Report on Phase 1 Causal Modeling for Schipol Airport

Roger Cooke, Oswaldo Morales, Dorota Kurowicka
EWI, TU Delft,
Jan 16, 2006

Introduction

This reports on the activities of EWI during the first 6 months of the CATS project. The project kicked of in July 2005, although preparatory work was initiated in the spring.

This report describes

- Development of mathematical models for distribution free continuous/discrete BBNs (DFCDBBNs)
- Development of elicitation protocols
- Development of software implementation for DFCDBBNs
- Prototype application for Controlled Flight into Terrain (CFIT)

Development of mathematical models

A Bayesian Belief Net (BBN) is a directed acyclic graph, together with an associated set of conditional probability distributions. The nodes of the graph represent random variables, which can be discrete or continuous, and the arcs represent causal relationships between variables. BBNs enable us to model high dimensional uncertainty distributions. The visual representation can be very useful in clarifying previously opaque assumptions about the dependencies between different variables.

The mathematical model development for CATS phase 1 is described in Appendix 1 " Distribution-Free Bayesian Belief Nets for Causal Models". The different types of BBNS, and motivation for DFCDBBNs are summarized as follows

Discrete BBN's
Discrete BBN's specify the marginal distributions for source nodes, and specify conditional probability tables for child nodes. They suffer three serious disadvantages:

1. Applications involving high complexity in data-sparse environments are severely limited by the excessive assessment burden which leads to rapid, informal and indefensible quantification.
2. The marginal distributions can often be retrieved from data, and are often continuous. This is often the most important information driving the model; dependence information is often less important. Discretizing this information into a small number of values sacrifices important data input.

3. Discrete BBNs take marginal distributions only for source nodes, marginals for other nodes are computed from the conditional probability tables. When these marginals are available from data, however, this imposes difficult constraints on the conditional probabilities. Thus in quantification with expert judgment, it would be impractical to configure the elicitation such that the experts would comply with the marginals.

Discrete Normal BBNs

Until recently, continuous BBNs were restricted to the joint normal distribution, where continuous nodes can have discrete parents but not discrete children. For each normal variable, the unconditional mean and (by assumption constant) conditional variance must be assessed. For each arc a conditional regression coefficient must be assessed. This is the answer to a question of the following type:

> *Suppose that ONE parent variable were moved up by ONE Normal Unit, by how many Normal Units would you expect the child to move?*

Discrete normal BBNs work well if indeed the normality assumptions hold. If not, then

1. The individual variables must be transformed to normal (requiring of course the marginal distributions).
2. The conditional variance *in normal units* must be constant, 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.
3. If a parent node is added, after quantification, then the previously assessed partial regression coefficients must be re-assessed.

To illustrate these issues, the densities for # of Missed Approach executions per 100,000 flights and of Visibility, as obtained from data for the prototype application are shown below. The horizontal units are the natural units of these two variables, and vertical units are the normal units. Normal units are indicted on the horizontal axis as the intervals between the arrowheads.

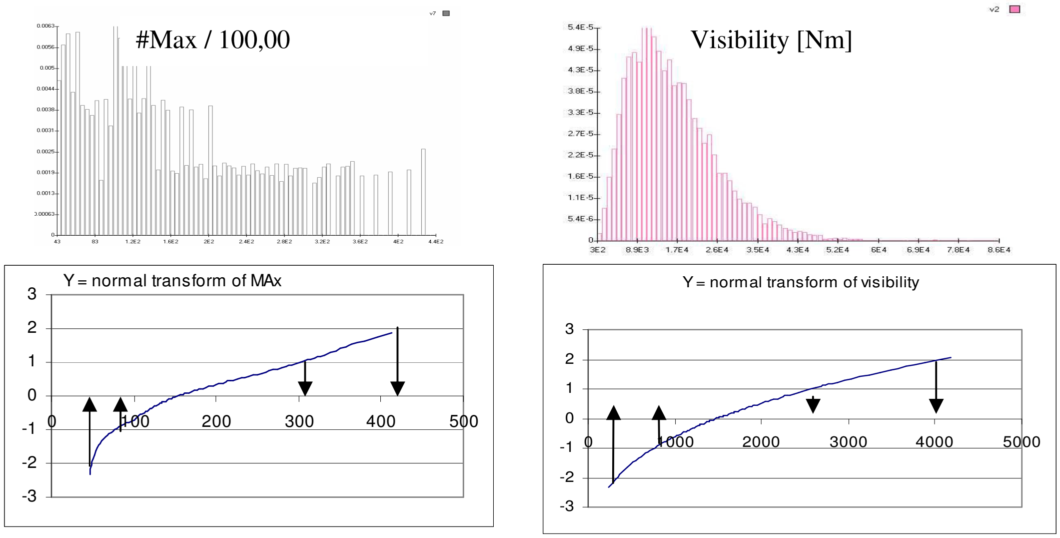

Figure 1. Normal transformations of Marginal Distributions.

Distrsibution free continuous/discrete BBNs
(Kurowicka 2004) introduce an approach to continuous BBNs using vines (Bedford2002) together with copulas that have *the zero independence property* In the procedure proposed here, nodes are associated with arbitrary continuous invertible distributions and arcs with conditional rank correlations, which are realized by the (conditional) copula, indexed by (conditional rank) correlation. The sampling procedure works with arbitrary conditional copulae satisfying the zero independence property. Thus it can happen that variables X, and Y are positively correlated when variable Z takes low values, but are negatively correlated when Z is high.

Use of non-constant conditional copulae would significantly complicate the Monte Carlo sampling and the quantification. The current platform supports only constant conditional copulae, as this is judged prudent for a first implementation. Given that the conditional copulae are constant, there are great advantages to using the joint normal copulae, which requires constant conditional copulae. Unlike the normal BBN, however, nodes and influences can be added or deleted without re-assessing previously assessed quantities.

The assessment burden for a DFCDBBN is thus:

1. One dimensional distributions for each node, hopefully obtained from data, otherwise from expert judgment
2. For each arc, a (conditional) rank correlation as indicated by the protocol in Appendix 1.

Development of elicitation protocols

Conditional rank correlations are difficult to assess directly. Instead these can be assessed by asking for (conditional) exceedence probabilities. To illustrate, we consider part of the BBN for CFIT shown below.

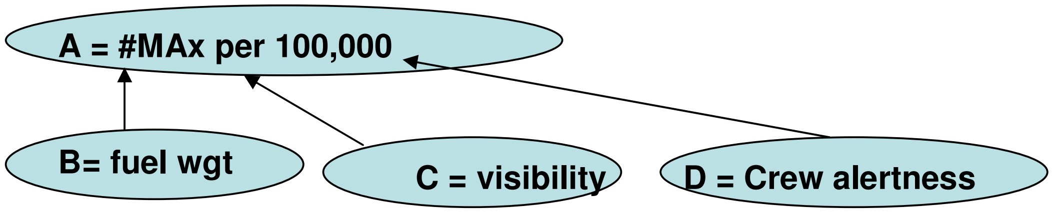

Figure 2. Hypothetic BBN for #Max per 100,000 flights.

The elicitation questions, in stylized format, are shown below.

Marginal for A:
Consider 100,000 flights under current conditions at Schipol Airport. On how many of these flights will a Missed Approach be executed?
5% __________; 50% (median)___________; 95%____________
Marginal for B: *idem*
Marginal for C: *idem*
Dependence:
On 50,000 flights, B is above its median. Consider 100,000 flights with B above its median, what is the probability that A is above its median?"____

Suppose B and C are above their medians on 100,000 flights, what is the probability that A is above its median? _____

Suppose B C and D are above their medians on 100,000 flights, what is the probability that A is above its median?_____

These conditional probabilities are not independent, and the answer to one question constrains the possible answers to the others. A mock-up software elicitation support tool has been developed to help visualize these constraints, of which a screen shot is shown below:

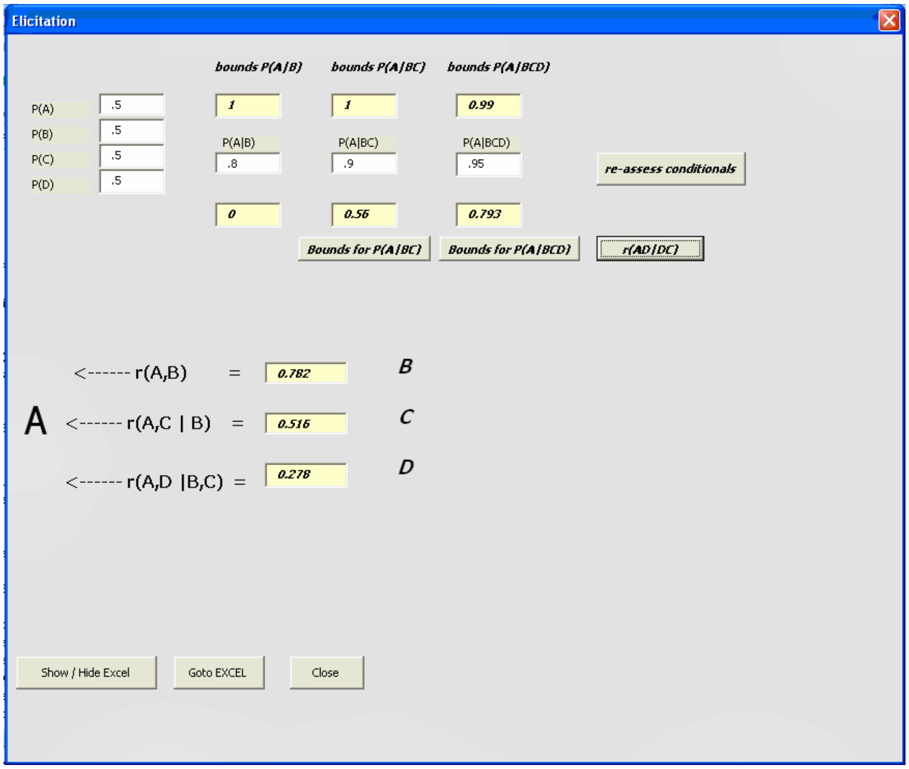

Figure 3. Mock-up software tool for support on the elicitation of dependencies.

The choice of P(A | B) is unconstrained. When the user assess this as 0.80, the choice of P(A | BC) is constrained to lie between 1 and 0.56, etc. In this way the user gets an intuitive grasp of the meaning of "influence" between the nodes. Note that the conditional rank correlations which realize the conditional probabilities are shown. These are the numbers entered into the DFCDBBN. A prototype elicitation, using the methods previously described was performed at NLR on December 20th with a single expert who is a pilot for KLM and a researcher for NLR. The results from the elicitation are summarized in Table 1. These will be used for further demonstration.

Value of Interest	Index	P	r
$r_{7,6}$	7: Missed approach, 6: Separation	0.15	-0.88
$r_{7,5\mid 6}$	7: Missed approach, 5: Cross Wind | 6: Separation	0.18	0.20
$r_{7,4\mid 6,5}$	7: Missed approach, 4: Speed Dev. | 6: Separation, 5: Cross Wind	0.20	0.12
$r_{7,3\mid 6,5,4}$	7: Missed approach, 3: Crew Alertness | 6: Separation, 5: Cross Wind, 4: Speed Dev.	0.24	0.23
$r_{7,2\mid 6,5,4,3}$	7: Missed approach, 2: Visibility | 6: Separation, 5: Cross Wind, 4: Speed Dev., 3: Crew Alertness	0.22	-0.11
$r_{7,1\mid 6,5,4,3,2}$	7: Missed approach, 1: Fuel Weight | 6: Separation, 5: Cross Wind, 4: Speed Dev., 3: Crew Alertness 2: Visibility	0.24	0.11

Table 1. Results from elicitation of dependences for the CFIT model.

Development of software implementation

A software tool has been developed as a satellite program for the Monte Carlo program UNICORN. So far the software has implemented already the algorithms described in the appendix of this report and a first version of the interface is readily available for the users.

The steps to work with the software are as follows:

1. Specify in UNICORN the variables that will enter the BBN later. Names and description for each variable are entered here. Continuous parametric distributions, discrete distributions and distributions specified in a ".dis" file are supported. In this part of the software the nodes of the BBN representing marginal distributions are generated (Figure 4).
2. From UNICORN go to the Run menu and click on Bayesian Belief Net. This command takes the user immediately to the BBN software and generates the marginal distributions in the required format for the satellite software for BBN. (Figure 4)

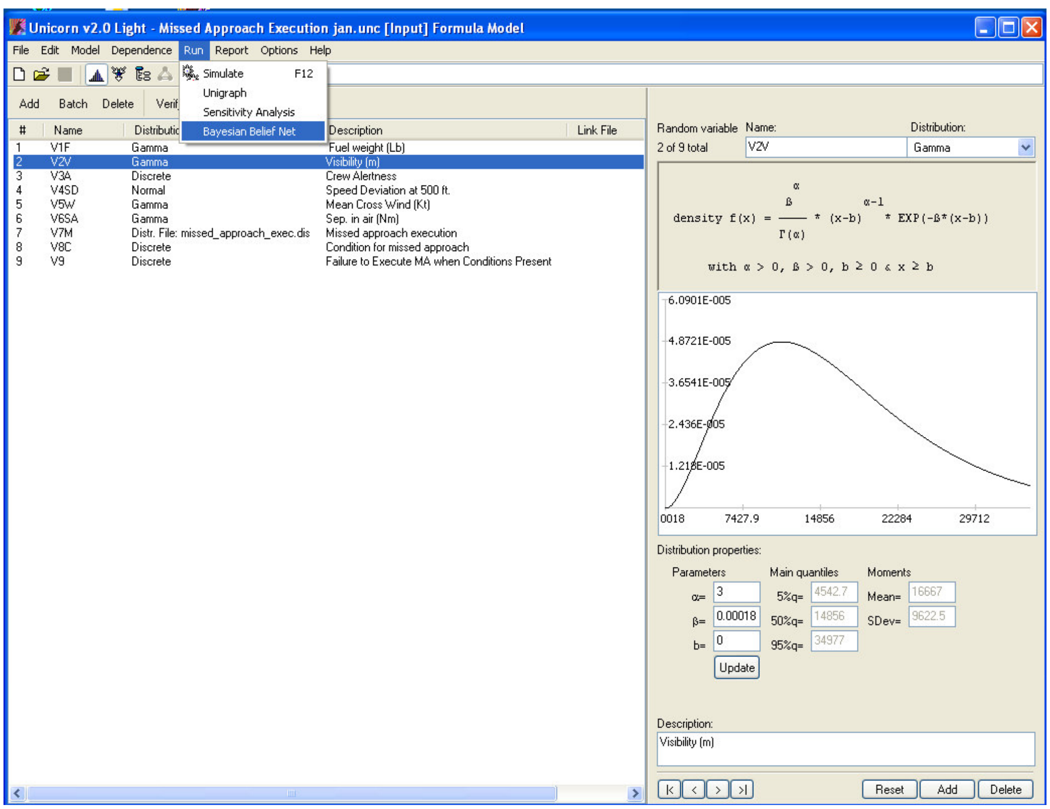

Figure 4. Specifying Nodes of the BBN in UNICORN.

Prototype application to CFIT

3. The first version of the interface is presented in Figure 5. In the *Random Variables* panel, the variables specified in UNICORN are placed with the same names and description entered in the previous step.
4. Drag the variables from the *Random Variable* panel to the blank window in the left. This operation will create the nodes of the BBN, the user may check that the marginal distributions correspond to those specified in UNICORN by right clicking on each node and selecting the option *Hystogram* from the menu.
5. Click on the "Add Arc" button to finish the construction of the BBN. An arc is added from any child to its parents by firs clicking in the child node and then in the parent node. The dependence info, i.e. the values for the conditional rank correlations and the order in which the rank correlations will enter the model are specified by right clicking on any parent node and selecting *Dependence Info* from the menu.
6. A detailed explanation is not given here, as the interface is still under development. The following screen shot shows the CFIT BBN and shows a histogram for the node Visibility.

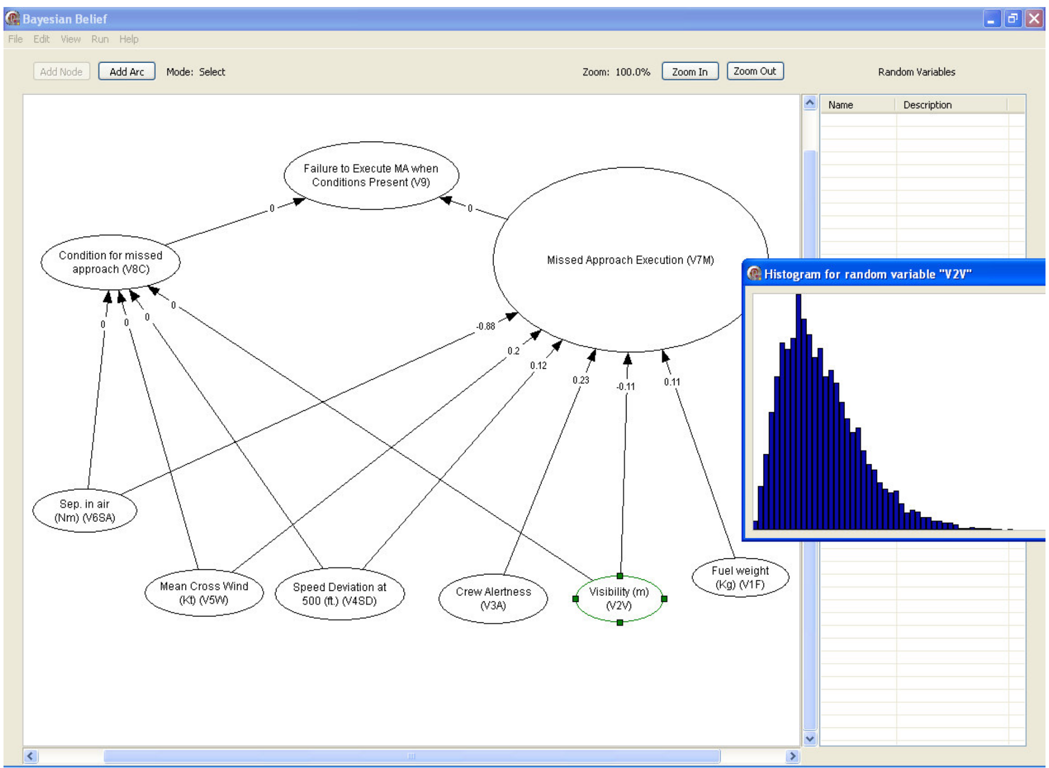

Figure 5. Satellite Unicorn BBN Software Implementation.

7. When the user runs the model from the main menu, a sample file is created. For the moment two formats for the sample file are available. The first is a ".cas" file, that is the required format of a case file for Netica and the second one is a ".sam" file that is the required format for UNICORN.

8. After the BBN is sampled, the ".cas" file can be exported to NETICA for further analysis and decision support. A discretized version of the model should be prepared beforehand in Netica, and Netica will use the sample file created in the previous peace of software to populate the probability tables required by the discrete version of the model. This joint distribution preserves the dependence information specified by the arcs of the BBN and is consistent with the marginal distributions specified in UNICORN.

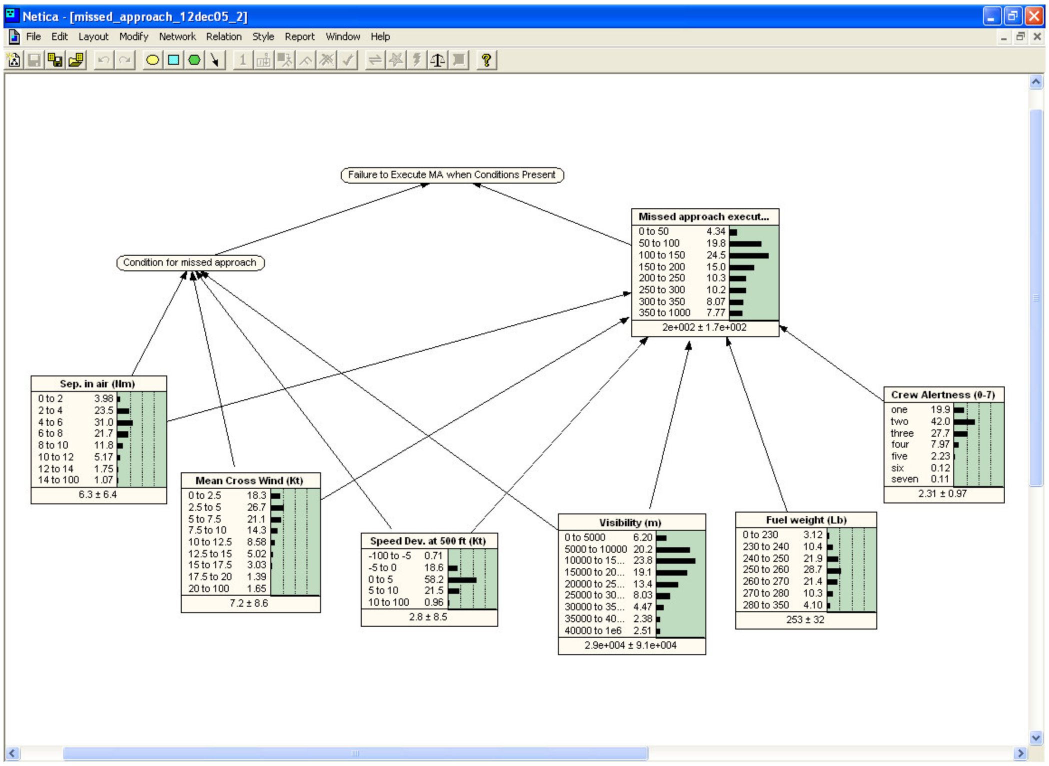

Figure 6. Discrete Version of the MAx Model in Netica.

9. Fast updating algorithms are implemented in Netica that allow the user to observe how a marginal distribution changes whenever information about some of the variables in the joint distribution becomes available. The reader may compare the node for Missed approach execution in Figures 6 and 7. The updating takes only few seconds to be performed.

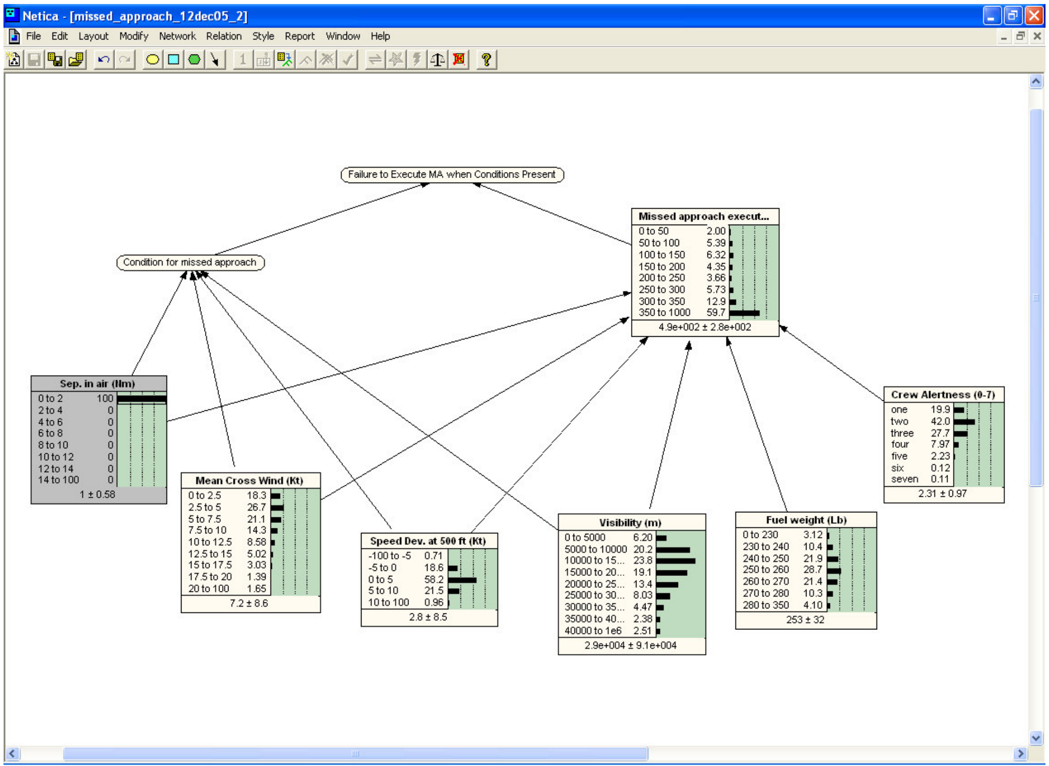

Figure 7. Discrete Version of the MAx Model in Netica conditioning on low separation in air.

As stated previously the model itself and the software application are under development. Improvements will be discussed on further versions of this report.