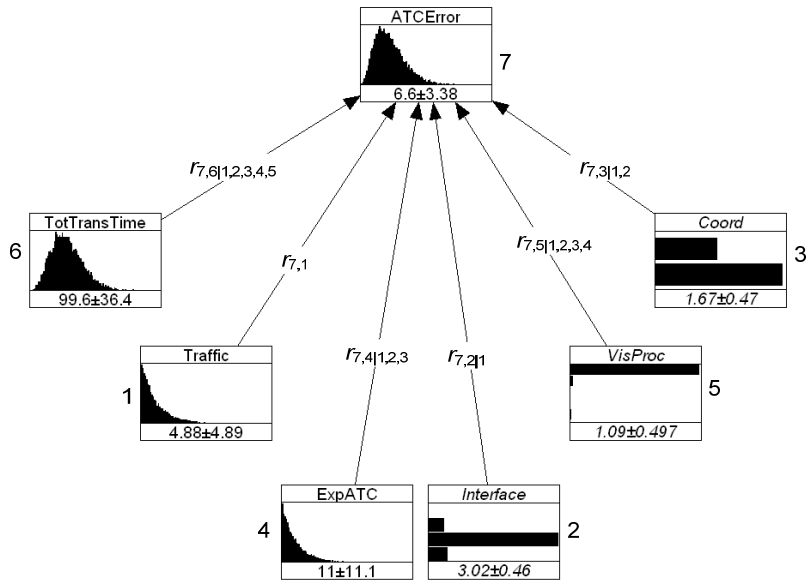


Figure 4.2 BBN representing the ATC performance Model

Node #	Definition	Marginal distribution source
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 ATCos takes place in the same room, 2 - The communication with other ATCos does not take place in the same room.	Data
4	Number of years working as an ATCo in the same position.	Data
5	Five states variable. From 1 - normal operations to 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	Number of errors per 1e6 flights. A distribution is fitted to data provided by DNV on average number of human errors per million flights and some of the percentiles of the error distribution.	DNV FT

Table 4.4 Description of variables from the ATC Model

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 6 would represent a base event in the Fault Trees. Whenever the ATC performance is of interest in the FTs an instance of node 6 (corrected according to the own marginal distribution as computed in the FTs) will appear in the CATS model.

An elicitation protocol was designed for obtaining the dependence information required by the model shown in figure 4.2. A total of 5 questions for retrieving the dependence information and 12 calibration variables were asked to 6 experts[4]. 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. Calculations are performed with the EXCALIBUR software developed at the TU Delft.

Results of scoring experts and Relative Information to the DM
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.00131 Calibration Power: 1

Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight	Normaliz.we without DM	Normaliz.we with DM
A	0.1012	0.5633	0.5034	10	0.05095	0.5208	0.2015
B	0.04706	1.03	0.9588	10	0.04512	0.4612	0.1784
C	0.00131	1.423	1.349	10	0.001767	0.01806	0.006987
D	2.795E-009	1.669	1.655	10	0	0	0
E	2.501E-006	1.017	0.9624	10	0	0	0
GWDM	0.6827	0.3094	0.2271	10	0.1551		0.6131

(c) 1989-2005 TU Delft

Table 4.5. Expert's and GW Decision Maker's Performance.

Table 4.5 shows the resulting scores for the five experts in this study plus one Decision Maker. 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 3.62E+07 (1.30E+04 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 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 sixth column gives the "un-normlaized 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. The equal weight decision maker (EWDM) is better calibrated than each expert individually. However information scores derived from the EWDM are poor. They are the lowest amongst all experts (that is including the EWDM as an expert) in both all variables and calibration questions alone.

[4] All experts are ATC controllers and hence 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. $r_{7,6|1,2,3\,4\,5}$ was elicited later from a single expert who is not an ATC but rather a risk analyst at NLR. Later versions of the model will incorporate more experts.

For the GWDM all experts with a calibration score less than the significance level (0.00131) found by the optimization procedure are outweighed as reflected by the zeros in columns 6, 7 and 8. 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 item weights decision maker 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.

As stated before, to elicit the rank correlations in figure 4.2 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 value[5]. 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 Appendix 2 and relation (A2.3). The combination of the three expert's individual assessments was done as described in AND [16]. The results of the combination scheme are presented in table 4.6 bellow.

Probability[6]		(Un)conditional rank correlation	
P_1	0.48	$r_{7,4\|1,2,3}$	-0.060
P_2	0.44	$r_{7,1}$	-0.179
P_3	0.47	$r_{7,2\|1}$	-0.210
P_4	0.58	$r_{7,3\|1,2}$	0.180
P_5	0.52	$r_{7,5\|1,2,3,4}$	0.020
-	-	$r_{7,6\|1,2,3,4,5}$	0.180

Table 4.6. GWDM Dependence Information for the ATC model

[5] 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.

[6] P_1 = P(ATC error ≥ median | Experience ≥ median), P_2 = P(ATC error ≥ median | Traffic ≥ median), P_1 = P(ATC error ≥ median | MMI ≥ 4), P_1 = P(ATC error ≥ median | Communication & Coordination = 2), P_1 = P(ATC error ≥ median | Visibility Procedure ≥2). $r_{7,6|1,2,3\,4\,5}$ was elicited later from a single expert who is not an ATC but rather a risk analyst at NLR. Later versions of the model will incorporate more experts.

4.3 The maintenance crew performance model.

The Air Traffic Control Performance model (ATCP) is the third and last of the generic models that have been developed by the EWI-NLR 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 model is shown in figure 4.3 and the variables taken into account are briefly described below (table 4.7) according to their labeling in figure 4.3.

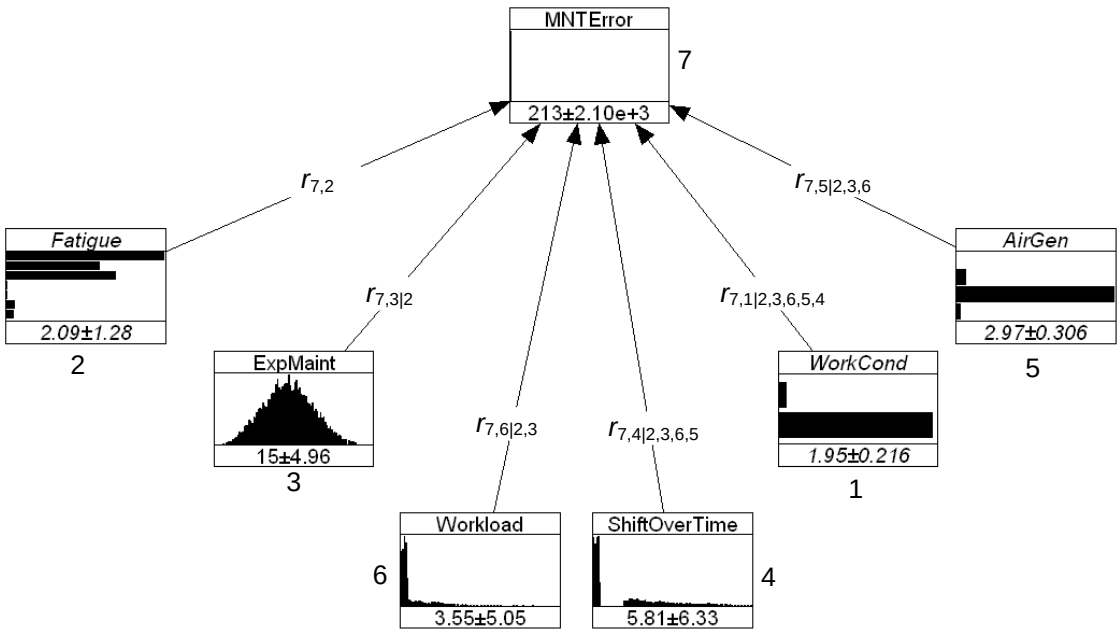


Figure 4.3 BBN representing the maintenance crew performance Model

Unlike the previous two models, only one expert was readily available for the quantification of the model[7]. The meeting with this single expert, who is a maintenance engineer for NLR, took place on May the 28th 2008. The expert was asked 21 questions in total, where 7 questions were used to assign marginal distributions and 11 to elicit calibration variables. The rest were used to obtain the rank and conditional rank correlations required (Figure 5). The expert was asked to specify his 5%, 50% and 95% quantile of his uncertainty distribution for each variable of interest. Since a single expert was used, no combination was required. The results of the dependence elicitation are shown in table 4.8.

[7] It is desirable to have more experts quantifying the model for CATS II.

Node #	Definition	Marginal distribution source
1	Whether the work is performed at the ramp (outside - 1) or in the hangar (inside - 2)	SEJ
2	Stanford sleepiness scale, where 1 is feeling active, vital, alert or wide awake and 7 is no longer fighting sleep; having dream-like thoughts	SEJ
3	Years in current position	Data
4	Time available to transfer a job (min)	SEJ
5	Aircraft generation is a scale from 1 to 4 where 4 is the most recent generation of aircrafts.	Data
6	Estimated delay in release of the aircraft (hrs)	SEJ
7	Number of errors per demand. A distribution is fitted to data provided by DNV on average number of human errors per million flights and some of the percentiles of the error distribution.	DNV FT

Table 4.7 Description of variables from the maintenance crew performance model

Probability[8]		(Un)conditional rank correlation	
P_1	0.7	$r_{7,2}$	0.23
P_2	0.423	$r_{7,3\|2}$	-0.24
P_3	0.538	$r_{7,6\|2,3}$	0.12
P_4	0.435	$r_{7,5\|2,3,6}$	-0.07
P_5	0.477	$r_{7,4\|2,3,6,5}$	-0.07
P_6	0.48	$r_{7,1\|\ 2,\ 3,\ 6,\ 5,\ 4}$	-0.02

Table 4.8. Dependence Information for the Maintenance model

[8] P_1 = P(MNT Error > median | Fatigue > 4), P_2 = P(MNT Error > median | Experience > median), P_3 = P(MNT Error > median | Workload > median), P_4 = P(MNT Error > median |Aircraft generation = 4), P_5 = P(MT Error > median |Shift overlap time > median), P_6 = P(MNT Error > median | Working condition = 1)

5. The CATS Bayesian belief net.

As previously mentioned, the CATS model integrates ESDs, FTs, and BBN into one single continuous-discrete BBN. This section is devoted to a description of the procedure to build up the CATS model. The three models presented in sections 4.1 – 4.3 are represented for 3 different flight phases; these are Take Off (TO) En-Route (ER) and Approach & Landing (AL). In cite DNV 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 cite DNV.

5.1 ESDs and FTs for the CATS model

The quantification of ESDs is presented in [18], this is used in [32] to quantify the FTs that later compute the accident probability. The FTs (and consequently ESDs) presented in [32] are translated into functional nodes as explained in section 2.7.

In figure 5.1 one may see ESD1 aircraft system failure for the TO flight phase. In total four AND gates and four OR gates represent the FT and ESD. Fifteen base events are influenced by the maintenance performance model presented in section 4.3 and 3 base events by the flight crew performance model from subsection 4.1. No influence of the ATC performance model is observed in this particular FT. To translate this information into a BBN the process is:

- Find a distribution of the probability of base events per demand according to their variability (see next section)
- Connect underneath the corresponding dependence model from subsections 4.1-4.3 with the corresponding dependence info. These nodes will be ancestors of base events in the Fault Trees
- Write down in descendent nodes of each base event the arithmetic formulas that translate a FT into a BBN (subsection 2.7). These will be functional nodes in the BBN

These steps are repeated for all 31 ESDs presented in table 5.1

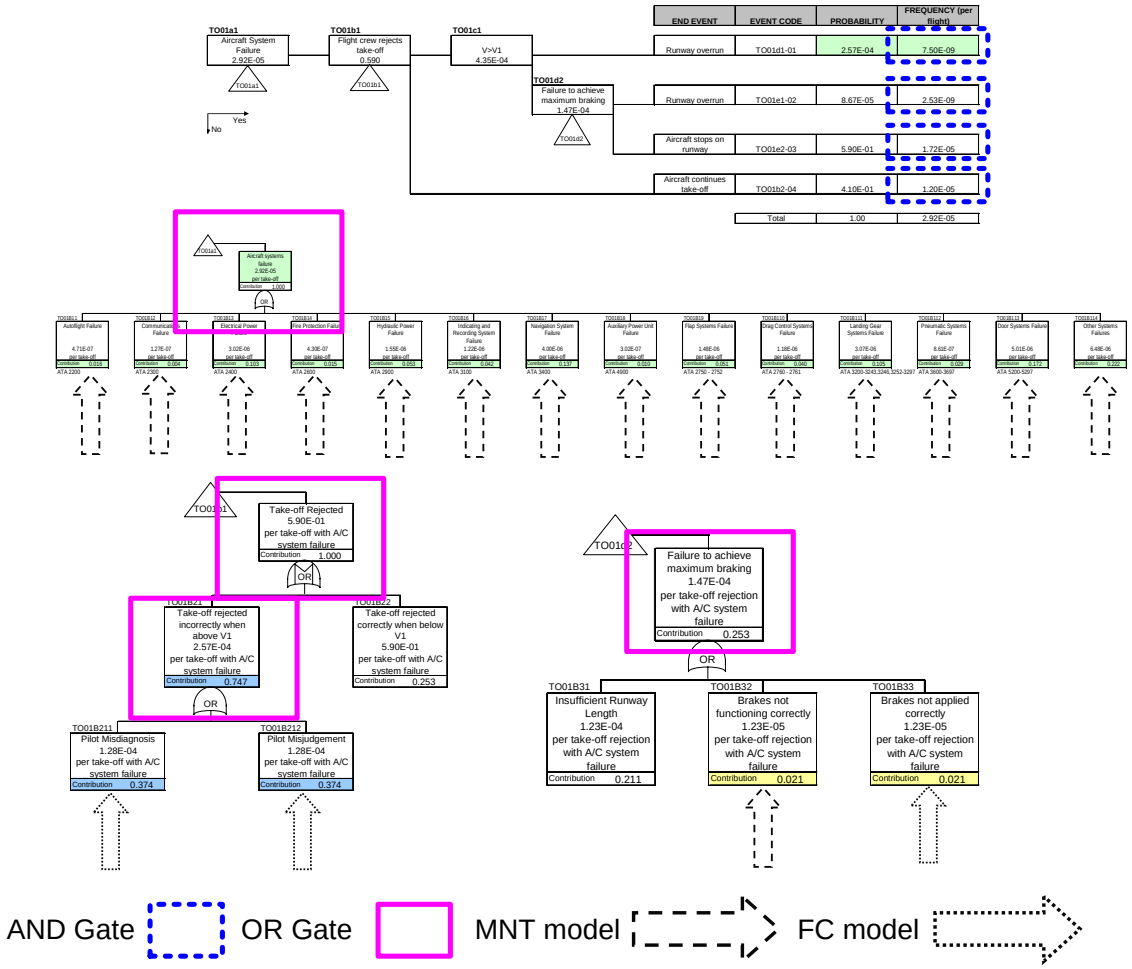


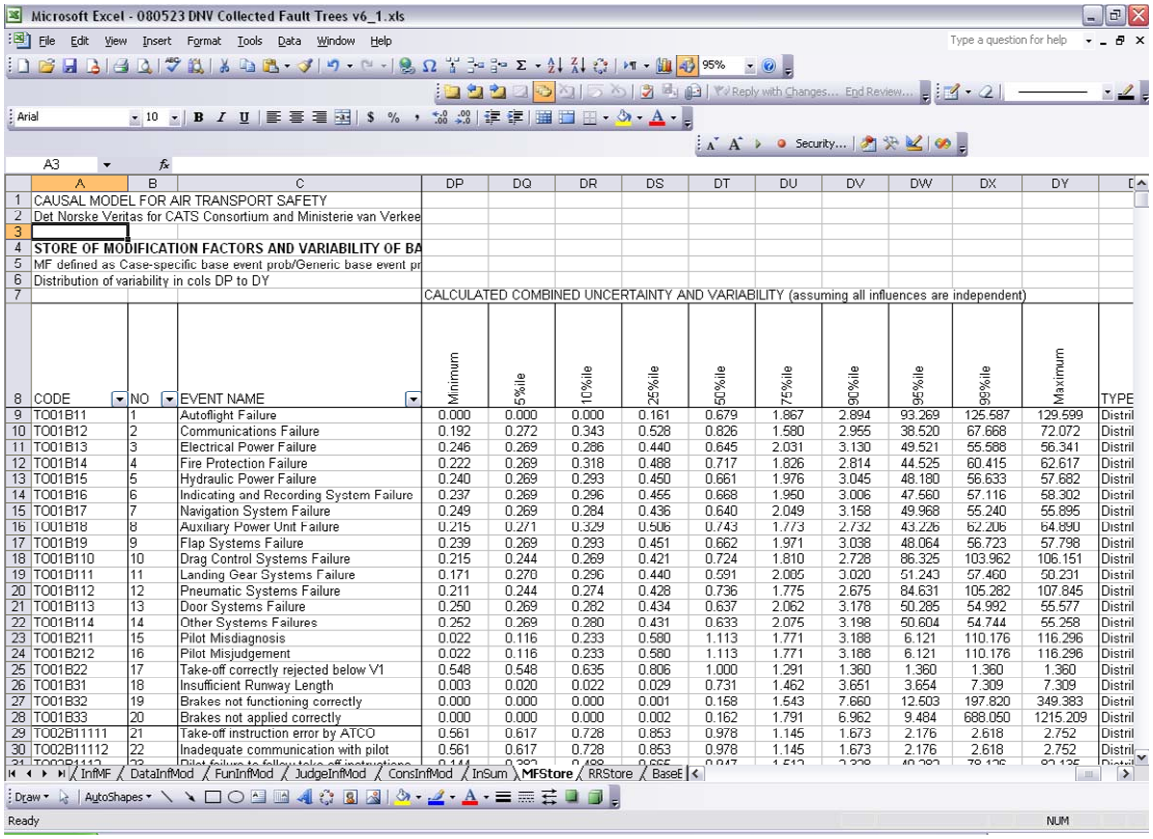
Figure 5.1 ESD1 Aircraft system failure and its corresponding FT showing.

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

Table 5.1. Thirty one ESDs represented in the current version of the CATS model

5.2 The error distributions

Base events in the FTs are influenced by the human performance models presented in section 4. In fact, the probabilities presented in DNV's FTs represent the expected probability of a given human error[9]. Other percentiles over the distribution of error probability are also given by DNV.



CAUSAL MODEL FOR AIR TRANSPORT SAFETY
Det Norske Veritas for CATS Consortium and Ministerie van Verkee
STORE OF MODIFICATION FACTORS AND VARIABILITY OF BA
MF defined as Case-specific base event prob/Generic base event pr
Distribution of variability in cols DP to DY

CALCULATED COMBINED UNCERTAINTY AND VARIABILITY (assuming all influences are independent)

CODE	NO	EVENT NAME	Minimum	5%ile	10%ile	25%ile	50%ile	75%ile	90%ile	95%ile	99%ile	Maximum	TYPE
TO01B11	1	Autoflight Failure	0.000	0.000	0.000	0.161	0.679	1.867	2.894	93.269	125.587	129.599	Distri
TO01B12	2	Communications Failure	0.192	0.272	0.343	0.528	0.826	1.580	2.955	38.520	67.668	72.072	Distri
TO01B13	3	Electrical Power Failure	0.246	0.269	0.286	0.440	0.645	2.031	3.130	49.521	55.588	56.341	Distri
TO01B14	4	Fire Protection Failure	0.222	0.269	0.318	0.488	0.717	1.826	2.814	44.525	60.415	62.617	Distri
TO01B15	5	Hydraulic Power Failure	0.240	0.269	0.293	0.450	0.661	1.976	3.045	48.180	56.633	57.682	Distri
TO01B16	6	Indicating and Recording System Failure	0.237	0.269	0.296	0.455	0.668	1.950	3.006	47.560	57.116	58.302	Distri
TO01B17	7	Navigation System Failure	0.249	0.269	0.284	0.436	0.640	2.049	3.158	49.968	55.240	55.895	Distri
TO01B18	8	Auxiliary Power Unit Failure	0.215	0.271	0.329	0.506	0.743	1.773	2.732	43.226	62.206	64.890	Distri
TO01B19	9	Flap Systems Failure	0.239	0.269	0.293	0.451	0.662	1.971	3.038	48.064	56.723	57.798	Distri
TO01B110	10	Drag Control Systems Failure	0.215	0.244	0.269	0.421	0.724	1.810	2.728	86.325	103.962	106.151	Distri
TO01B111	11	Landing Gear Systems Failure	0.171	0.270	0.296	0.440	0.591	2.065	3.020	51.243	57.460	50.231	Distri
TO01B112	12	Pneumatic Systems Failure	0.211	0.244	0.274	0.428	0.736	1.775	2.675	84.631	105.282	107.845	Distri
TO01B113	13	Door Systems Failure	0.250	0.269	0.282	0.434	0.637	2.062	3.178	50.285	54.992	55.577	Distri
TO01B114	14	Other Systems Failures	0.252	0.269	0.280	0.431	0.633	2.075	3.198	50.604	54.744	55.258	Distri
TO01B211	15	Pilot Misdiagnosis	0.022	0.116	0.233	0.580	1.113	1.771	3.188	6.121	110.176	116.296	Distri
TO01B212	16	Pilot Misjudgement	0.022	0.116	0.233	0.580	1.113	1.771	3.188	6.121	110.176	116.296	Distri
TO01B22	17	Take-off correctly rejected below V1	0.548	0.548	0.635	0.806	1.000	1.291	1.360	1.360	1.360	1.360	Distri
TO01B31	18	Insufficient Runway Length	0.003	0.020	0.022	0.029	0.731	1.462	3.651	3.654	7.309	7.309	Distri
TO01B32	19	Brakes not functioning correctly	0.000	0.000	0.000	0.001	0.158	1.543	7.660	12.503	197.820	349.383	Distri
TO01B33	20	Brakes not applied correctly	0.000	0.000	0.000	0.002	0.162	1.791	6.962	9.484	688.050	1215.209	Distri
TO02B11111	21	Take-off instruction error by ATCO	0.561	0.617	0.728	0.853	0.978	1.145	1.673	2.176	2.618	2.752	Distri
TO02B11112	22	Inadequate communication with pilot	0.561	0.617	0.728	0.853	0.978	1.145	1.673	2.176	2.618	2.752	Distri

Figure 5.2 Percentiles of the FT's base event probability distribution from DNV

Figure 5.2 says that the minimum value that the base event probability TO01B11 is 0*E(TO01B11)=0 where E(X) denotes the expectation of X. In the same way the maximum value of TO01B11 should be equal to 129.599*E(TO01B11). Other quantiles may be read in the same way. This information is used to fit a parametric distribution to this data to represent the distribution over the base event probability. In total there are 743 basic events considered over all thirty one ESDs from table 5.1 in the model as is presented in Figure 1.1.

[9] There might be events which do not represent a human error but rather something else going wrong in the aviation system.

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 shown in Figure 6 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 could be obtained.*

The fitting procedure to obtain the parametric distribution for each base event is described briefly next:

- Obtain the minimally informative distribution (with respect to log uniform background measure to avoid negative values) that fits this information. As stated previously, this distribution is a distribution over the probability of error. Observe that the minimally informative solution will always capture the percentiles specified by DNV, however, still might give inconsistent results. For example, part of the distribution may be outside the (0,1) interval and this distribution will not have in general the expectation provided by DNV as the probability of base events in the FTs. For this reasons we then,
- Find the parameters of a Weibull, Gamma, Log-normal or beta distribution that minimizes the sum of squared difference between the minimally informative solution found in step 1 and the parametric distribution such that,
    - The expectation of the parametric distribution is equal to DNVs estimate,
    - The distribution lies in the interval (0,1) and,
    - The 0.9999999999999999 percentile of the distribution is less or equal than the maximum value provided by DNV.

The procedure described above was applied in a first pass to ESD1, ESD26 and ESD19. Naturally, the results of the fitting procedure vary widely across base events. In general it was observed that Weibull and Gamma characterized better the data provided by DNV and at the end to improve speed the procedure described above was applied searching for Gamma and Weibull parameters only to the 743 base events provided by DNV.

Where no convergence was possible, a search on Beta and LogNormal parameters was also implemented. In general, since the mean and maximum value of DNVs data entered as constraints in the optimization procedure described above they are well captured by

the fitted distributions. However some fits may still be far from DNVs quintile data. Two examples are presented in figures 5.3 and 5.4 for illustration purposes only. In both cases the Weibull distribution was minimum in terms of the sum of squared difference.

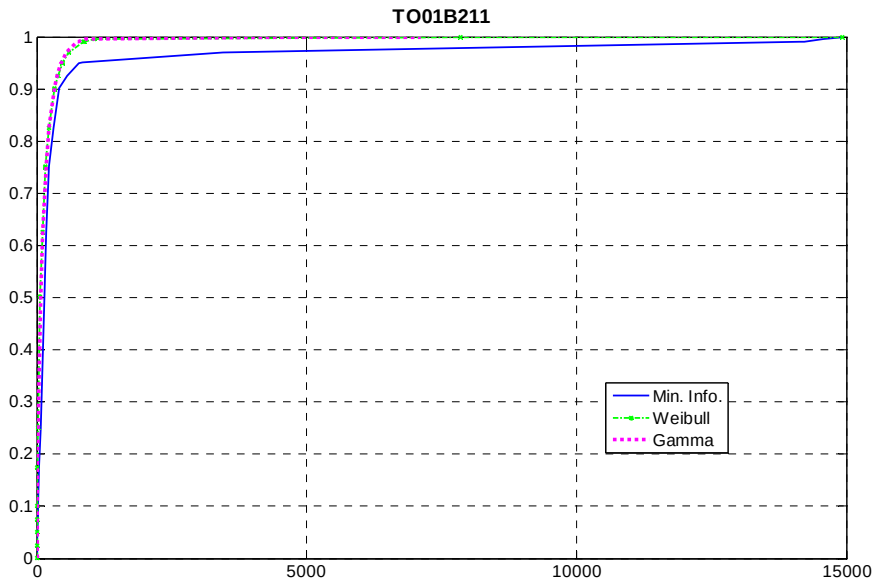


Figure 5.3 Different fits to the percentiles of the FTs base event probability distribution from DNV for the base event TO01B211

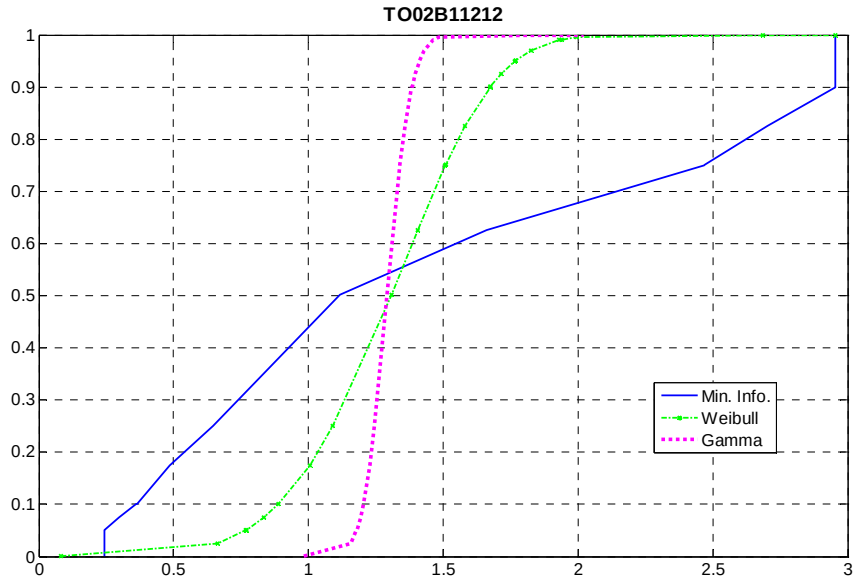


Figure 5.4 Different fits to the percentiles of the FTs base event probability distribution from DNV for the base event TO02B11212

5.2 Building up the model

Once a distribution for each of the 743 base events of interest has been found, the next step is to attach the adequate dependence information between base events. From figure 5.1 it may be observed that each base event represents an instance of one of the three human reliability models from chapter 4.

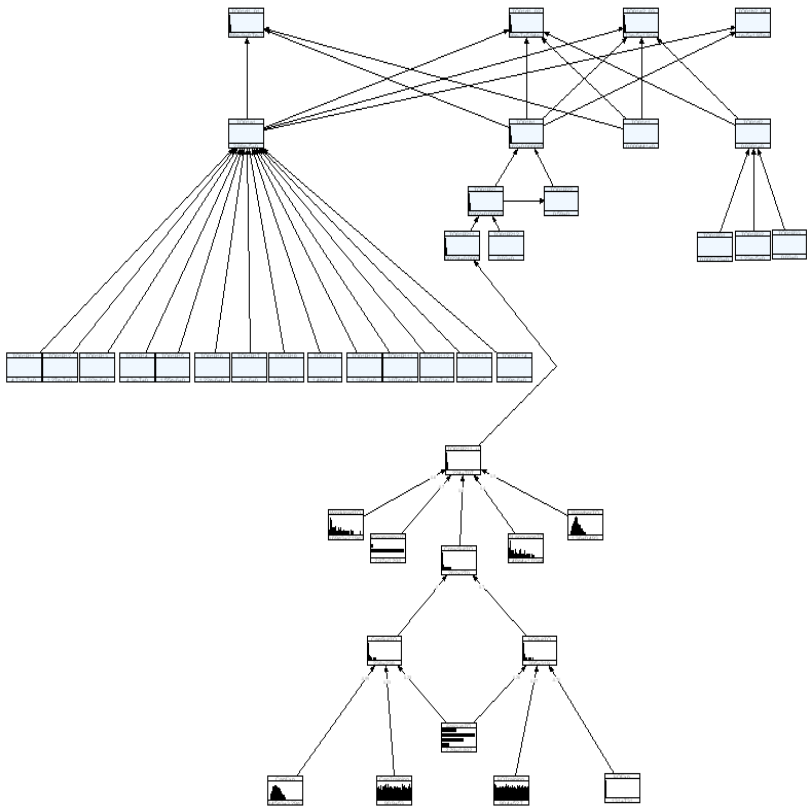


Figure 5.5 ESD1 with one instance of the FC performance model.

Figure 5.5 shows ESD1 with a single instance of the FC performance model from section 4.1. Since according to figure 5.1 there are in total three base events in ESD1 influenced by the flight crew performance, then when including the remaining two FC base events the model should look as in figure 5.6

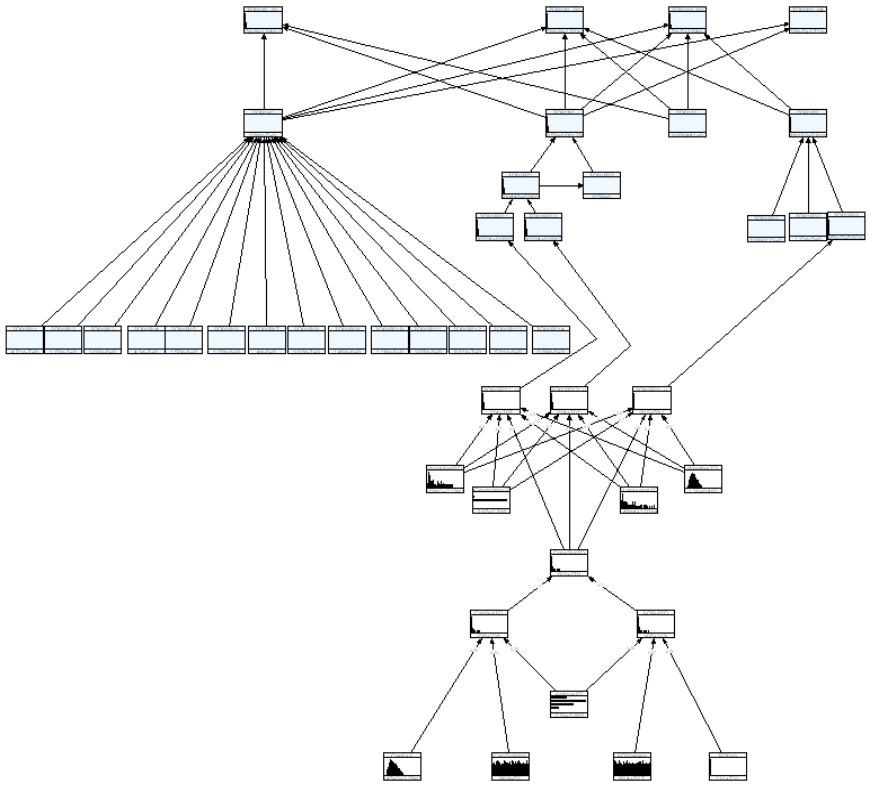


Figure 5.6 ESD1 with three instance of the FC performance model.

Also from figure 5.1 it may be observed that the FC performance model is not the only one influencing basic events in ESD1. The fifteen base events influenced by the maintenance crew performance model should also be included in the BBN representation of ESD1. The complete representation of ESD1 is shown in figure 5.6

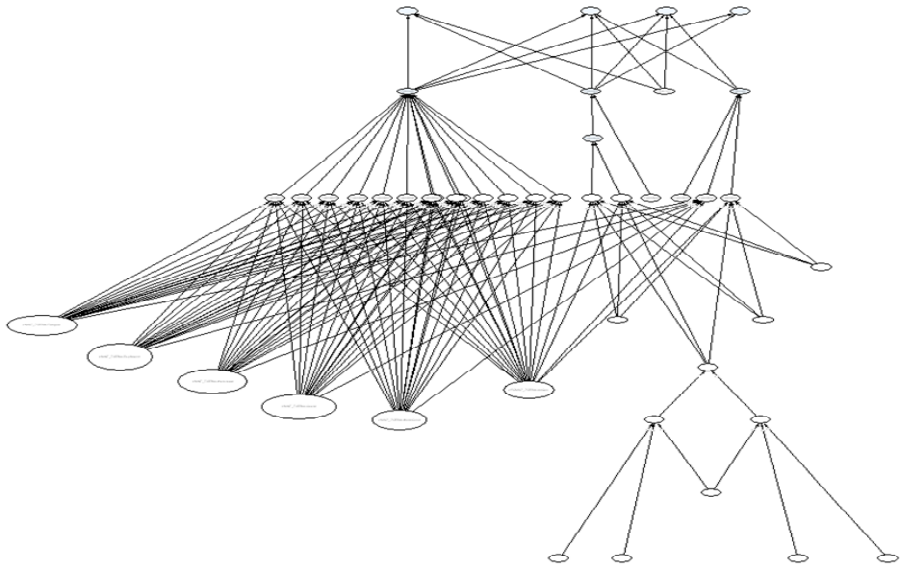


Figure 5.6 BBN representation of figure 5.1 with human reliability models expressing non-independence of base events in the FTs.

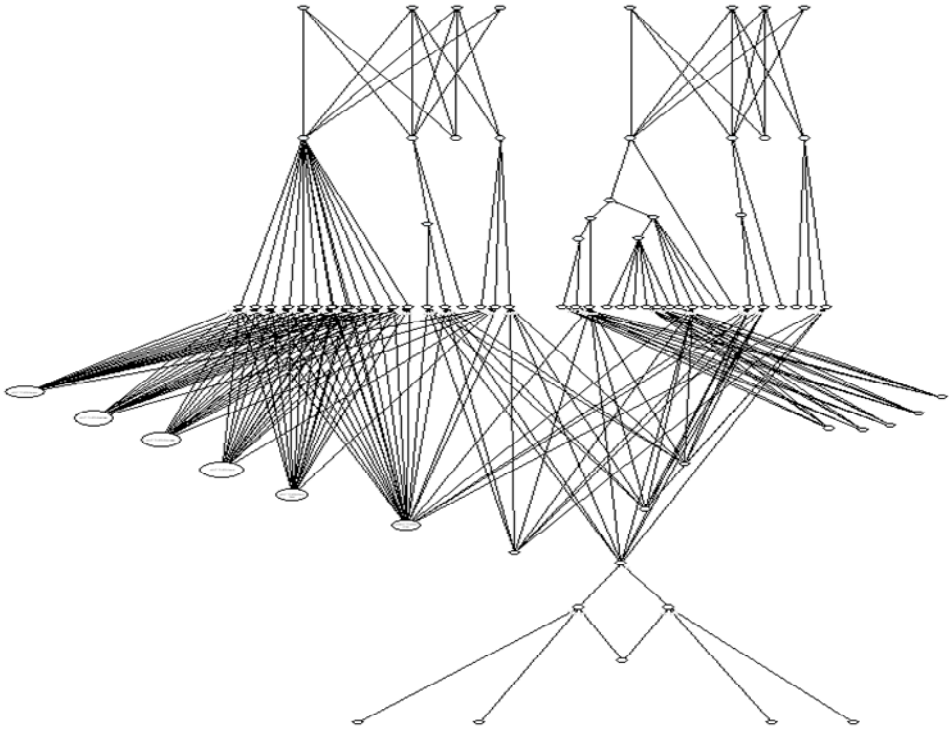


Figure 5.7 BBN representation of figure ESD1 and ESD2 with human reliability models expressing non-independence of base events in the FTs.

If the same process is repeated for ESD2 the model should look as in figure 7. This process has to be repeated for the 31 ESDs from table 5.1. The reader should observe that some nodes change through flight phases and some do not. For example, experience in the ATC anf FC models is consider not to change across flight phases. On the other hand the FC model would have one instance of weather per flight phase. A complete list of the variables as entered in UniNet may is presented in Appendix 3.

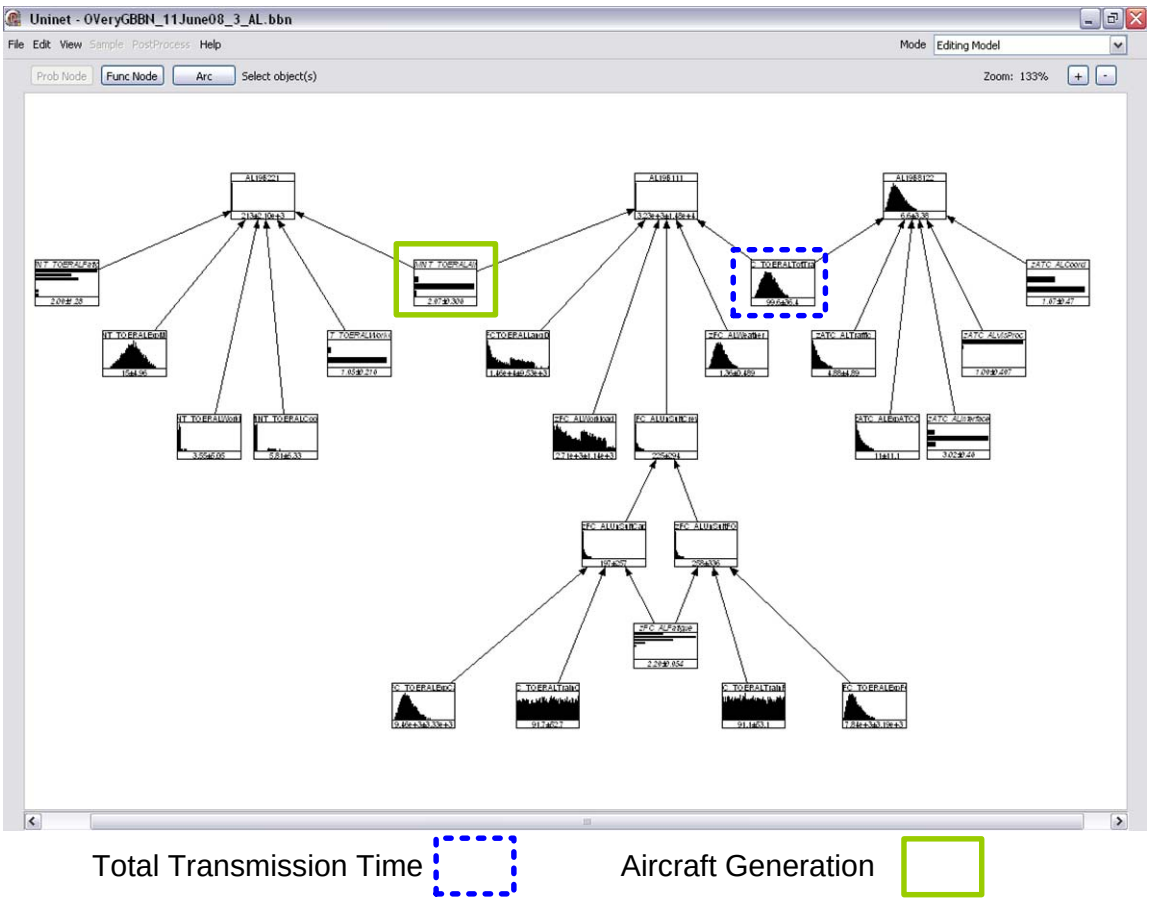


Figure 5.8 One instance of each of the ATC, FC and MNT performance models in the Approach and Landing flight phases.

In the CATS model the flight crew performance model and the ATC performance model share in common the total transmission time in the AL flight phase. The maintenance model and the flight crew model share the aircraft generation node in *all* flight phases. This situation is summarized in figure 5.8

5.3 The complete model

The complete CATS model is presented in figures 5.9 and 2.2 at the beginning of this report. The model has been integrated with the methods described in this report and the references at the end of the document. The model, at the date of publication of this document, consists of 834 probabilistic nodes, 532 functional and 4,756 arcs.

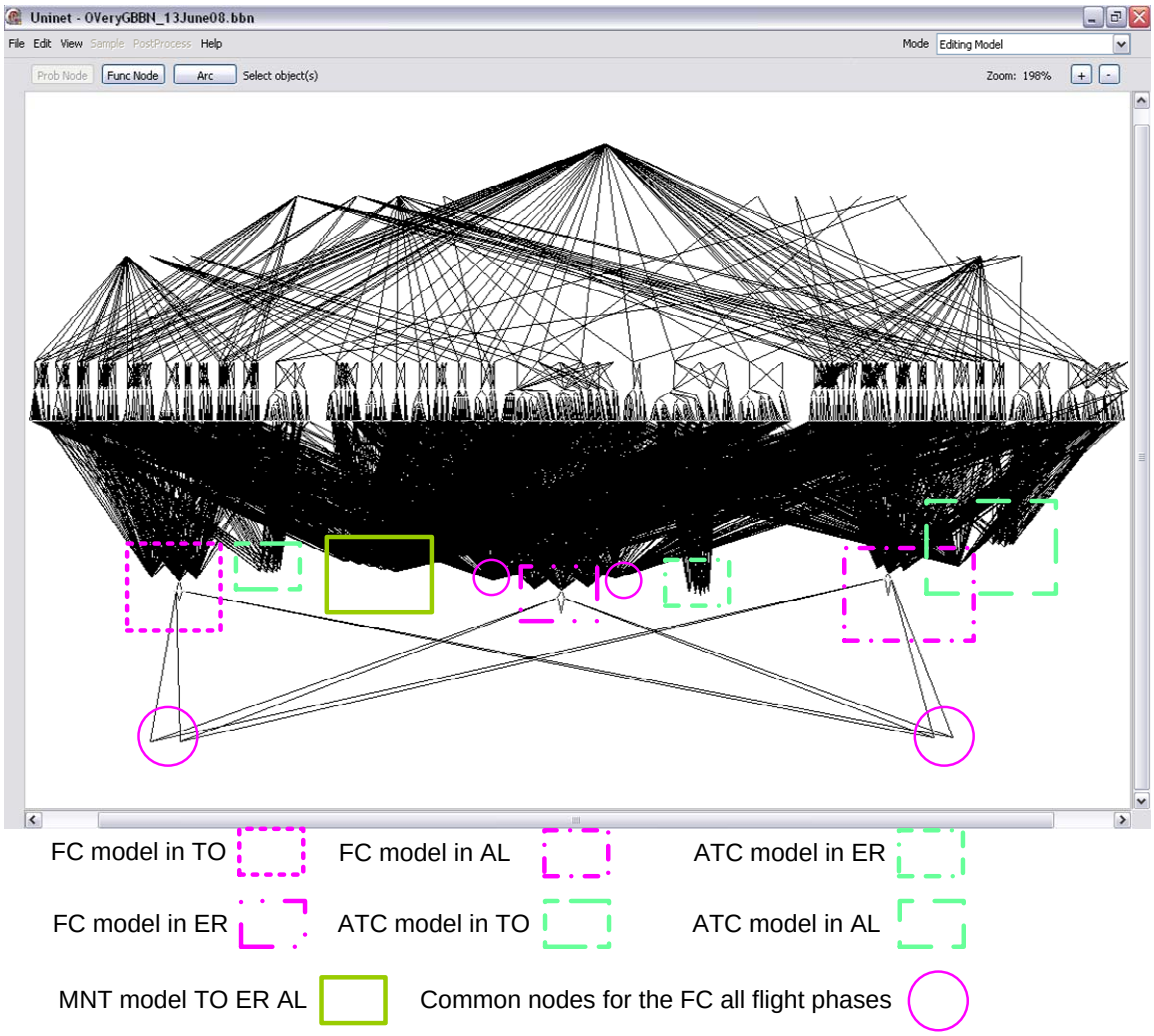


Figure 5.9 BBN representing the CATS model with human reliability models in different flight phases attached to the base events.

Appendices.

Appendix 1. Summary list of EWI products related to the CATS project

Documents:

- Report on phase 1 CATS
- Report on phase 2 CATS
- Report on phase 3 CATS
- Report on phase 4 CATS
- Report on phase 5 CATS
- Report on phase 6 CATS
- Description of the Expert elicitation Results for the Flight Crew Performance Model
- Description of the Expert elicitation Results for the ATC Performance Model
- System level Risk Analysis of new Merging and Spacing Protocols.
- Aviation Risks with continuous-discrete distribution free BBNs (Master Thesis) Katarzyna Agata Krugla
- Quantification of non-parametric continuous BBNs with ex-pert judgment (Master Thesis) I.D. Jagielska.

Software:

- UniNet (Continuous BBNs)
- UniExp (Elicitation of Rank and Conditional Rank Correlations)

Appendix 2. Elicitation procedures for rank and conditional rank correlations

Two types of methods for the elicitation of dependence measures will be briefly discussed. For a morge general overview on elicitation procedures for rank and unconditional rank correlations the reader is referred to [15] and [16]**.** Examples of the two methods will be discussed with the BBN in figure A2-1. The six marginal distributions (one required for each node) may be computed from data from separate sources or with the classical model for expert judgment [9]. Variables X_1 to X_5 are independent of each other. Four conditional rank correlations and one unconditional rank correlation are required.

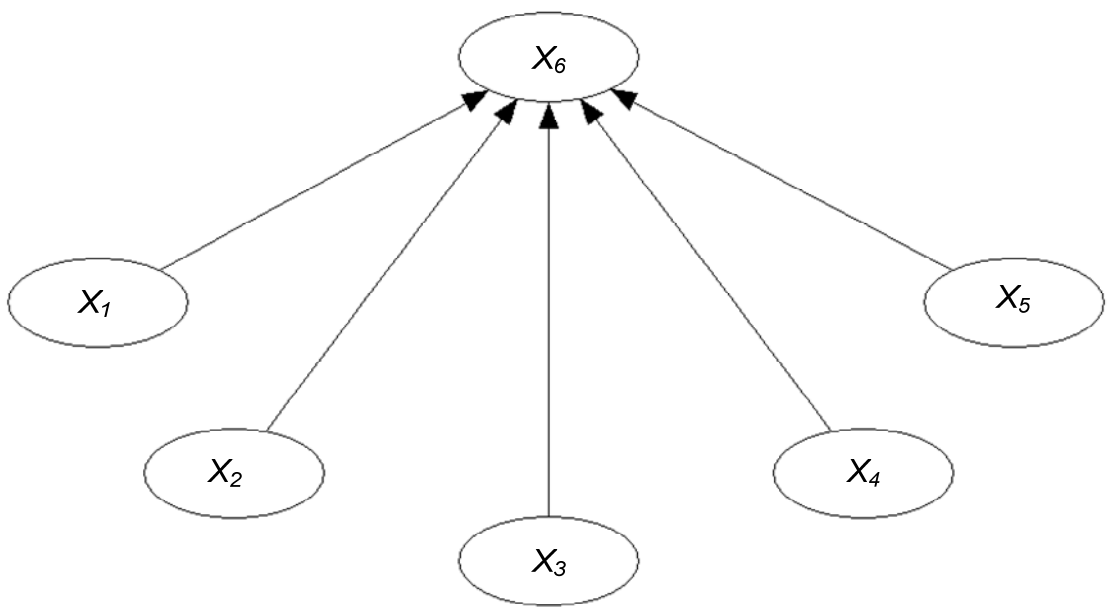


Figure A2-1 BBN on 6 Nodes

- Probabilistic Approaches: Experts are queried 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 a rank correlation. Denote $r_{6,1}^{e_i}$ the rank correlation between X_6 and X_1 for expert i = 1,...,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 $x_{j,50}^{e_i}$. Similarly the kth percentile of variable X_j is denoted as $x_{j,k}^{e_i}$. The cumulative

distribution function for variable X_j from expert e_i will be denoted as $F_{X_j}^{e_i}$. The required questions for figure A2-1 could be:

$$
\begin{aligned}
P_1^{e_i} &= P\left(X_6 > x_6^{e_i} \mid X_1 > x_1^{e_i}\right) \\
&= P\left(F_{X_6}^{e_i}(X_6) > 0.5 \mid F_{X_1}^{e_i}(X_1) > 0.5\right) \\
P_2^{e_i} &= P\left(X_6 > x_6^{e_i} \mid X_1 > x_1^{e_i}, X_2 > x_2^{e_i}\right) \\
&= P\left(F_{X_6}^{e_i}(X_6) > 0.5 \mid F_{X_1}^{e_i}(X_1) > 0.5, F_{X_2}^{e_i}(X_2) > 0.5\right) \\
P_3^{e_i} &= P\left(X_6 > x_6^{e_i} \mid X_1 > x_1^{e_i}, X_2 > x_2^{e_i}, X_3 > x_3^{e_i}\right) \\
&= P\left(F_{X_6}^{e_i}(X_6) > 0.5 \mid F_{X_1}^{e_i}(X_1) > 0.5, F_{X_2}^{e_i}(X_2) > 0.5, F_{X_3}^{e_i}(X_3) > 0.5\right) \\
P_4^{e_i} &= P\left(X_6 > x_6^{e_i} \mid X_1 > x_1^{e_i}, X_2 > x_2^{e_i}, X_3 > x_3^{e_i}, X_4 > x_4^{e_i}\right) \\
&= P\left(F_{X_6}^{e_i}(X_6) > 0.5 \mid F_{X_1}^{e_i}(X_1) > 0.5, F_{X_2}^{e_i}(X_2) > 0.5, F_{X_3}^{e_i}(X_3) > 0.5, F_{X_4}^{e_i}(X_4) > 0.5\right) \\
P_5^{e_i} &= P\left(X_6 > x_6^{e_i} \mid X_1 > x_1^{e_i}, X_2 > x_2^{e_i}, X_3 > x_3^{e_i}, X_4 > x_4^{e_i}, X_5 > x_5^{e_i}\right) \\
&= P\left(F_{X_6}^{e_i}(X_6) > 0.5 \mid F_{X_1}^{e_i}(X_1) > 0.5, F_{X_2}^{e_i}(X_2) > 0.5, F_{X_3}^{e_i}(X_3) > 0.5, F_{X_4}^{e_i}(X_4) > 0.5, F_{X_5}^{e_i}(X_5) > 0.5\right)
\end{aligned}
\tag{A2.1}
$$

Notice that the recommended choice for the percentile used in the probabilities stated in relation (A2.1) is the median; however any other percentile $x_{j,k}^{e_i}$ may be used. In particular other percentiles are necessary for discrete variables. Notice also that as stated before other probabilistic statements could be elicited in relation (A2.1) according to the analysts preference. For example shorter conditioning sets might be considered. Another option would be to elicit joint distributions instead of conditional probabilities of exceedance.

Once estimates as in relation (A2.1) are available to the analyst, the corresponding (un)conditional rank correlations may be computed for each expert (relation(A2.2)). In [15] a more extensive discussion about relation (A2.2) is provided.

$$
\begin{aligned}
P_1^{e_i} &\rightarrow r_{6,1}^{e_i} \\
P_2^{e_i} &\rightarrow r_{6,2|1}^{e_i} \\
P_3^{e_i} &\rightarrow r_{6,3|1,2}^{e_i} \\
P_4^{e_i} &\rightarrow r_{6,4|1,2,3}^{e_i} \\
P_5^{e_i} &\rightarrow r_{6,5|1,2,3,4}^{e_i}
\end{aligned}
\tag{A2.2}
$$

- Statistical Approach: Another option is to let experts directly assess a rank correlation. In particular we could let experts rank variables $X_1,\ldots,X_5$ according to the one which they regard has the largest rank correlation (in absolute value) with X_6. For each expert, the variable with the largest rank correlation (in absolute value) with X_6 will be renamed as V_1. The second largest V_2 and so on until the smallest V_5.

This ranking will in general be different for each expert. Experts could then be queried the following questions:

$$
\begin{aligned}
P_1^{e_i} &= P\left(X_6 > x_6^{e_i} \mid V_1 > v_1^{e_i}\right) \\
R_2^{e_i} &= r_{6,V_2}^{e_i} \Big/ r_{6,V_1}^{e_i} \\
R_3^{e_i} &= r_{6,V_3}^{e_i} \Big/ r_{6,V_1}^{e_i} \\
R_4^{e_i} &= r_{6,V_4}^{e_i} \Big/ r_{6,V_1}^{e_i} \\
R_5^{e_i} &= r_{6,V_5}^{e_i} \Big/ r_{6,V_1}^{e_i}
\end{aligned}
\qquad \text{(A2.3)}
$$

The largest rank correlation is still elicited through a probabilistic statement. $R_2^{e_i}$ in relation (A2.3) denotes the ratio of the second largest 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 (A2.3) is the median, however any other percentile $x_{j,k}^{e_i}$ or $v_{j,k}^{e_i}$ may be used. As stated before, other probabilistic statements could be elicited for $P_1^{e_i}$ in relation (A2.3) according to the analyst's preference.

$$
\begin{aligned}
P_1^{e_i} &\rightarrow r_{6,1}^{e_i} \\
P_2^{e_i} &\rightarrow r_{6,2|1}^{e_i} \\
P_3^{e_i} &\rightarrow r_{6,3|1,2}^{e_i} \\
P_4^{e_i} &\rightarrow r_{6,4|1,2,3}^{e_i} \\
P_5^{e_i} &\rightarrow r_{6,5|1,2,3,4}^{e_i}
\end{aligned}
\qquad \text{(A2.4)}
$$

Once estimates in relation (A2.3) are available, the desired estimates could be computed for each expert. The computation of the required (un)conditional rank correlations in relation (A2.4) follows the same arguments as before [15] [16].

One argument in favor of the elicitation of probabilistic statements is that their elicitation has proven to be feasible in previous studies [14] and [15] with 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 (A2.1)) experts tend to object the elicitation of probabilities. As mentioned previously this could be avoided by eliciting conditional probabilities with smaller number of conditioning variables.

The second method presented previously, combines the elicitation of one probabilistic statement with ratios of unconditional rank correlations. The 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 in [16] and will not be addressed here.

Appendix 3. Human reliability model nodes entered in the complete CATS model

ParameterCode	DisplayName	ParameterDescription
zATC_ALCoord	ATC Coordination A&L	Communtication/Coordination 1-does not take place in the same room 2-takes place in the same room (Approach Landing)
zATC_ALExpATCO	ATC ATCO Experience A&L	Number of years working as an ATCo at current position (Approach Landing)
zATC_ALInterface	ATC Interface A&L	1-Radio only 2-Radio and Primary Radar 3-radio, primary and secondary radar 4-radio, primary, secondary and additional tools (e.g. A-SMGCS,STCA) (Approach Landing)
zATC_ALTraffic	ATC Traffic A&L	Number of aircraft simultaneously under control by ATCo (Approach & Landing)
zATC_ALVisProc	ATC Visual Procedure at A&L	1-Normal operations 2-Operating under BZO A 3-Operating under BZO B 4-Operating under BZO C 5-Operating under BZO D (Approach Landing)
zATC_ERCoord	ATC Coordination En route	Communtication/Coordination 1-does not take place in the same room 2-takes place in the same room (Take Off)
zATC_ERExpATCO	ATC ATCO Experience En route	Number of years working as an ATCo at current position (Take Off)
zATC_ERInterface	ATC Interface En route	1-Radio only 2-Radio and Primary Radar 3-radio, primary and secondary radar 4-radio, primary, secondary and additional tools (e.g. A-SMGCS,STCA) (Take Off)
zATC_ERTraffic	ATC Traffic En route	Number of aircraft simultaneously under control by ATCo (Take Off)

zATC_ERVisProc	ATC Visual procedures En route	1-Normal operations 2-Operating under BZO A 3-Operating under BZO B 4-Operating under BZO C 5-Operating under BZO D (Take Off)
zATC_TOCoord	ATC Coordination at take-off	Communtication/Coordination 1-does not take place in the same room 2-takes place in the same room (Take Off)
zATC_TOExpATCO	ATC ATCO Experience at take-off	Number of years working as an ATCo at current position (Take Off)
zATC_TOInterface	ATC Interface at take-off	1-Radio only 2-Radio and Primary Radar 3-radio, primary and secondary radar 4-radio, primary, secondary and additional tools (e.g. A-SMGCS,STCA) (Take Off)
zATC_TOTraffic	ATC Traffic at take-off	Number of aircraft simultaneously under control by ATCo (Take Off)
zATC_TOVisProc	ATC Visual Procedure at take-off	1-Normal operations 2-Operating under BZO A 3-Operating under BZO B 4-Operating under BZO C 5-Operating under BZO D (Take Off)
zFCATC_ERALTotTransTime	Total air/ground comms time	The Total Transmission Time, i.e. the total duration (in seconds) of the air/ground communications, per aircraft
zFC_ALFatigue	Flight crew Fatigue at A&L	Stanford Sleepiness Scale 1 completely awake to 7 sleep onset soon (Approach-Landing)
zFC_ALUnSuitCap	Flight Crew Captain Unsuitability A&L	Number of Captains failing their proficiency check test per 10,000 Captains (Approach-Landing)
zFC_ALUnSuitCrew	Flight Crew Unsuitability A&L	Number of Captains or/and First Officers failing their proficiency check test per 10,000 Captains or/and First Officers (Approach-Landing)
zFC_ALUnSuitFO	Flight Crew FO Unsuitability A&L	Number of First Officers failing their proficiency check test per 10,000 First Officers (Approach-Landing)

zFC_ALWeather	Weather at cockpit A&L	Rainfall rate (mm/hr) translated into airborne weather radar in a cockpit at (Approach-Landing)
zFC_ALWorkload	Flight Crew Workload A&L	Number of times the crew members encounter an abnormal/emergency situation per 100,000 flights (Approach-Landing)
zFC_ERFatigue	Flight crew Fatigue En route	Stanford Sleepiness Scale 1 completely awake to 7 sleep onset soon (En-Route)
zFC_ERUnSuitCap	Flight Crew Captain Unsuitability En route	Number of Captains failing their proficiency check test per 10,000 Captains (En-Route)
zFC_ERUnSuitCrew	Flight Crew Unsuitability En route	Number of Captains or/and First Officers failing their proficiency check test per 10,000 Captains or/and First Officers (En-Route)
zFC_ERUnSuitFO	Flight Crew FO Unsuitability En route	Number of First Officers failing their proficiency check test per 10,000 First Officers (En-Route)
zFC_ERWeather	Weather at cockpit En route	Rainfall rate (mm/hr) translated into airborne weather radar in a cockpit at (En-Route)
zFC_ERWorkload	Flight Crew Workload En route	Number of times the crew members encounter an abnormal/emergency situation per 100,000 flights (En-Route)
zFC_TOERALExpCap	Flight Crew Captain Experience	Total number of hours flown since the pilot's license obtained by captains
zFC_TOERALExpFO	Flight crew FO Experience	Total number of hours flown since the pilots license obtained by first officers.
zFC_TOERALTrainCap	Flight crew Captain Training	Number of days passed since last recurrence training for Captains.
zFC_TOERALTrainFO	Flight crew FO Training	Number of days passed since last recurrence training for First Officers.
zFC_TOFatigue	Flight Crew Fatigue at take-off	Stanford Sleepiness Scale 1 completely awake to 7 sleep onset soon (Take Off)

zFC_TOUnSuitCap	Flight Crew Captain Unsuitability at take-off	Number of Captains failing their proficiency check test per 10,000 Captains (Take Off)
zFC_TOUnSuitCrew	Flight Crew Unsuitability at take-off	Number of Captains or/and First Officers failing their proficiency check test per 10,000 Captains or/and First Officers (Take Off)
zFC_TOUnSuitFO	Flight Crew FO Unsuitability at take-off	Number of First Officers failing their proficiency check test per 10,000 First Officers (Take Off)
zFC_TOWeather	Weather at cockpit at take-off	Rainfall rate (mm/hr) translated into airborne weather radar in a cockpit at (Take Off)
zFC_TOWorkload	Flight Crew Workload during take off	Number of times the crew members encounter an abnormal/emergency situation per 100,000 flights (Take Off)
zFCMNT_TOERALAirGen	Aircraft Generation	Aircraft generation is a scale from 1 to 4 where 4 is the most recent generation of aircrafts
zFCTOERALLangDif	Flight Crew language Difference	Number of flights in which the pilot and first officer will have a different mother tongue per 100,000 flights
zMNT_TOERALCoord	Maintenance Shift Overlap Time	Time available to transfer a job (min)
zMNT_TOERALExpMaint	Maintenance Experience	Years in current position
zMNT_TOERALFatigue	Maintenance Fatigue	Stanford sleepness scale, where 1 is feeling active, vital, alert or wide awake and 7 is no longer fighting sleep; having dream-like thoughts
zMNT_TOERALWorkCond	Maintenance Work Location	Whether the work is performed at the ramp (outside - 1) or in the hangar (inside - 2)
zMNT_TOERALWorkload	Maintenance Workload	Estimated delay in release of the aircraft (hrs)

Bibliography.

[1] A. Bobbio, L. Portinale, M. Minichino, E. Ciancamerla. *Comparing Fault Trees and Bayesian Networks for dependability Análisis*. In M. Felici, K.Kanoun, A. Pasquini (Eds.) SAFECOMP'99, LNCS 1698, pp. 310-322, 1999

[2] A. Bobbio, L. Portinale, M. Minichino, E. Ciancamerla. *Improving the analysis of dependable systems by mapping fault trees into Bayesian Networks.* Reliability Engineering & System Safety. 71 (2001) 249-60.

[3] B.J.M. Ale, L.J. Bellamy, R.M. Cooke, L.H.J. Goossens, A.R. Hale, A.L.C. Roelen, E. Smith. *Towards a causal model for air transport safety an ongoing research project*. Safety Science. 2006, vol. 44 nr. 8 pp. 657-673

[4] R.M. Cooke, O. Morales and D. Kurowicka *Vines in overview*, in Proceedings of the Third Brazilian Conference on Statistical Modelling in Insurance and Finance. 2007, pp. 2-20.

[5] B.J.M. Ale, L.J. Bellamy, Boom R.van der, J. Cooper, R.M. Cooke, L.H.J. Goossens, A.R. Hale, D. Kurowicka, O. Morales, A.L.C. Roelen and J. Spouge. *Further developments of a causal model for air transport safety (CATS); building the mathematical heart.* in ESREL 2007. 1431-1439.

[6] D. Kurowicka, R.M. Cooke. *Distribution-Free Continuous Bayesian Belief Nets.* Modern Statistical and mathematical Methods in Reliability. (2005) 309-323. Wilson A., Limnios N., Keller-McNulty S. and Armijo Y. eds.

[7] A. Hanea, D. Kurowicka, R.M. Cooke. *Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets.* Quality and reliability Engineering International. 2006 (22) 709-729.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann. 1988

[9] R.M. Cooke. *Experts in uncertainty*. Oxford University Press. 1991

[10] R.M. Cooke, L.H.J. Goossens. *TU Delft expert jugement data base.* Reliability Engineering & System Safety. 2008 (93) 657-674

[11] A. O'Hagan. *Research in elicitation.* Bayesian Statistics and its applications. 2005 375-382U. Singh and D.K. Dey eds.. Anamaya, New Delhi

[12]G.W. Clemen and et al. *Assesing Dependencies: Some Experimental Results.* Management Science 2000 Informs. 2000 (46) 8, 1100-1115

[13] G.W. Clemen and et al. *Correlations and Copulas for Decision and Risk Analysis.* Management Science. 1999 (45) 208-224

[14] B.C.P. Kraan. *Probabilistic Inversion in Uncertainty Analysis and Related Topics.* Delft University of Technology. 2002

[15] O. Morales, D. Kurowicka, A. Roelen. *Eliciting Conditional and Unconditional Rank Correlations from Conditional Probabilities.* Reliability Engineering and System Safety. 2008 (93) 699-710

[16] O. Morales, D. Kurowicka, RM Cooke, G Baren van. *Expert Elicitation Methods of Rank and Conditional Rank Correlations: An Example with Human Reliability Models in the Aviation Industry.* TU Delft Working paper. 2008

[17] A.L.C. Roelen, G.B. van Baren, J.W. Smeltink, P.H. Lin and O. Morales, *A generic flight crew performance model for application in a causal model of air transport*, Nationaal Lucht- en Ruimtevaartlaboratorium (National Aerospace Laboratory NLR). 2007. NLR-CR-2007-562

[18] A.L.C. Roelen, B.A. van Doorn, J.W. Smeltink, M.J. Verbeek, R. Wever. *Quantification of event sequence diagrams for a causal risk model of commercial air transport*. Nationaal Lucht- en Ruimtevaartlaboratorium National Aerospace Laboratory NLR. 2007. NLR-CR-2006-520

[19] W.E. Vesely, F.F. Goldberg, N.H. Roberts and D.F. Haasl. *Fault Tree Handbook.* U.S. Nuclear regulatory Commission. 1981 NUREG-0492

[20] Kurowicka D., and Cooke R.M., A Parametrization of Positive Definite Matrices in Terms of Partial Correlation Vines", in Linear Algebra and its Application 372, 225-251, 2003.

[21] Kurowicka D., and Cooke R.M., Completion Problem with Partial Correlation Vines, Linear Algebra and Its Applications,, vol. 418, nr 1, pp 188-200, 2006_B.

[22] Hanea, A.M., Kurowicka, D. Cooke, R.M. and Ababei, D.A.A. “Ordinal data mining with non-parametric continuous Bayesian belief nets”, Dept. of Mathematics, T.U. Delft, 2008.

[23] Bedford, T.J. , and Cooke, R.M. , "Vines - a new graphical model for dependent Random Variables"- in Annals of Statistics, vol. 30, No.4, August 2002.

[24] Bedford, T.J. and Cooke, R.M. "Probability density decomposition for conditionally dependent random variables modelled by Vines, " Annals of Mathematics and Artificial Intelligence, Volume 32 pp 245-268, 2001.

[25] Cooke, R.M. , and Goossens, L.H.J.. "Procedures Guide for Structured Expert Judgement in Accidents Consequence Modeling", Radiation Protection Dosimetry Vol. 90 No. 3.2000, pp 303-311

[26] Hanea, A.M. “Algorithms for non-parametric Bayesian Belief Nets”, PhD thesis, Dept. of Mathematics, Delft University of Technology, 2008.

[27] Joe, H. , Relative entropy measures of multivariate dependence, Journal of the American Statistical

[28] Kurowicka, D. , and Cooke, R.M. Uncertainty Analysis and High Dimensional Dependence Modeling Wiley, New York, 2006-A.

[29] Kurowicka, D. , Cooke, R.M. , Sampling algorithms for generating joint uniform distributions using the vine-copula method, , Comp. Stat. & Data Analysis, Volume 51 , Issue 6 March 2007.

[30] Lewandowski, D. Cooke, R.M. and Duintjer Tebbens, R.J. "Sample Based Estimation of Correlation Ratio with Polynomial Approximation" Transactions on Modeling and Computer Simulation, vol. 18 no.1, article 3, 2007.

[31] A.L.C. Roelen, G.B. van Baren, P.H. Lin, O. Morales, D. Kurowicka, R.M. Cooke A generic air traffic controller performance model for application in a causal model of air transport Nationaal Lucht- en Ruimtevaartlaboratorium National Aerospace Laboratory NLR. 2007 NLR-CR-2007-593

[32] 080523 DNV Collected Fault Trees v6_1.xls (Excel file).

[33] FAA Aerospace Forecasts Fiscal Years 2007-2020, US Department of Transportation, Federal Aviation Administration, Aviation Policy and Plans.

[34] Civil Aviation Safety Data, Inspectie Verkeer en WaterStaat. Paostbus 575, 2130 AN Hoofddorp, www.ivw.nl

Acronyms.

ATC	Air traffic control
ATCo	Air traffic controller
BBN	Bayesian belief net
CATS	Causal Model for Air Transport Safety
CFIT	Control flight into terrain
DNV	Det Norske Veritas
ESD	Event Sequence Diagram
FT	Fault Tree
NLR	National Aerospace Laboratory
TUD	Technische Universiteit Delft / Delft University of Technology
WQ	White Queen
DIAM	Delft Institute of Applied Mathematics
EEMCS	Electrical Engineering Computer Sciences and Mathematics