

Special Issue

Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets

A. M. Hanea*,† D. Kurowicka and R. M. Cooke

Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Bayesian belief nets (BBNs) have become a popular tool for specifying high-dimensional probabilistic models. Commercial tools with an advanced graphical user interface that support BBNs construction and inference are available. Thus, building and working with BBNs is very efficient as long as one is not forced to quantify complex BBNs. A high assessment burden of discrete BBNs is often caused by the discretization of continuous variables. Until recently, continuous BBNs were restricted to the joint normal distribution. We present the ‘copula–vine’ approach to continuous BBNs. This approach is quite general and allows traceable and defensible quantification methods, but it comes at a price: these BBNs must be evaluated by Monte Carlo simulation. Updating such a BBN requires re-sampling the whole structure. The advantages of fast updating algorithms for discrete BBNs are decisive. A hybrid method advanced here samples the continuous BBN once, and then discretizes this so as to enable fast updating. This combines the reduced assessment burden and modelling flexibility of the continuous BBNs with the fast updating algorithms of discrete BBNs. Sampling large complex structures only once can still involve time consuming numerical calculations. Therefore a new sampling protocol based on normal vines is developed. Normal vines are used to realize the dependence structure specified via (conditional) rank correlations on the continuous BBN. We will emphasize the advantages of this method by means of examples. Copyright © 2006 John Wiley & Sons, Ltd.

Received 2 February 2006; Revised 18 April 2006; Accepted 23 May 2006

KEY WORDS: Bayesian belief nets; dependence modelling; vines; multivariate probability distribution

1. INTRODUCTION: DISCRETE, NORMAL AND NON-PARAMETRIC BBNs

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

*Correspondence to: A. M. Hanea, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

†E-mail: a.hanea@ewi.tudelft.nl

high-dimensional uncertainty distributions. The visual representation can be very useful in clarifying previously opaque assumptions about the dependencies between different variables.

1.1. Discrete BBNs

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

- Applications involving high complexity in data-sparse environments are severely limited by the excessive assessment burden which leads to rapid, informal and indefensible quantification. Thus, if a child node has seven parents, and each node is discretized to take three possible values, then the conditional probability table for the child node contains 6561 entries, i.e. 6561 conditional probabilities must be acquired and maintained. This assessment burden can only be reduced by a drastic discretization of the nodes, or simplification of the model. In many cases continuous nodes would be more appropriate.
- The marginal distributions can often be retrieved from data, but not the full interactions between children and parent nodes. These marginal distributions often represent the most important information driving the model; dependence information is often less important. Thus the construction of conditional probability tables should not molest any available data input. Rough discretization of course does exactly that.
- 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, 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.

Some of the drawbacks listed above are mentioned in Cowell *et al.*¹ and a typical example of how things can go wrong in modelling complex problems with discrete BBNs is given by Edwards².

1.2. Normal and discrete-normal BBNs

Continuous BBNs developed for joint normal variables interpret ‘influence’ of the parents on a child as partial regression coefficients when the child is regressed on the parents^{3,4}. 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?*’

One can also construct a discrete-continuous model¹ in which continuous nodes can have discrete parents but not discrete children[‡] and the conditional distribution of the continuous variables given the discrete variables is multivariate normal.

The price of the normal and discrete-normal BBNs is the restriction to the joint normal distribution, and, in the absence of data, to experts who can estimate partial regression coefficients and conditional variances. If the normality assumption does not hold, then:

- the individual variables must be transformed to normals (requiring of course the marginal distributions);
- 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);
- if a parent node is added or removed, after quantification, then the previously assessed partial regression coefficients must be re-assessed (this reflects the fact that partial regression coefficients depend on the set of regressors).

Hence, circumventing the restriction to joint normality is primarily of theoretical interest.

[‡]Theoretically there is no need for such a restriction. However in applications, if this restriction is violated, some conditional marginals become mixtures of normals and this extension is technically demanding¹.

1.3. Non-parametric BBNs

In Kurowicka and Cooke⁵ the authors introduced an approach to continuous BBNs using vines^{6,7} together with copulae that represent (conditional) independence as zero (conditional) rank correlation. In the procedure proposed here, nodes are associated with arbitrary continuous invertible distributions and arcs with (conditional) rank correlations, which are realized by the chosen copula. No joint distribution is assumed, which makes the BBN non-parametric. In order to quantify BBNs using this approach, one needs to specify all one-dimensional marginal distributions and a number of (conditional) rank correlations equal to the number of arcs in the BBN. A way of stipulating a joint distribution is by sampling it, hence we use a sampling algorithm, fully described in Section 3, to specify and analyse the BBN structure.

The sampling procedure works with arbitrary conditional copulae. 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. This behaviour indicates that it would be appropriate to use non-constant conditional copulae (hence non-constant conditional correlations), but the use of such copulae would significantly complicate the Monte Carlo sampling and the assessment. We will therefore restrict our study to *constant* conditional rank correlations. This is a prudent choice in a first implementation.

Conditional rank correlations are not elicited directly or estimated from data directly. Rather, given a copula, these can be obtained from conditional exceedence probabilities. Thus suppose node A has parents B and C . According to the protocol described in Section 3, we need the rank correlation r_{AB} and the conditional rank correlation $r_{AC|B}$. We extract these from answers to the following two questions.

- ‘Suppose that B was observed to be above its median, what is the probability that A is also above its median?’
- ‘Suppose that B and C were both observed to be above their medians, what is the probability that A is also above its median?’

The conditional rank correlations, obtained in the way described above, can be realized using any copula that represents (conditional) independence as zero (conditional) rank correlation. Given that the conditional copulae are constant, there are great advantages to using the joint normal copulae (see Section 5).

Section 2 of this paper presents a number of definitions and preliminary results necessary in the development of the methods introduced here. Section 3 introduces the details of non-parametric BBNs using the copula–vine modelling approach. The protocol for assigning constant (conditional) rank correlations to the edges of a BBN is presented. The main result from this section (Theorem 3.1) shows that these assignments, the choice of the copula, and the marginals uniquely determine the joint distribution. Moreover, the (conditional) rank correlations assigned to the edges of a BBN are *algebraically independent*. The dependence structure is meaningful for any such quantification, and need not be revised if the univariate distributions are changed.

In case of complex structures, some conditional distributions have to be calculated. These calculations consist of evaluating multiple integrals. For each sample (from the sample file), one needs to calculate the numerical value of these multiple integrals and this can be very time consuming. In the case of a large number of variables, one would have to be prepared to run the model for a few days. Moreover, updating such a BBN requires re-sampling the whole structure every time new evidence becomes available. In these cases the advantages of fast updating algorithms for discrete BBNs are decisive. This motivates the hybrid approach advanced in Section 4 of this paper, which consists of combining the reduced assessment burden and modelling flexibility of the continuous BBNs with the fast updating algorithms of discrete BBNs. This is done using vine sampling together with existing discrete BBNs software. Even though this procedure has some attractive features, remarkably in the fact that the structure needs to be sampled just once, it also has disadvantages. We will present a small example in which sampling just this one time becomes very complicated and time consuming under certain conditions.

To overcome this problem, a new method for sampling the BBN is proposed in Section 5. The influences will still be represented as (conditional) rank correlations and they will be realized by the joint normal copula. We use the joint normal distribution to realize the dependence structure specified via (conditional) rank correlations on the continuous BBN. We will transform the variables to standard normals and the (conditional) rank correlations

to (conditional) product moment correlations via Pearson's formula⁸. Using the properties of the multivariate normal distribution, we now have a partial correlation vine specification that uniquely determines the correlation matrix, as shown by Bedford and Cooke⁷. Furthermore, we can induce this product moment correlation structure using well-known methods, and transform back to the original variables.

In doing so we circumvent the expensive numerical evaluations of multiple integrals that are necessary in the copula–vine method. The normal copula–vine approach requires only calculating new partial correlations. These calculations are not very time consuming, so the computational time is reduced considerably.

Section 6 of this paper describes a very important feature of the normal copula–vine method, namely that conditioning can be performed analytically. Since all the calculations are performed on a joint normal vine, any conditional distribution will also be a normal with known mean and variance. Finding the conditional distribution of the corresponding original variable will just be a matter of transforming it back using the inverse distribution function of this variable and the standard normal distribution function. Examples and comparisons will be presented throughout the paper for a better understanding of the methods.

2. DEFINITIONS AND CONCEPTS

In this section we introduce notations and terminology needed to describe the approach proposed for non-parametric BBNs using copula–vine modelling. All definitions presented here can be found in Kurowicka and Cooke⁹. We will start by defining *copulae* and *vines*.

The *copula* of two continuous random variables X and Y is the joint distribution of $F_X(X)$ and $F_Y(Y)$, where F_X , F_Y are the cumulative distribution functions of X , Y respectively. The copula of (X, Y) is a distribution on $[0, 1]^2$ with uniform marginal distributions. A copula is said to have the *zero independence property* if zero correlation entails the independent copula.

A graphical model called vines was introduced by Bedford and Cooke⁷. A vine on n variables is a nested set of trees. The edges of the j th tree are the nodes of the $(j + 1)$ th tree. \mathcal{V} is called a *regular vine* on n elements if:

1. $\mathcal{V} = (T_1, \dots, T_n)$;
2. T_1 is a tree with nodes $N_1 = 1, \dots, n$ and edges E_1 , and for $i = 2, \dots, n - 1$, T_i is a tree with nodes $N_i = E_{i-1}$;
3. For $i = 2, \dots, n - 1$, $a, b \in E_i$, $\#a \Delta b = 2$, where Δ denotes the symmetric difference.

We will distinguish two particular regular vines.

- A regular vine is called a D-vine if each node in T_1 has the degree at most 2 (see Figure 1(a));
- A regular vine is called a C-vine if each tree T_i has a unique node of degree $n - i$. The node with maximal degree in T_1 is called the root (see Figure 1(b)).

In the next sections we will restrict our attention only to D-vines.

Each edge of a regular vine may be associated with a constant (conditional) rank correlation which can be arbitrarily chosen in the interval $[-1, 1]$ (see Figure 1). Using a copula to realize these (conditional) rank correlations, a joint distribution satisfying the copula–vine specification can be constructed and it will always be consistent. For rigorous definitions and proofs we refer to Kurowicka and Cooke⁹. Each vine branch may also be associated with partial correlations. First, we will recall the definition of the partial correlation¹⁰. If X_1, \dots, X_n are random variables, the *partial correlation* of X_1 and X_2 given X_3, \dots, X_n is

$$\rho_{12;3,\dots,n} = \frac{\rho_{12;4,\dots,n} - \rho_{13;4,\dots,n} \cdot \rho_{23;4,\dots,n}}{((1 - \rho_{13;4,\dots,n}^2) \cdot (1 - \rho_{23;4,\dots,n}^2))^{1/2}} \quad (2.1)$$

Partial correlations can be computed from correlations using the above recursive formula.

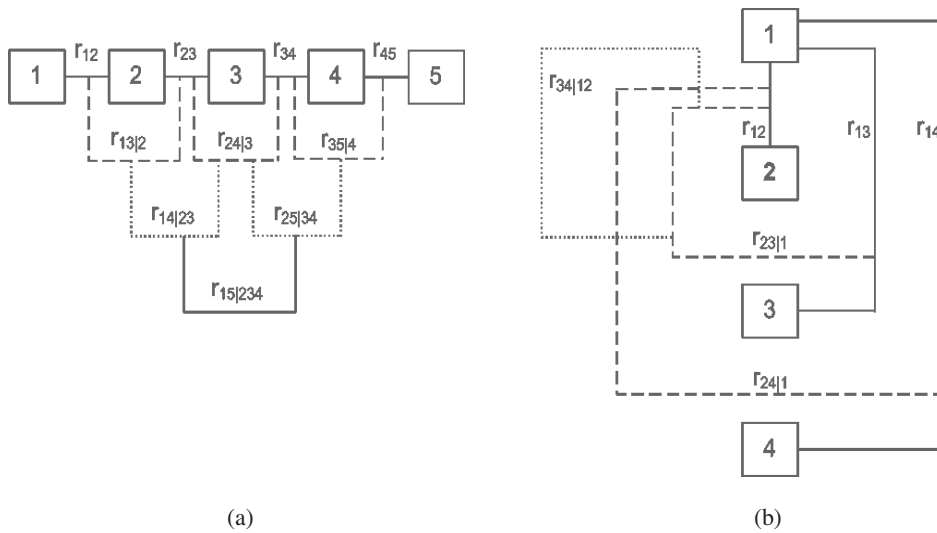


Figure 1. (a) A D-vine on five variables and (b) a C-vine on four variables showing the (conditional) rank correlations associated with the edges.

A complete partial correlation vine specification is a regular vine with a partial correlation specified for each edge[§]. One can remark that in defining a complete partial correlation vine specification there is no reference to a copula as, in general, the partial correlation is not a property of a copula. However, this is not the case for the bivariate normal distribution, hence a complete normal partial correlation specification is a special case of a regular vine specification. We note that for the joint normal variables the partial correlation is just equal to the conditional product moment correlation. The following theorem shows how the notion of a regular vine can be used to construct a joint normal distribution⁷.

Theorem 2.1. *Given any complete partial correlation vine specification for standard normal random variables X_1, \dots, X_n , there is a unique joint normal distribution for X_1, \dots, X_n satisfying all partial correlation specifications.*

Another important result from Bedford and Cooke⁷ is that each partial correlation vine specification uniquely determines the correlation matrix, even without the assumption of joint normality.

Theorem 2.2. *For any regular vine on n elements there is a one to one correspondence between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specifications for the vine.*

In the end of this section we will summarize some well-known facts about the multivariate normal distribution:

- If (X, Y) has the bivariate normal distribution, a necessary and sufficient condition for X and Y to be independent is that $\rho(X, Y) = 0$.
- In the joint normal distribution's case, one can find the relationship between product moment correlation (ρ) and rank correlation (r) using Pearson's transformation.

Proposition 2.1. (Pearson⁸) *Let (X, Y) be a random vector with the joint normal distribution, then*

$$\rho(X, Y) = 2 \sin\left(\frac{\pi}{6} \cdot r(X, Y)\right)$$

[§]For the exact definition of a partial correlation vine specification we refer to Bedford and Cooke⁷.

- With respect to conditioning a multivariate normal, let X be a n -dimensional random vector with *multivariate normal distribution*. Let the vector μ be the expected value of X , and V be its covariance matrix. For a fixed $k < n$ consider the following partition of X , μ and V :

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \quad V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix}$$

where $X_a = (X_1, \dots, X_k)'$, $X_b = (X_{k+1}, \dots, X_n)'$, $\mu_a = (\mu_1, \dots, \mu_k)'$, $\mu_b = (\mu_{k+1}, \dots, \mu_n)'$, $V_{ii} = \text{var}(X_{i \in \{a,b\}})$ and $V_{ab} = \text{cov}(X_a, X_b)$. The conditional variance of X_b given X_a is denoted by $\text{var}_{b|a}(X_b)$.

Proposition 2.2. (Marginal and conditional density function¹¹) *If the partitioned random vector follows the distribution*

$$(X_a, X_b) \sim N \left[(\mu_a, \mu_b), V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix} \right]$$

then:

- (i) *the marginal distribution of X_a is normal with mean μ_a and variance V_{aa} ;*
- (ii) *the conditional distribution of $(X_b|X_a)$ is normal with the mean*

$$E_{b|a}(X_b) = \mu_b + B_{b|a} \cdot (x_a - \mu_a), \quad \text{where } B_{b|a} = V_{ba} V_{aa}^{-1}$$

and the variance

$$\text{var}_{b|a}(X_b) = V_{bb|a} = V_{bb} - V_{ba} V_{aa}^{-1} V_{ab}$$

3. CONTINUOUS BBNS AND VINES

Having put all definitions and concepts in place, let us now consider a non-parametric continuous BBN. As we already said, in this case, the nodes represent continuous univariate random variables and the arcs are associated with (conditional) parent–child rank correlations. We assume throughout that all univariate distributions have been transformed to uniform distributions on $(0, 1)$. The high-dimensional joint distribution is specified using the copula–vine approach. Any copula with invertible conditional cumulative distribution function may be used as long as the chosen copula represents (conditional) independence as zero (conditional) correlation. The conditional rank correlations assigned to the arcs of the BBN are algebraically independent, and there are tested protocols for their use in structured expert judgement (these protocols are presented by Kurowicka and Cooke⁹ (ch. 2)). We note that quantifying BBNs in this way requires assessing all (continuous, invertible) one dimensional marginal distributions. One can assign (conditional) rank correlations to the arcs of a BBN according to the protocol presented by Kurowicka and Cooke⁵. The conditional rank correlations need not be constant, although they are taken to be constant in the following examples. In contrast, in Section 5, where we introduce normal vines, the conditional rank correlations must be constant. We will illustrate the protocol for assigning (conditional) rank correlations to the arcs of a BBN with an example.

Example 3.1. Let us consider the undirected cycle on four variables from Figure 2. There are two sampling orders for this structure: 1, 2, 3, 4, or 1, 3, 2, 4. Let us choose 1, 2, 3, 4. The factorization of the joint distribution is

$$P(1)P(2|1)P(3|\underline{2}1)P(4|3\underline{2}1) \tag{3.1}$$

The underscored nodes are those which are not necessary in sampling the conditioned variable. Hence, the (conditional) rank correlations that need to be assigned to the edges of this BBN are[¶] $\{r_{21}, r_{31}, r_{42}, r_{43|2}\}$.

[¶]One could as well specify $\{r_{21}, r_{31}, r_{43}, r_{42|3}\}$ instead.

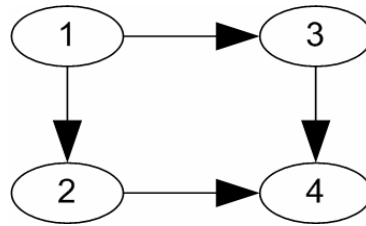


Figure 2. BBN with four nodes and four arcs

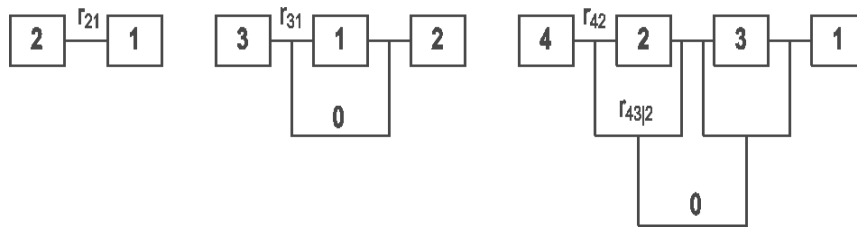


Figure 3. $\mathcal{D}^2, \mathcal{D}^3, \mathcal{D}^4$ for Example 3.1

For each term i ($i = 1, \dots, 4$) of the factorization (3.1), a D-vine on i variables is built. This D-vine is denoted by \mathcal{D}^i and it contains the variable i , the non-underscored variables, and the underscored variables, in this order. Figure 3 shows the D-vines built for variables 2, 3, 4.

Building the D-vines is not a necessary step in specifying the rank correlations^{||}, but it is essential in proving the main result for continuous BBNs. In order to formulate this result, we need a more general setting. For a BBN on n variables the factorization of the joint distribution in the standard way (following the sampling order $1, \dots, n$) is

$$P(1, \dots, n) = P(1)P(2|1)P(3|2, 1) \dots P(n|n - 1, \dots, 1) \tag{3.2}$$

In this factorization, we will underscore the nodes from each conditioning set, which are not parents of the conditioned variable. For each term i with parents (non-underscored variables) $i_1 \dots i_{p(i)}$ in Equation (3.2), we associate the arc $i_{p(i)-k} \rightarrow i$ with the conditional rank correlation:

$$\begin{cases} r(i, i_{p(i)}), & k = 0 \\ r(i, i_{p(i)-k}|i_{p(i)}, \dots, i_{p(i)-k+1}), & 1 \leq k \leq p(i) - 1 \end{cases} \tag{3.3}$$

The assignment is vacuous if $\{i_1 \dots i_{p(i)}\} = \emptyset$. Assigning (conditional) rank correlations for $i = 1, \dots, n$, every arc in the BBN is assigned a (conditional) rank correlation between parent and child.

The following theorem shows that these assignments uniquely determine the joint distribution and are algebraically independent.

Theorem 3.1. *Given the following conditions, the joint distribution of the n variables is uniquely determined:*

1. a directed acyclic graph (DAG) with n nodes specifying conditional independence relationships in a BBN;
2. n variables, assigned to the nodes, with continuous invertible distribution functions;
3. the specification (3.3), $i = 1, \dots, n$, of conditional rank correlations on the arcs of the BBN;
4. a copula realizing all correlations $[-1, 1]$ for which correlation 0 entails independence.

^{||}These are assigned directly to the arcs of the BBN. Each arc is associated with a (conditional) parent-child rank correlation.

This joint distribution satisfies the characteristic factorization (3.2) and the conditional rank correlations in (3.3) are algebraically independent.

Proof. Given that all univariate distributions are known, continuous, invertible functions, one can use them to transform each variable to a uniform on (0, 1). Hence, we can assume, without any loss of generality, that all univariate distributions are uniform distributions on (0, 1).

The first term in (3.3) is determined vacuously. We assume the joint distribution for $\{1, \dots, i-1\}$ has been determined. The i th term of the factorization (3.2) involves $i-1$ conditional variables, of which $\{i_{p(i)+1}, \dots, i_{i-1}\}$ are conditionally independent of i given $\{i_1, \dots, i_{p(i)}\}$. We assign

$$r(i, i_j | i_1, \dots, i_{p(i)}) = 0, \quad i_{p(i)} < i_j \leq i-1 \quad (3.4)$$

Then the conditional rank correlations (3.3) and (3.4) are exactly those on \mathcal{D}^i involving variable i . The other conditional bivariate distributions on \mathcal{D}^i are already determined. It follows that the distribution on $\{1, \dots, i\}$ is uniquely determined. Since zero conditional rank correlation implies conditional independence,

$$P(1, \dots, i) = P(i | 1 \dots i-1) P(1, \dots, i-1) = P(i | i_1 \dots i_{p(i)}) P(1, \dots, i-1)$$

from which it follows that the factorization (3.2) holds. \square

After specifying the joint distribution, we will now show how to sample it. In order to sample a BBN structure we will use the procedures for vines. We can sample X_i using the sampling procedure for the vine \mathcal{D}^i . When using vines to sample a continuous BBN, it is not in general possible to keep the same order of variables in successive D-vines. In other words, we will have to re-order the variables before constructing \mathcal{D}^{i+1} and sampling X_{i+1} , and this will involve calculating some conditional distributions. We will present the sampling procedure for BBNs using the structure from Example 3.1. In Figure 3, one can notice that the D-vine for the third variable is $\mathcal{D}^3 = D(3, 1, 2)$, and the order of the variables from \mathcal{D}^4 must be $D(4, 3, 2, 1)$. Hence, this BBN cannot be represented as just one D-vine. The procedure, which is presented below, starts with sampling four independent, uniform (0,1) variables, say U_1, \dots, U_4 ,

$$\begin{aligned} x_1 &= u_1 \\ x_2 &= F_{r_{21};x_1}^{-1}(u_2) \\ x_3 &= F_{r_{31};x_1}^{-1}(F_{r_{32|1};F_{r_{21};x_1}}^{-1}(u_3)) \\ x_4 &= F_{r_{42};x_2}^{-1}(F_{r_{43|2};F_{r_{32};x_2}}^{-1}(F_{r_{41|32};F_{r_{21|3};F_{r_{32};x_3}}^{-1}(F_{r_{31};x_3}}^{-1}(u_4))) \end{aligned}$$

where $F_{r_{ijk};X_i}(X_j)$ denotes the cumulative distribution function of X_j , given X_i under the conditional copula with correlation $r_{ij|k}$.

The BBN structure reads the conditional independence of X_3 and X_2 given X_1 ($r_{32|1} = 0$), and of X_4 and X_1 given X_2, X_3 ($r_{41|32} = 0$), hence

$$F_{r_{32|1};F_{r_{21};x_1}}^{-1}(u_3) = u_3 \quad \text{and} \quad F_{r_{41|32};F_{r_{21|3};F_{r_{32};x_3}}^{-1}(F_{r_{31};x_3}}^{-1}(u_4)) = u_4$$

Consequently, using these conditional independence properties, the sampling procedure can be simplified as

$$\begin{aligned} x_1 &= u_1 \\ x_2 &= F_{r_{21};x_1}^{-1}(u_2) \\ x_3 &= F_{r_{31};x_1}^{-1}(u_3) \\ x_4 &= F_{r_{42};x_2}^{-1}(F_{r_{43|2};F_{r_{32};x_2}}^{-1}(u_4)) \end{aligned}$$

We shorten the notation by dropping the ‘ r ’ terms and write $F_{j|i}(x_j)$ instead of $F_{r_{ij};x_i}(x_j)$. The conditional distribution $F_{3|2}(x_3)$ is not given, so it must be found by calculating

$$F_{3|2}(x_3) = \int_0^{x_3} \int_0^1 c_{21}(x_2, x_1) c_{31}(v, x_1) dx_1 dv$$

where c_{i1} is the density of the chosen copula with correlation r_{i1} , $i \in \{2, 3\}$. In our study we will use Frank’s¹² copula**.

For each sample, one will need to calculate the numerical value of the double integral^{††}. In this case, when only one double integral needs to be evaluated, it can be easily performed without excessive computational burden.

If some of the variables become certain, the results of sampling this model—conditional on their values—are obtained either by sampling again the structure (cumulative approach), or by using the density approach. We will present a summary of both methods; for details we refer to Kurowicka and Cooke⁹.

Let us assume we learn $X_2 = 0.85$. In the cumulative approach the sampling procedure becomes

$$\begin{aligned} x_1 &= F_{1|2;x_2}^{-1}(u_1) \\ x_2 &= 0.85 \\ x_3 &= F_{3|1;x_1}^{-1}(u_3) \\ x_4 &= F_{4|2;x_2}^{-1}(F_{4|32;F_{3|2}(x_3)}^{-1}(u_4)) \end{aligned}$$

In the density approach, the joint density can be evaluated as follows⁷:

$$g(x_1, \dots, x_4) = c_{21}(x_2, x_1) c_{31}(x_3, x_1) c_{42}(x_4, x_2) c_{43|2}(F_{4|2}(x_4), F_{3|2}(x_3))$$

The conditionalization is made using $x_2 = 0.85$ in the previous formula. Whichever of the two methods is preferred, the double integral still needs to be evaluated for each sample and for any new conditionalization.

If the BBN consists of a cycle* of five variables, and the same sampling procedure is applied, a triple integral will have to be calculated. The larger the cycle is, the larger the number of multiple integrals that have to be numerically evaluated. And yet, this is not the worst that can happen[†]; an example of such a situation will be presented in Section 5 of this paper.

The BBNs that resemble real life problems will often be quite large, and may well contain cycles of five or more variables. Updating such a structure is performed by re-sampling the network each time new evidence is obtained. In the case of a large number of variables, one would have to be prepared to run the model for a few days. To overcome this limitation we would like to combine the vine approach with the continuous BBNs, with the benefits of the discrete BBNs software. This is done in Section 4.

4. HYBRID METHOD

Sampling a large BBN structure every time new evidence becomes available does not seem a very good idea in terms of computational time. On the other hand, sampling it just once, and employing the easiness of use, flexibility, good visualization, and fast updating of a commercial BBN tool, provides an elegant solution to this problem. The hybrid method proposed in this paper can be summarized as follows:

**The reasons for this choice are: it has the zero independence property; it realizes a specified rank correlation without adding to much information to the product of the margins; its density covers the entire unit square; it has tractable functional forms for the density; conditional distribution and inverse of the conditional distribution.

††All numerical results in this paper are obtained using Matlab.

*Whenever we speak of cycles, we mean undirected cycles.

†More examples of BBN structures in which additional numerical calculations are needed are presented by Kurowicka and Cooke⁹ (ch. 6).

1. quantify nodes of a BBN as continuous univariate random variables and arcs as parent–child rank correlations;
2. sample this structure creating a large sample file;
3. use this sample file (in a commercial BBN tool) to build conditional probability tables for a discretized version of the continuous BBN;
4. use the commercial tool to visualize the network and perform fast updating for the discretized BBN.

Most often, when continuous non-parametric BBNs have to be quantified, their discretized version is used instead. A large number of states should be used for each node, in order for the quantification to be meaningful. This leads to huge conditional probability tables that must be filled in, in a consistent manner and, hence, to rapid, informal, and indefensible quantification. In contrast, the first step of the hybrid method can significantly reduce the assessment burden, while preserving the interpretation of arrows as influences. Not only is the degree of realism greater in the continuous model, but also the quantification requires only the marginal distributions and a reduced number of algebraically independent (conditional) rank correlations. After quantifying the continuous model, the discretized version of the model is used. Discretizing the nodes in fairly many states will ensure preserving the dependence structure specified via (conditional) rank correlations. The conditional probability tables for the discretized version of the model are immediately constructed by simply importing the sample file in a commercial BBN tool (third step of the hybrid method). The main use of the BBNs is updating on the basis of newly available information. We have shown how this can be performed using the copula–vine method and what its disadvantages are. This motivates the fourth step of the hybrid method which offers immediate updating.

There is a large variety of BBN software tools. Some of them are free (e.g. Bayda, BNT, BUGS, GeNIe) and others are commercial, although most of the latter have free versions that are restricted in various ways¹³. In our experience, the commercial tools have some advantages over the free versions, either from the functionality point of view or even because the graphical user interface is sometimes not included in the free software. Two of the most popular commercial tools for BBNs are Hugin¹⁴ and Netica¹⁵. They both provide an elegant graphical user interface and their main features are very similar (at least the features that we need to use in our study). Comparing the technical support, the number of references, the efficiency of algorithms, and the price, we chose Netica for our further study.

In order to perform the third step of the hybrid method, a network has to be pre-prepared in Netica. This will contain the nodes of the BBN, each discretized in a certain—not necessarily small—number of states, together with the connections. The way in which variables are discretized is a choice of the analyst. To preserve the information about the dependence structure in the sample file, a large number of discretization intervals is preferred. On the other hand, when the number of discretization intervals increases for each variable, the size of the conditional probability tables that Netica constructs from the sample file increases as well. There is a trade-off between the number of discretization intervals and the size of the conditional probability tables. After a few comparisons (for particular cases) between the choices of 5, 10 and 20 discretization intervals (for each variable), one can observe that the dependence structure assigned by the experts is maintained up to a difference of order 10^{-3} in the case where the variables are discretized in 10 intervals each and that the sample file imported in Netica does not need to be of extraordinary size. Based on this result, the variables from the following examples will be discretized each in 10 intervals. Another choice that one has to make, also with respect to the discretization, is the size of the discretization intervals. The variables can be discretized in equal intervals, or according to the quantiles of their distributions, or at random. The third choice is of course not very useful. After the sample file is imported in Netica, the marginal distributions can be visualized (via the option ‘Style/Belief bars’). If the variables are discretized in equal intervals, the shape representing each variable corresponds to the shape its real distribution. If, on the other hand, the variables are discretized according to their quantiles, Netica will show uniform marginals. We will illustrate the method described above by means of an extensive example.

Example 4.1. (Flight crew alertness) In Figure 4, the flight crew alertness model is given. A discrete form of this model was first presented by Roelen *et al.*¹⁶ and an adapted version of it was discussed by Kurowicka and Cooke⁵. In the original model all chance nodes were discretized to take one of two values *OK* or *NotOK*.

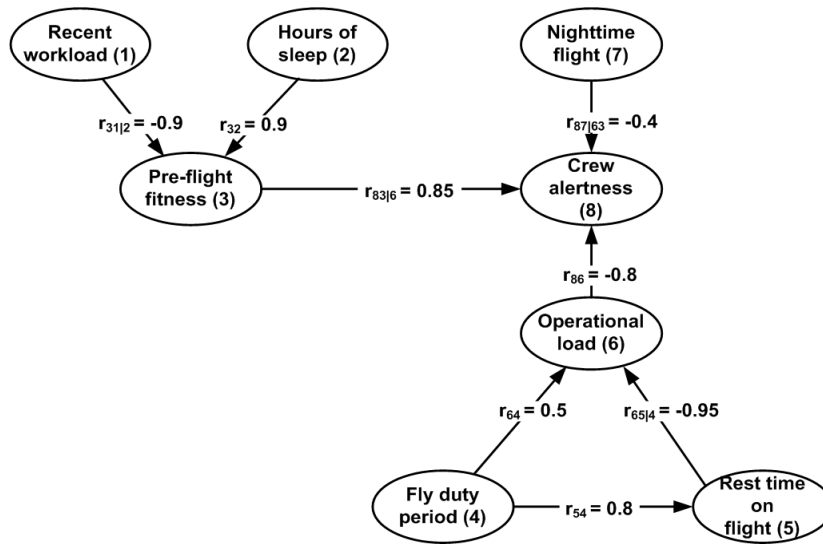


Figure 4. Flight crew alertness model. A number is assigned to each variable (on the right-hand side of each name)

The names of nodes have been altered to indicate how, with greater realism, these can be modelled as continuous variables. Alertness is measured by performance on a simple tracking test programmed on a palmtop computer. Crew members did this test during breaks in-flight under various conditions. The results are scored on an increasing scale and can be modelled as a continuous variable. The alertness of the crew is influenced by a number of factors such as: how much time the crew slept before the flight, the recent work load, the number of hours flown up until this moment in the flight (flight duty period), pre-flight fitness, etc. Figure 4 resembles the latest version of the model.

In order to use the hybrid method described in the beginning of this section, continuous distributions for each node and (conditional) rank correlations for each arc must be gathered from existing data or expert judgement. The distribution functions are used to transform the variables to uniforms on (0, 1). The (conditional) rank correlations assigned to each arc of the BBN are chosen by Kurowicka and Cooke⁵ for illustrative purposes. The marginal distributions are chosen to be uniforms on (0, 1). The sampling order is 1, 2, 3, 4, 5, 6, 7, 8. The sampling procedure uses Frank’s copula, and does not require any additional calculations:

$$\begin{aligned}
 x_1 &= u_1 \\
 x_2 &= u_2 \\
 x_3 &= F_{3|2:x_2}^{-1}(F_{3|21:x_1}^{-1}(u_3)) \\
 x_4 &= u_4 \\
 x_5 &= F_{5|4:x_4}^{-1}(u_5) \\
 x_6 &= F_{6|4:x_4}^{-1}(F_{6|54:F_{5|4}(x_5)}^{-1}(u_6)) \\
 x_7 &= u_7 \\
 x_8 &= F_{8|6:x_6}^{-1}(F_{8|63:x_3}^{-1}(F_{8|763:x_7}^{-1}(u_8)))
 \end{aligned}$$

Figure 5 shows the BBN from Example 4.1, modelled in Netica. The variables are uniform on the (0, 1) interval, and each is discretized in 10 states. A case file containing 8×10^5 samples, obtained using the sampling procedure described, was imported in Netica via the option ‘Relation/Incorporate Case File’. This automatically creates the conditional probability tables.

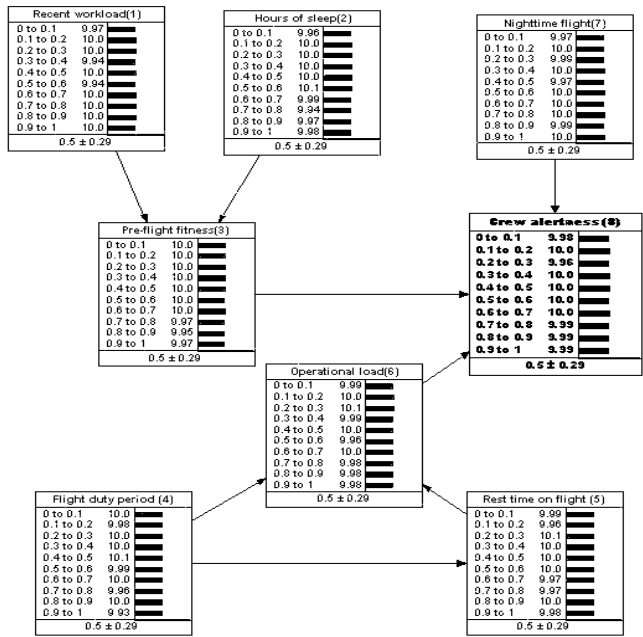


Figure 5. Flight crew alertness model in Netica

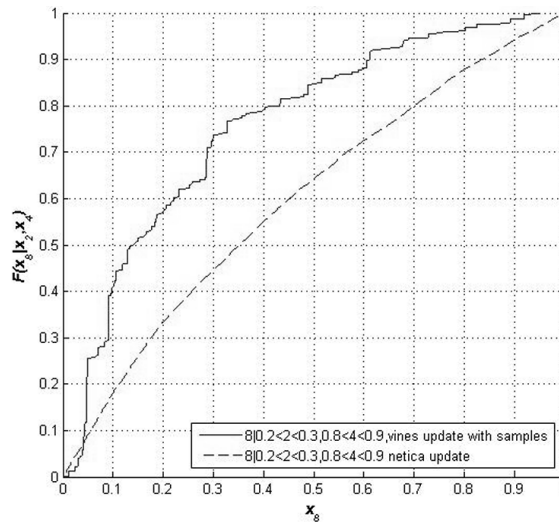


Figure 6. Distribution of $X_8|X_2, X_4$. Comparison of updating results in vines and Netica using 10^4 samples

The quantification of the discretized BBN would require 12 140 probabilities, whereas the quantification with continuous nodes requires only eight algebraically independent (conditional) rank correlations and eight margins.

The main use of BBNs in decision support is updating on the basis of possible observations. Let us suppose that we have some information about how much the crew slept before the flight and about the flight duty period of the crew. Figures 6 and 7 present the distribution of the crew alertness in the situation when the crew’s hours of sleep are between the 20th and the 30th percentiles (the crew did not have enough sleep) and the flight duty period is between the 80th and 90th percentiles (the flight duty period is long).

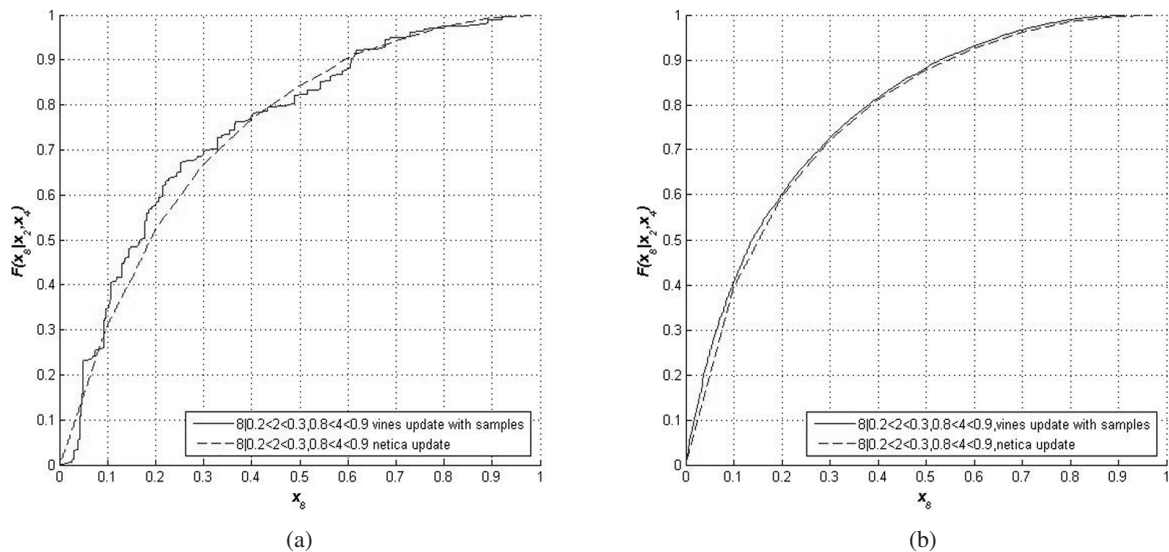


Figure 7. Distribution of $X_8 | X_2, X_4$. Comparison of updating results in vines and Netica (a) using 10^4 from 8×10^5 samples and (b) using 8×10^5 samples

The conditional distribution of the ‘Flight crew alertness(8)’ from Figures 6 and 7 is obtained in two ways:

- using the vines-Netica updating;
- using the vines updating with the density approach.

After the sample file is imported in Netica, we condition on ‘Hours of sleep’ $\in [0.2, 0.3]$ and ‘Fly duty period’ $\in [0.8, 0.9]$. We can use Netica to generate samples from the conditional distribution of ‘Crew alertness’. In the same manner, we sample from ‘Hours of sleep’ $\in [0.2, 0.3]$ and ‘Fly duty period’ $\in [0.8, 0.9]$ and save the samples that Netica generates via the option ‘Network/Simulate Cases’. In the simulation for vines updating, we will have to re-sample the structure, in the same conditions. For better results of the comparisons, we use the samples that we saved from Netica, in the simulation for updating with vines.

In Figure 6, the conditional probability tables from Netica were built using 10^4 samples. The agreement between the two methods is very poor. For example, one can notice from both curves that the combination of the two factors (not enough sleep and a long flight duty period) has an alarming effect on the crew alertness. The difference is that in vines-updating, with probability 50%, alertness is less than or equal to the 15th percentile of its unconditional distribution[‡], whereas in vines-Netica updating with probability 50% alertness is less than or equal to the 35th percentile of its unconditional distribution. This disagreement is due to the number of samples from which Netica calculates the conditional probability tables (10^4). There are 10^3 different input vectors for node 8, each requiring 10 probabilities for the distribution of 8 given the input. With 10^4 samples, we expect each of the 10^3 different inputs to occur 10 times, and we expect a distribution on 10 outcomes to be very poorly estimated with 10 samples. Moreover, updating with vines does not produce a very smooth and accurate curve, also because the simulation was performed with 10^4 samples.

In Figure 7(a), the sample file imported in Netica contains 8×10^5 samples which allows a very good estimation of the conditional distribution of ‘Crew alertness’. Another 10^4 samples for ‘Hours of sleep’ $\in [0.2, 0.3]$ and 10^4 for ‘Fly duty period’ $\in [0.8, 0.9]$ are saved from Netica and used in the vines updating. The curves start to look very similar indeed, but the curve corresponding to vines updating is still not smooth because of the number of samples. If we do everything with the entire sample file of 8×10^5 samples,

[‡] ‘Crew alertness’ is an uniform variable, and therefore its unconditional distribution function is the diagonal of the unit square.

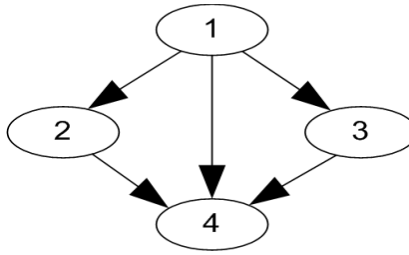


Figure 8. BBN with four nodes and five arcs

the agreement between the two conditional distributions is impeccable (see Figure 7(b)). This motivates the use of a very large sample file.

For a BBN with nodes that require a large number of inputs (large number of parent nodes, discretized in fairly many states) the sample files should also be very large. The main advantage is that this huge sample file needs to be done only once.

Note, however, that in some cases it might happen that sampling the structure, even just once, will cause problems, as we already mentioned in Section 3. We will further present a BBN structure, which at a first glance, seems very easy to deal with, in the sense that it offers a lot of information about the dependence structure.

Example 4.2. Let us consider the BBN from Figure 8. If the set of (conditional) rank correlations that can be elicited is either $\{r_{21}, r_{31}, r_{42}, r_{41|2}, r_{43|21}\}$ or $\{r_{21}, r_{31}, r_{43}, r_{41|3}, r_{42|31}\}$ then the BBN can be represented as one D-vine, and so the sampling procedure does not require any extra calculations. If, for some reason, these rank correlations cannot be specified, and the only correlations available are: $\{r_{21}, r_{31}, r_{43}, r_{42|3}, r_{41|32}\}$ the situation worsens considerably.

The BBN can no longer be represented as one D-vine, since the order of the variables in \mathcal{D}^3 is 3, 1, 2, and in \mathcal{D}^4 is 4, 3, 2, 1. To sample X_4 , one needs to calculate

$$x_4 = F_{4|3;x_3}^{-1}(F_{4|23;F_{2|3}(x_2)}^{-1}(F_{4|123;F_{1|23}(x_1)}^{-1}(u_4)))$$

The conditional distribution $F_{2|3}(x_2)$, can be found by evaluating a double integral as in Example 3.1. Furthermore, $F_{1|23}(x_1)$ needs to be calculated. This is, in fact, the conditional distribution of $F_{1|2}(x_1)$, given $F_{3|2}(x_3)$. Even though all the information needed seems to be available, evaluating the joint distribution of these two quantile functions turns out to be very difficult. Moreover, at each step of its evaluation, one should calculate the numerical value of the double integral for $F_{3|2}(x_3)$. This is a task that takes time and patience.

If this kind of calculation is necessary for such a small BBN, it is very likely that more complicated calculations will be involved in larger structures. The time spent to solve this sort of problems would be, by far, much longer than one can afford.

5. NORMAL COPULA-VINE APPROACH

All the troubles discussed until now are caused by the different sampling order of variables from one vine to another. To avoid these problems we advance here a new way of realizing the rank correlation specification on a regular vine using the joint normal distribution.

Let us start with a rank correlation vine specification on the variables X_1, \dots, X_n , with continuous, invertible distribution functions F_1, \dots, F_n . We adopt the following protocol.

1. Transform X_1, \dots, X_n to the standard normal variables Y_1, \dots, Y_n via the transformation $Y_i = \Phi^{-1}(F_i(X_i))$ ($\forall i$) ($i = 1, \dots, n$), where Φ is the cumulative distribution function of the standard normal distribution.

2. Construct the vine for the standard normal variables Y_1, \dots, Y_n . Since $\Phi^{-1}(F_i(X_i))$ are strictly increasing transformations, the same (conditional) rank correlations correspond to the edges of this vine.
3. To each edge of this vine assign $\rho_{i,j|D} = 2 \sin(\pi \cdot r_{i,j|D}/6)$, where $\{i, j\}$ and D are the conditioned and conditioning sets, respectively, of the edge, and $r_{i,j|D}$ is the conditional correlation assigned to the corresponding edge from the initial vine. We now have a complete partial correlation vine specification[§] for Y_1, \dots, Y_n . Theorem 2.1 ensures that there is a unique joint normal distribution for Y_1, \dots, Y_n satisfying all partial correlation specifications. Moreover there is a unique correlation matrix determined by this vine (Theorem 2.2).
4. Compute the correlation matrix R using the recursive formula (2.1).
5. Sample the joint normal distribution of Y_1, \dots, Y_n , with correlation matrix R (Tong¹⁷).
6. For each sample, calculate $(F_1^{-1}(\Phi(y_1^j)), F_2^{-1}(\Phi(y_2^j)), \dots, F_n^{-1}(\Phi(y_n^j)))$, where $((y_1^j), (y_2^j), \dots, (y_n^j))$ is the j th sample from the previous step.

In this way we realize the joint distribution of the initial variables X_1, \dots, X_n , together with the dependence structure specified.

The normal copula–vine method might seem very similar to the joint normal transform method presented by Ghosh and Henderson¹⁸ and Imam and Helton¹⁹, but the presence of vines is crucial in avoiding the problems encountered in the latter method. In the joint normal transform approach, the rank correlation matrix must be first specified and then induced by transforming distributions to standard normals and generating a dependence structure using the linear properties of the joint normal. In absence of data, specifying a rank correlation matrix can be a very difficult task. Moreover, it is not always possible to find a product moment correlation matrix generating a given rank correlation matrix via Pearson’s transformation, as shown by Kurowicka and Cooke⁹ (ch. 4). Using the normal copula–vine approach we avoid these problems because we do not specify a rank correlation matrix, but rather a rank correlation vine. Moreover, for such a specification all assignments of numbers between -1 and 1 are consistent.

In case of a BBN that cannot be represented as one vine, we can make use of the protocol described above. Everything is calculated on the joint normal vine, hence we can re-order the variables and recompute all partial correlations needed. We expect a dramatic decrease in the computational time using this method. We note that the assumption of constant conditional rank correlations, previously a matter of convenience, is now required.

Furthermore, we will present comparisons between the normal copula vine method and the copula–vine method together with Netica updating, using the BBN from Example 3.1.

The marginal distributions of X_1, X_2, X_3, X_4 are uniform on the interval $(0, 1)$. We sample the structure both with the copula–vine, and the normal copula–vine approach. Hence, we produce two sample files, each containing 10^5 samples. The resulting files are imported in Netica, and conditioning is performed in both cases. Figure 9 presents the conditional distribution of the variable X_4 given that $X_1 \in [0.1, 0.2]$ and $X_2 \in [0.3, 0.4]$, obtained using the sample files produced with the two methods. One can notice a small disagreement between the two conditional distributions. If we think of the normal copula–vine method in terms of the copula–vine method, where we made use of the normal copula, we can say that the difference between the two conditional distributions from Figure 9 is due to the different choice of the copula.

Another way of comparing these methods is to calculate and compare the two sample correlation matrices. The matrices presented below correspond to the sample file obtained using the copula–vine approach (left) and the sample file generated with the normal copula–vine method (right):

$$\begin{pmatrix} 1 & 0.4031 & 0.7028 & 0.3746 \\ 0.4031 & 1 & 0.2843 & 0.2028 \\ 0.7028 & 0.2843 & 1 & 0.5201 \\ 0.3746 & 0.2028 & 0.5201 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0.4000 & 0.6974 & 0.3843 \\ 0.4000 & 1 & 0.2837 & 0.1985 \\ 0.6974 & 0.2837 & 1 & 0.5271 \\ 0.3843 & 0.1985 & 0.5271 & 1 \end{pmatrix}$$

[§]As we already mentioned, conditional and partial correlations are equal for normal variables.

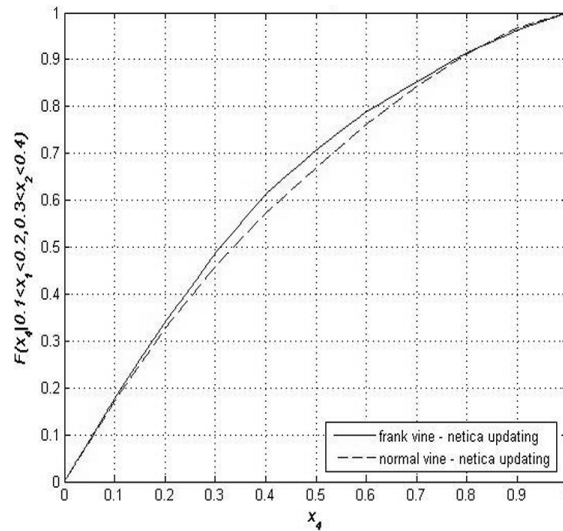


Figure 9. The distribution of $X_4|X_1, X_2$. Frank's copula-vine versus normal copula-vine (conditioning in Netica using 10^5 samples)

Comparing these two matrices one can observe differences of order 10^{-3} , which represent a reasonable result taking into account the sampling errors.

The main advantage of the normal copula-vine method is that the simulation runs for a few seconds, whereas with the previous sampling algorithm (in which a double integral needs to be numerically evaluated for each sample) the results were available in hours. Both methods were implemented in Matlab for a fair comparison of the computational times. The normal copula-vine method is implemented in a new software application, called UniNet[¶]. UniNet allows for quantification of non-parametric continuous/discrete BBNs. The program has a friendly interface and the simulations are very fast (faster than Matlab).

The same kind of results we find when we examine the structure from Example 4.1. Figure 10 shows the conditional distribution of the variable 'Crew alertness(8)' given that 'Hour of sleep(2)' is in the interval $[0.2, 0.3]$ and 'Fly duty period(4)' $\in [0.8, 0.9]$. We can again notice that the choice of the copula produces a small discrepancy between the curves. Comparing the two sample correlation matrices for this example we find that the maximum difference is 8×10^{-3} .

6. ANALYTICAL UPDATING

A very important feature of the normal copula-vine method is that conditioning can be performed analytically. Since all the calculations are performed on a joint normal vine, any conditional distribution will also be normal with mean and variance given by the formulas in Proposition 2.2. Finding the conditional distribution of the corresponding original variable will just be a matter of transforming it back using the inverse distribution function of this variable and the standard normal distribution function.

Proposition 6.1. *Let X_1 and X_2 be random variables with continuous, invertible distribution functions F_1 and F_2 . Let Y_1 and Y_2 be the transformation of X_1 and X_2 to standard normal variables. Then the conditional distribution $X_1|X_2$ can be calculated as $F_1^{-1}(\Phi(Y_1|Y_2))$, where Φ is the cumulative distribution function of the standard normal distribution.*

[¶]Figures 2, 4 and 8 are made using UniNet.

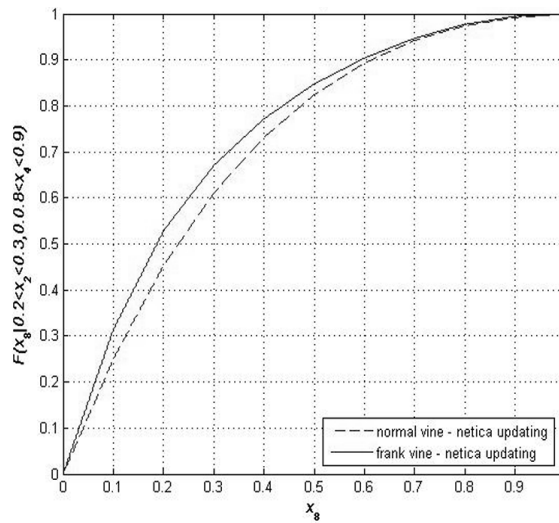


Figure 10. The distribution of $X_8|X_2, X_4$. Frank’s copula–vine versus normal copula–vine (conditioning in Netica using 8×10^5 samples)

Proof. For $i \in \{1, 2\}$, $Y_i = \Phi^{-1}(F_i(X_i))$, and therefore we can write $X_i = F_i^{-1}(\Phi(Y_i))$.

Remark. For A, B, C random variables and f a function such that $A = f(B)$, then $A|C = f(B|C)$.

We will use the above remark for $X_1, X_2, Y_1, Y_2, F_1^{-1} \circ \Phi$ and x_2 an arbitrary value of X_2 :

$$\begin{aligned} X_1|(X_2 = x_2) &= F_1^{-1}(\Phi(Y_1|(X_2 = x_2))) = F_1^{-1}(\Phi(Y_1|(F_2^{-1}(\Phi(Y_2) = x_2)))) \\ &= F_1^{-1}(\Phi(Y_1|(Y_2 = \Phi^{-1}(F_2(x_2)))))) = F_1^{-1}(\Phi(Y_1|(Y_2 = y_2))) \end{aligned}$$

where we denoted $\Phi^{-1}(F_2(x_2)) = y_2$. □

Let us illustrate this result on the BBN structure from Example 4.1. In Figure 11 we present a comparison between updating in the normal copula–vine method and the copula–vine method. The updating is performed both in Netica and analytically.

As one would expect, the pairs of curves corresponding to the two methods (copula–vine and normal copula–vine) follow exactly the same patterns regardless of the way we perform conditioning. The distance between the pairs of curves is caused by the different choice of the copula.

We will now consider (for the same example) the univariate distributions to be standard normals instead of uniforms on (0, 1). The same kind of comparisons as before are performed. In doing so, a new model should be pre-prepared in Netica. The differences between the new model and that presented in Figure 5 are the range of the variables and the discretization intervals. We will keep the same number of intervals for the discrete version of each variable, only they will not be equally sized anymore. The variables are discretized with respect to the quantiles of their distributions.

We conditionalize on ‘Hours of sleep’ between its 0.2 and 0.3 quantiles and ‘Fly duty period’ between its 0.8 and 0.9 quantiles^{||}. The conditional distribution of the ‘Crew alertness’, obtained with the methods previously discussed, is presented in Figure 12(a).

^{||}In the previous comparisons(for uniform marginals) the conditioning was ‘Hours of sleep’ $\in [0.2, 0.3]$ and ‘Fly duty period’ $\in [0.8, 0.9]$.

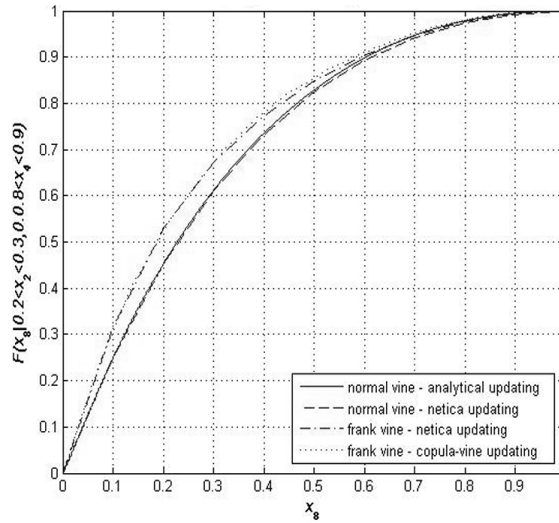


Figure 11. The distribution of $X_8|X_2, X_4$. Comparison of updating results in Frank’s copula–vine (using Netica and the copula–vine updating) versus updating in normal copula–vine (using Netica and analytically). All univariate distributions are uniforms on (0, 1)

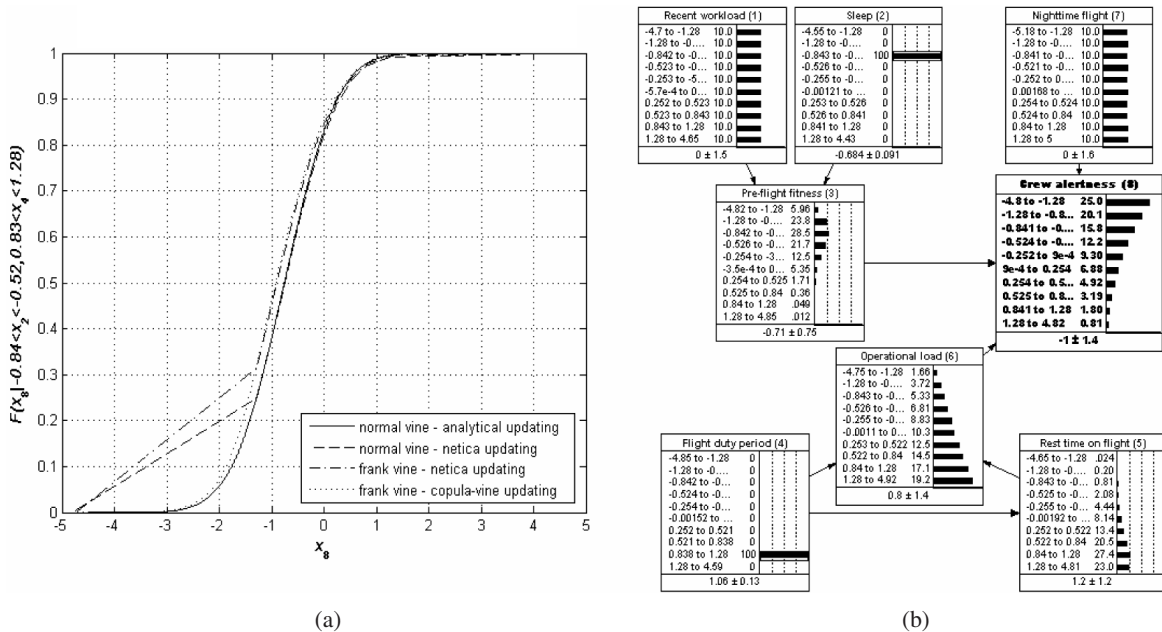


Figure 12. The distribution of $X_8|X_2, X_4$. (a) Comparison of updating results in Frank’s copula–vine (using Netica and the copula–vine updating) versus updating in normal copula–vine (using Netica and analytically). (b) The *Flight Crew Alertness* model in Netica. All univariate distributions are standard normals

The curves nicely agree everywhere, except for the first interval of the discretization, where the results given by Netica updating, in both methods, are completely different from the results of the analytical updating**.

**The word ‘analytical’ is appropriate only for the normal copula–vine method. For the copula–vine method, updating is performed via re-sampling the structure.

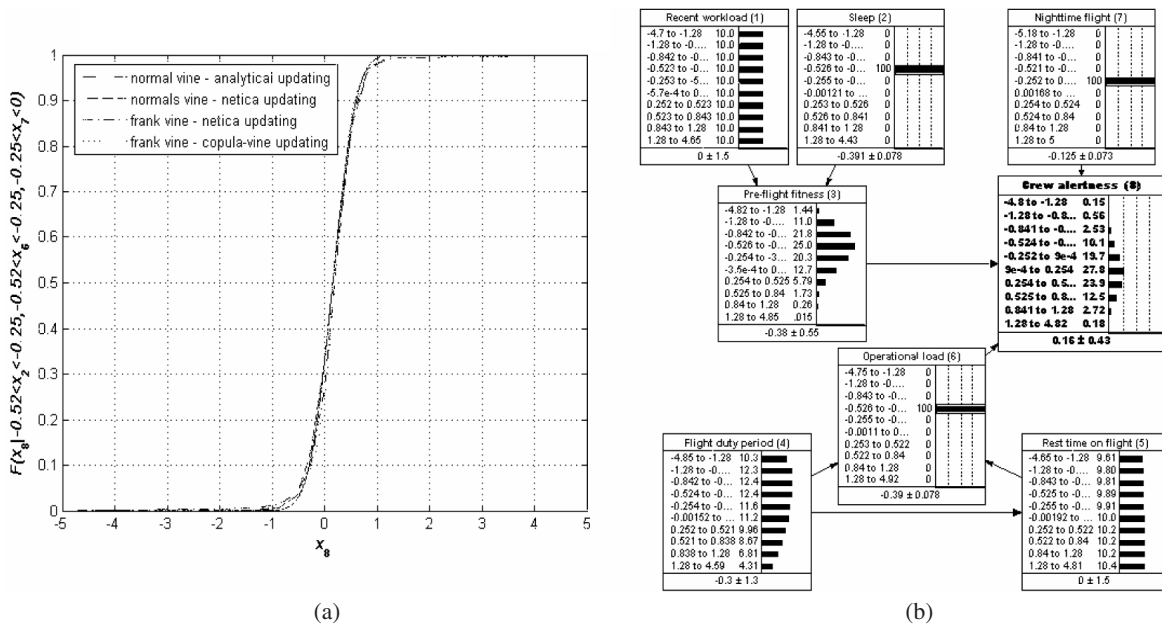


Figure 13. The distribution of $X_8|X_2, X_6, X_7$. (a) Comparison of updating results in Frank’s copula–vine (using Netica and the copula–vine updating) versus updating in normal copula–vine (using Netica and analytically). (b) The *Flight Crew Alertness* model in Netica. All univariate distributions are standard normals

As already stated, the discretization of the variables was made according to their quantiles, hence the first interval and the last one (for each variable) are much wider than the rest of the intervals. This can be noticed in Figure 12(b), which shows the *Flight Crew Alertness* structure in Netica, after we updated the model. A sample file of 8×10^5 , obtained with the normal copula–vine method, was imported in Netica in order to create the conditional probability tables. For the variable ‘Crew alertness’, the first and the last discretization intervals are approximately 12 times wider than the rest of the intervals from its discretization (see Figure 12(b)). In order to plot the conditional distribution of ‘Crew alertness’ given by Netica, one will need to generate samples from it. Netica simply samples uniformly from each discretization interval, taking into account its probability. The information that most of the samples from the first interval should be concentrated in its right-hand side, is not included. Therefore, the first part of each of the curves given by Netica does not resemble reality. The same kind of discrepancy would happen in the last interval if its probability were larger.

We will further consider another updating of the same model. We conditionalize on ‘Hours of sleep’ between its 0.3 and 0.4 quantiles, ‘Operational load’ between its 0.3 and 0.4, and ‘Nighttime flight’ between its 0.4 and 0.5 quantiles. Figure 13(b) shows the structure in Netica, after updating. Looking at the conditional probability of ‘Crew alertness’, one can notice that the first and the last discretization intervals have very small probabilities. In these conditions, we expect the curves for the conditional distribution of the ‘Crew alertness’, obtained with the four methods, to be very similar on the entire domain. As Figure 13(a) shows, there is perfect agreement between the methods.

Although in most cases Netica updating has a reasonable outcome, in some particular cases, its results are not to be trusted. Therefore the opportunity to perform analytical updating is a key advantage of the normal vine method.

7. CONCLUSIONS

We have presented the copula–vine approach to continuous non-parametric BBNs, addressed its computational problems and proposed ways to solve them.

Quite often, real life problems are modelled with large BBNs, which might contain large cycles. In most cases, in the copula–vine approach extra calculations are needed in order to sample a structure. These calculations are numerical evaluations of multiple integrals, which are very time consuming. Moreover, updating such a structure is performed by re-sampling the network each time a new policy is evaluated. Hence, even though the copula–vine approach to continuous non-parametric BBNs provides a practical and flexible way of stipulating a high-dimensional distribution, it has the disadvantage that a large model could run for days.

In this paper, we have shown how to overcome this limitation by making use of the advantages of BBNs software. We sampled the structure just once, imported the sample file in Netica and performed fast updating for the discretized version of the BBN, without losing information about the dependence structure specified in a coherent and defensible way. We have chosen Netica for the reasons described earlier (in Section 4) but any software can be used for the purposes presented in this paper, as long as it has the same features as Netica. The results presented here are independent of the choice of software.

For a large BBN, which contains nodes with many parents, discretized in fairly many states, a very large sample file is needed in Netica. Furthermore, we presented a very simple example in which unexpected problems appeared and made this one time sampling very complicated and time consuming. If this happens for a small BBN, it is very likely that for a larger calculation, more difficulties will emerge.

These problems are caused by the different ordering of variables in the D-vines that represent the BBN structure. We solved this issue by realizing the rank correlation specification on a joint normal vine. We transformed the rank correlation vine to a partial correlation vine on standard normal variables; computed the correlation matrix and sampled from a joint normal distribution with standard normal margins and the calculated correlation matrix. We transformed back, and in this way, we realized the joint distribution of the initial variables with the specified dependence structure. In the joint normal vine we can re-order the variables and compute the re-ordered partial correlations, using the properties of the joint normal distribution. Hence, no extra calculations are involved. The main advantage is that the computational time reduces from *hours* to *seconds*.

The normal copula–vine method can be used to sample any non-parametric continuous BBN and stipulate its joint distribution in a fast and flexible way.

A very attractive feature of the normal vine sampling algorithm is that conditioning can be performed analytically. However, the advantages of porting to a familiar BBN platform with friendly support and post processing options should not be underestimated.

The theory presented here can be extended to include ‘ordinal’ variables; that is variables which can be written as monotone transforms of uniform variables, perhaps taking finitely many values. The dependence structure must be defined with respect to the uniform variates. The case of dichotomous variables (and generally variables with few states), and the rank correlations between two such variables, is still an open issue. We still have to answer questions such as ‘is it possible to realize all rank correlations between this kind of variables?’. Another subject to investigate is the possibility of specifying functional relationships between variables.

REFERENCES

1. Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. *Probabilistic Networks and Expert Systems (Statistics for Engineering and Information Sciences)*. Springer: New York, 1999.
2. Edwards W. Tool for and experiences with Bayesian normative modeling. *Journal of the American Psychological Association* 1998; **53**(4):416–428.
3. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann: San Mateo, CA, 1988.
4. Shachter RD, Kenley CR. Gaussian influence diagrams. *Management Science* 1989; **35**(5):527–550.
5. Kurowicka D, Cooke RM. Distribution-free continuous Bayesian belief nets. *Proceedings of the Mathematical Methods in Reliability Conference*, June 2004.
6. Cooke RM. Markov and entropy properties of tree- and vine-dependent variables. *Proceedings of the Section on Bayesian Statistical Science*, 1997. American Statistical Association: Alexandria, VA, 1997.
7. Bedford TJ, Cooke RM. Vines—a new graphical model for dependent random variables. *Annals of Statistics* 2002; **30**(4):1031–1068.

8. Pearson K. *Mathematical Contributions to the Theory of Evolution IV (Drapers' Company Research Memoirs: Biometric Series)*. Cambridge University Press: Cambridge, 1907.
9. Kurowicka D, Cooke RM. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley: New York, 2006.
10. Yule GU, Kendall MG. *An Introduction to the Theory of Statistics* (14th edn). Charles Griffin: Belmont, CA, 1965.
11. Whittaker J. *Graphical Models in Applied Multivariate Statistics*. Wiley: Chichester, 1990.
12. Frank MJ. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae* 1979; **19**:194–226.
13. Murphy KP. An introduction to graphical models. *Technical Report*, Intel Research, 2002.
14. Hugin. <http://www.hugin.com> [20 June 2006].
15. Netica. <http://www.norsys.com> [20 June 2006].
16. Roelen ALC, Wever R, Hale AR, Goossenes LHJ, Cooke RM, Lopuhaa R, Simons M, Valk PJJ. Casual modeling using Bayesian belief nets for integrated safety at airports. *Risk Decision and Policy* 2004; **9**:207–222.
17. Tong YL. *The Multivariate Normal Distribution*. Springer: New York, 1990.
18. Ghosh S, Henderson SG. Properties of the NOTRA method in higher dimensions. *Proceedings of the 2002 Winter Simulation Conference*, Yücesan E, Chen CH, Snowdon JL, Charnes JM (eds.). IEEE Press: Piscataway, NJ, 2002; 263–269.
19. Iman R, Helton J. A comparison of uncertainty and sensitivity analysis techniques for computer models. *Technical Report NUREG/CR-3904 SAND84-1461 RG*, Sandia National Laboratory, Albuquerque, NM, 1985.

Authors' biographies

Anca M. Hanea graduated from the Faculty of Mathematics, University of Bucharest, in 2001. She received her Master's degree in Risk and Environmental Modelling from the Delft University of Technology in 2004. She is a PhD candidate at the Delft University of Technology, Department of Mathematics. Her research interest is non-parametric continuous/discrete BBNs.

Dorota Kurowicka studied mathematics at Gdansk University in Gdansk and specialized in numerical methods. Her Master thesis was completed in 1990. She performed PhD research at the Gdynia Maritime Academy and the Delft University of Technology, and received her PhD from Delft in 2000. Since then she has been employed as Assistant Professor at the Delft University of Technology, Department of Mathematics. Her current research is carried out in the area of uncertainty and sensitivity analysis.

Roger M. Cooke received his bachelors and doctoral degrees in mathematics and philosophy at Yale University. He taught logic and philosophy of science at Yale University, the University of Amsterdam and the Delft University of Technology. In 1994 he was appointed to the chair of Applications of Decision Theory in the Department of Mathematics at Delft. His current interests are methodological aspects of risk analysis.