

# *Vines and Continuous Non-parametric Bayesian Belief Nets with Emphasis on Model Learning*

**D. Kurowicka; Roger M. Cooke**

Delft University of Technology;  
Resources for the Future and Delft University of Technology

Regular vines are graphical structures for picking out bivariate and conditional bivariate margins which uniquely determine a joint distribution. They enable much greater modelling flexibility than familiar families of multivariate distributions, and they easily translate into sampling algorithms. Recent work (Aas *et al* 2009) has focused on estimating parameters for the (conditional) bivariate copulas once a particular regular vine is chosen. Another line of research (Kurowicka and Joe 2010) has focused on optimally choosing a regular vine. Mathematical results on vines have also motivated developments in the theory of directed acyclic graphs (DAGs), or Bayesian belief nets (BBNs), leading to continuous non-parametric Bayesian belief nets (Hanea *et al* 2006). Whereas regular vines provide adequate tools for analysing distributions whose dimensionality runs into tens of variables, BBNs have proven adequate for distributions with several hundred variables. This capability is achieved by restricting to the normal copula; and the usefulness of the graphical representation depends on having a great deal of conditional independence. Under these restrictions, the BBNs can offer modelling advantages. Recent applications of regular vines and BBNs to multivariate data sets and multivariate spatial distributions have yielded very encouraging results (see Kurowicka and Joe (2010) for broad exposition of vines and its applications). This article is based on Cooke *et al* (2007). The

A short (100 words or so) biography is needed for the front matter of the book.

Change OK?

Change OK?

Author: please check running head on odd pages - abbreviated form of title OK?

Changes to this paragraph OK? Book style is for un-numbered sections.

first three sections summarise definitions of regular vines and summarise important results. The fourth section (see page 620 onwards) is devoted to non-parametric continuous BBNs and the fifth section (see page 624) studies model learning for BBNs and regular vines. Optimisation heuristics are illustrated with hourly wind speed data from seven wind stations in The Netherlands in the sixth section (see page 629) and we present our conclusions in the final section.

### DEFINITIONS AND NOTATION

The product moment correlation of random variables  $X, Y$  with finite expectations  $E(X), E(Y)$  and finite variances  $\sigma_X^2, \sigma_Y^2$ , also called linear or Pearson correlation is defined as

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

The product moment correlation can be also defined in terms of regression coefficients as follows. Let us consider  $X$  and  $Y$  with means zero. Let  $b_{XY}$  minimise

$$E((X - b_{XY}Y)^2)$$

Word added – OK?

and let  $b_{YX}$  minimise

$$E((Y - b_{YX}X)^2)$$

Then

$$\rho(X, Y) = \text{sgn}(b_{XY}) \sqrt{b_{XY} b_{YX}}$$

The product moment correlation measures degree of linear relationship between random variables.

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

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

The rank correlation measures degree of monotonic relationship between random variables. It always exists and does not depend on marginal distributions.

By definition, product moment and rank correlations are equal for uniform variables, but in general they are different.

The partial correlation can be defined in terms of partial regression coefficients. Consider variables  $X_i$  with zero mean and standard

deviations  $\sigma_i, i = 1, \dots, n$ , and let the numbers  $b_{12;3,\dots,n}, \dots, b_{1n;2,\dots,n-1}$  minimise

$$E((X_1 - b_{12;3,\dots,n}X_2 - \dots - b_{1n;2,\dots,n-1}X_n)^2)$$

$$\rho_{12;3,\dots,n} = \text{sgn}(b_{12;3,\dots,n})\sqrt{b_{12;3,\dots,n}b_{21;3,\dots,n}}$$

etc.

Equivalently, we could define the partial correlation as

$$\rho_{12;3,\dots,n} = -\frac{C_{12}}{\sqrt{C_{11}C_{22}}}$$

where  $C_{ij}$  denotes the  $(i, j)$ th cofactor of the correlation matrix (that is, the determinant of the submatrix obtained by removing row  $i$  and column  $j$  multiplied by  $(-1)^{i+j}$ ). The partial correlation  $\rho_{12;3,\dots,n}$  can be interpreted as the correlation between the orthogonal projections of  $X_1$  and  $X_2$  on the plane orthogonal to the space spanned by  $X_3, \dots, X_n$ . Change OK?

Partial correlations can be computed from correlations with the following recursive formula (Yule and Kendall 1965)


$$\rho_{12;3,\dots,n} = \frac{\rho_{12;3,\dots,n-1} - \rho_{1n;3,\dots,n-1} \cdot \rho_{2n;3,\dots,n-1}}{\sqrt{1 - \rho_{1n;3,\dots,n-1}^2}\sqrt{1 - \rho_{2n;3,\dots,n-1}^2}}. \quad (24.1)$$

The conditional correlation of  $Y$  and  $Z$  given  $X$

$$\rho_{YZ|X} = \rho(Y | X, Z | X)$$

$$= \frac{E(YZ | X) - E(Y | X)E(Z | X)}{\sigma(Y | X)\sigma(Z | X)}$$

is the product moment correlation computed with the conditional distribution of  $Y$  and  $Z$  given  $X$ .

For joint no  distribution, partial and conditional correlations are equal. In general, however, partial and conditional correlations are not equal and the difference can be large.

In the following definition we consider a measure for assessing the difference between different probability densities.

**Definition 24.1 (relative information, mutual information).** Let  $f$  and  $g$  be densities on  $\mathbb{R}^n$  with  $f$  absolutely continuous with respect to  $g$ .

- The relative information of  $f$  with respect to  $g$  is

$$I(f | g) = \int \dots \int f(x_1, \dots, x_n) \ln \left( \frac{f(x_1, \dots, x_n)}{g(x_1, \dots, x_n)} \right) dx_1 \dots dx_n$$

- The mutual information of  $f$  is

$$MI(f) = I(f | \prod_{i=1}^n f_i)$$

where  $f_i$  is the  $i$ th univariate marginal density of  $f$  and  $\prod_{i=1}^n f_i$  is the independent distribution with univariate margins  $f_i$ .

Relative information is also called the Kullback–Leibler information, relative entropy or the directed divergence. Note that  $I(f | g) \neq I(g | f)$ . According to the definition above, the relative information measures the distance of the density  $g$  to density  $f$  with respect to an expectation over  $f$ . It does not depend on margins of  $f$ . The mutual information is also called the “information proper”. The mutual information will be used to capture general dependence in a set of multivariate data. The following makes the mutual information the appropriate measure of multivariate dependence.

**Proposition 24.2.**  $MI(f) = 0$  if and only if  $f = \prod f_i$  (independence) and  $MI(f) = \infty$  if  $f$  has positive mass on a set of  $\prod f_i$  measure zero (complete dependence).

For the normal distribution  $e^{-2MI(f)}$  is equal to the determinant of the correlation matrix (Whittaker 1990).

**Theorem 24.3.** Let  $f$  be a joint normal density with mean vector zero. Then

$$MI(f) = -\frac{1}{2} \ln(D)$$

where  $D$  is the determinant of the correlation matrix.

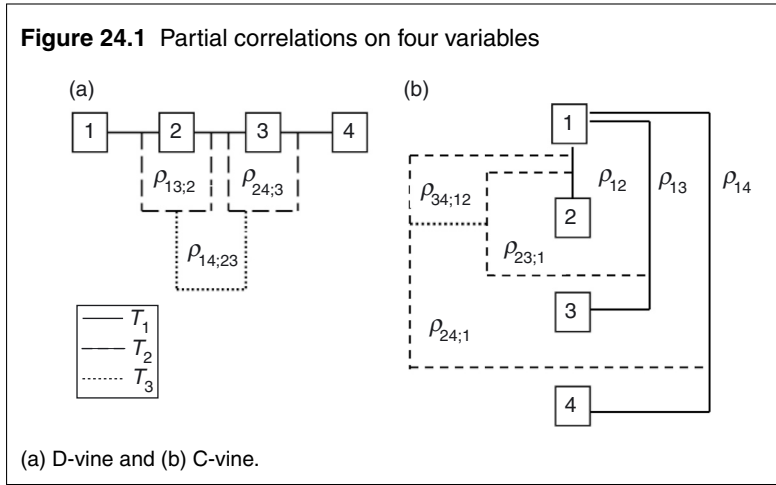
We do not possess something like an “empirical mutual information”. Rather, it must be estimated with kernel estimators, as suggested in Joe (1993). Hence, according to Theorem 24.3 and Proposition 24.2, the determinant of the correlation matrix can be considered as a “good” approximation of the mutual information specially in high-dimensional problems.

## VINES

Graphical models called “vines” were introduced in Cooke (1997), Bedford and Cooke (2002) and Kurowicka and Cooke (2003).

Greek 'Pi' OK here and later? Please check all Greek Pi and all products [] have been typeset correctly throughout.

I have assumed this 'e' denotes the exponential function and set it upright according to book style – OK? Please mark any that denote variable  $e$ .



**Definition and properties**

A vine  $\mathcal{V}$  on  $n$  variables is a nested set of connected trees  $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$ , where the edges of tree  $j$  are the nodes of tree  $j + 1$ ,  $j = 1, \dots, n - 2$ . A “regular vine” on  $n$  variables is a vine in which two edges in tree  $j$  are joined by an edge in tree  $j + 1$  only if these edges share a common node,  $j = 1, \dots, n - 2$  (see the proximity condition below). The formal definitions follow.

**Definition 24.4 (regular vine).**  $\mathcal{V}$  is a regular vine on  $n$  elements if the following hold.

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

A regular vine is called a canonical vine or “C-vine” if each tree  $T_i$  has a unique node of degree  $n - i$ , and hence has a maximum degree.<sup>1</sup> A regular vine is called a “D-vine” if all nodes in  $T_1$  have degree not higher than 2 (see Figure 24.1). There are  $n(n - 1)/2 = \binom{n}{2}$  edges in a regular vine on  $n$  variables. An edge in tree  $T_j$  is an unordered pair of nodes of  $T_j$  or, equivalently, an unordered pair of edges of  $T_{j-1}$ . By definition, the order of an edge in tree  $T_j$  is  $j - 1$ ,  $j = 1, \dots, n - 1$ .

Changes to sentence OK?

Change OK?

All figures have been relettered. Please check all figures and captions carefully.

The above definition is explained using the D-vine in Figure 24.1, where edges of each tree are distinguished by different line styles. There are three trees in the D-vine on four variables

Change OK?

$$\mathcal{V} = \{T_1, T_2, T_3\}$$

$$T_1 = (N_1, E_1): N_1 = \{1, 2, 3, 4\}; E_1 = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$$

$$T_2 = (N_2, E_2): N_2 = E_1; E_2 = \{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{3, 4\}\}\}$$

$$T_3 = (N_3, E_3): N_3 = E_2; E_3 = \{\{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{3, 4\}\}\}\}$$

A regular vine is just a way of identifying a set of conditional bivariate constraints. Conditional bivariate constraints are associated with each edge and are determined as follows: the variables reachable from a given edge via the membership relation are called the constraint set of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the conditioning variables, and the symmetric differences of the constraint sets are the conditioned variables. More precisely, the constraint, the conditioning and the conditioned set of an edge can be defined as follows.

**Definition 24.5 (constraint, conditioning and conditioned sets).**

For  $e \in E_i, i \leq n - 1$ , the constraint set associated with edge  $e$  is the complete union  $U_e^*$  of  $e$ , that is, the subset of  $\{1, \dots, n\}$  reachable from  $e$  by the membership relation.

For  $i = 1, \dots, n - 1, e \in E_i$ , if  $e = \{j, k\}$ , then the conditioning set associated with  $e$  is

$$D_e = U_j^* \cap U_k^*$$

and the conditioned set associated with  $e$  is

$$\{C_{e,j}, C_{e,k}\} = \{U_j^* \setminus D_e, U_k^* \setminus D_e\}$$

Note that, for  $e \in E_1$ , the conditioning set is empty. We can see that the order of an edge is the cardinality of its conditioning set. For  $e \in E_i, i \leq n - 1, e = \{j, k\}$  we have  $U_e^* = U_j^* \cup U_k^*$ .

Change OK?

For the D-vine in Figure 24.1 the complete union of  $j = \{1, 2\}$  is  $U_j^* = \{1, 2\}$ , and for  $k = \{2, 3\}$  it is  $U_k^* = \{2, 3\}$ . Hence, the conditioning set of the edge  $e = \{\{1, 2\}, \{2, 3\}\}$  in  $T_2$  is  $D_e = U_j^* \cap U_k^* = \{1, 2\} \cap \{2, 3\} = \{2\}$ . The conditioned set consists of  $C_{e,j} = U_j^* \setminus D_e = \{1, 2\} \setminus \{2\} = \{1\}$  and  $C_{e,k} = U_k^* \setminus D_e = \{2, 3\} \setminus \{2\} = \{3\}$ . The edge of  $T_2$  between  $\{1, 2\}$  and  $\{2, 3\}$  in Figure 24.1 shows the elements of the conditioned sets  $\{1\}, \{3\}$  before “;” and conditioning set  $\{2\}$  after “;”.

**Definition 24.6 (*m*-child; *m*-descendent).** If node  $e$  is an element of node  $f$ , we say that  $e$  is an “*m*-child” of  $f$ ; similarly, if  $e$  is reachable from  $f$  via the membership relation:  $e \in e_1 \in \dots \in f$ , we say that  $e$  is an “*m*-descendent” of  $f$ .

The following properties of vines are proved in Bedford and Cooke (2002) and Kurowicka and Cooke (2003, 2006b).

**Proposition 24.7.** Let  $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$  be a regular vine. Then

1. the number of edges is

$$\frac{n(n-1)}{2} = \binom{n}{2}$$

2. each conditioned set is a doubleton, and each pair of variables occurs exactly once as a conditioned set,
3. if two edges have the same conditioning set, then they are the same edge,
4. for any node  $K$  of order  $k > 0$  in a regular vine,<sup>2</sup> if variable  $i$  is a member of the conditioned set of  $K$ , then  $i$  is a member of the conditioned set of exactly one of the  $m$ -children of  $K$ , and the conditioning set of an  $m$ -child of  $K$  is a subset of the conditioning set of  $K$ .

Word added – OK?

Change from ‘lemma’ in endnote OK?

**Remark 24.8.** When  $e \in E(\mathcal{V})$  for a regular vine  $\mathcal{V}$ , we let  $D_e$  denote the conditioning set associated with  $e$  and  $\{e_1, e_2\}$  be the conditioned set.

Word added – OK?

### Partial correlation vine

The edges of a regular vine may be associated with partial correlations, with values chosen arbitrarily in the interval  $(-1, 1)$  in the following way.

To each  $e \in E(\mathcal{V})$  we associate

$$\rho_{e_1, e_2; D_e}$$

where  $\rho_{e_1, e_2; D_e} = \rho_{e_1, e_2}$  if  $D_e$  is vacuous. The result is called a partial correlation vine.

Is this a technical term? Otherwise please clarify phrase.

Theorem 24.9 shows<sup>3</sup> that each such partial correlation vine specification uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way. In other words, a regular vine provides a bijective mapping from  $(-1, 1)^{\binom{n}{2}}$  into the set of positive definite matrices with “1”s on the diagonal.

Citation moved to endnote – OK? Theorem number from cited work needed for clarity?

**Theorem 24.9.** 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 other words this theorem says that all assignments of the numbers between  $-1$  and  $1$  to the edges of a partial correlation regular vine are consistent, and all correlation matrices can be obtained this way.

If the set of partial correlations on the D-vine in Figure 24.1 is

$$\begin{bmatrix} * & \rho_{12} & \rho_{13;2} & \rho_{14;23} \\ & * & \rho_{23} & \rho_{24;3} \\ & & * & \rho_{34} \\ & & & * \end{bmatrix} = \begin{bmatrix} * & 0.1 & 0.2 & 0.3 \\ & * & 0.4 & 0.5 \\ & & * & 0.6 \\ & & & * \end{bmatrix}$$

then the correlation matrix corresponding to this partial correlation specification can be calculated as

$$\begin{bmatrix} 1 & 0.1 & 0.2224 & 0.3409 \\ & 1 & 0.4 & 0.6066 \\ & & 1 & 0.6 \\ & & & 1 \end{bmatrix} \quad (24.2)$$

A regular vine may thus be seen as a way of picking out partial correlations which uniquely determine the correlation matrix and which are algebraically independent. The partial correlations in a partial correlation vine need not satisfy any algebraic constraint like positive definiteness. The “completion problem” for partial correlation vines is therefore trivial. An incomplete specification of a partial correlation vine may be extended to a complete specification by assigning arbitrary numbers in the  $(-1, 1)$  interval to the unspecified edges in the vine. Moreover, the following theorem shows how to find the extension with maximal determinant; we have only to assign zero partial correlation to unspecified edges in the vine.

**Theorem 24.10.** Let  $D$  be the determinant of the  $n$ -dimensional correlation matrix ( $D > 0$ ). For any partial correlation vine

$$D = \prod_{e \in E(\mathcal{V})} (1 - \rho_{e_1, e_2; D_e}^2) \quad (24.3)$$



**Vine copula**

Each edge in a regular vine may also be associated with a conditional copula, that is, a conditional bivariate distribution with uniform margins. It is convenient to specify the conditional bivariate copulas by first assigning a constant conditional rank correlation to each edge of the vine. For  $i = 1, \dots, n - 1$ , with  $e \in E_i$  and  $\{e_1, e_2\}$  the conditioned variables of  $e$ ,  $D_e$  the conditioning variables of  $e$ , we associate

$$r_{e_1, e_2 | D_e}$$

The resulting structure is called a “conditional rank correlation vine”. These rank correlations can be calculated from data (if available) or assessed by experts (Morales Napoles *et al* 2007).


Given a conditional rank correlation vine, we choose a class of copulas indexed by correlation coefficients in the interval  $[-1, 1]$  and select the copulas with correlation corresponding to the conditional rank correlation assigned to the edge of the vine. A joint distribution satisfying the vine-copula specification can be constructed and sampled on the fly, and will preserve maximum entropy properties of the conditional bivariate distributions (Bedford and Cooke 2001; Cooke 1997). Notice that different families of copula can be assigned to the edges of a vine.

**Theorem 24.11.** Let  $\mathcal{V} = (T_1, \dots, T_{n-1})$  be a regular vine on  $n$  elements. For each edge  $e \in E(\mathcal{V})$ , let the conditional copula and copula density be  $C_{e_1, e_2 | D_e}$  and  $c_{e_1, e_2 | D_e}$ . Let the marginal distributions  $F_i$  with densities  $f_i, i = 1, \dots, n$  be given. Then the vine-dependent distribution is uniquely determined and has a density given by

$$f_{1 \dots n} = f_1 \cdots f_n \prod_{e \in E(\mathcal{V})} c_{e_1, e_2 | D_e}(F_{e_1 | D_e}, F_{e_2 | D_e}) \quad (24.4)$$

Moreover, every positive density can be represented as a vine-dependent distribution.

**Theorem 24.12.** Let

- $f$  be a positive density of random vector  $(X_1, \dots, X_n)$  with marginal densities  $(f_1, \dots, f_n)$ ,
- $F = (F_1, \dots, F_n)$  be cumulative ity functions with densities  $(f_1, \dots, f_n)$  and
- $\mathcal{V} = (T_1, \dots, T_{n-1})$  be a regular vine on  $n$  elements,

Spelled out as the abbreviation is not used elsewhere - OK?

- $B = \{c_{jk|D_{e(j,k)}} \mid e(j,k) \in \bigcup_{i=1}^{n-1} E_i, \text{ where } e(j,k) \text{ is the unique edge with conditioned set } \{j,k\} \text{ and conditioning set } D_{e(j,k)}, \text{ and } c_{jk|D_{e(j,k)}} \text{ is a copula density of conditional distribution } f_{j,k|D_e} = f_{\{j,k,D_e\}} / f_{D_e}\}$

Is this all one set containing text, or are there unmatched braces missing around the first and last equations?

Word added – OK?

Then  $f$  is equal to density  $f_{j,k|D_e}$  by Equation 24.4, and hence the vine-dependent distribution.

The above two theorems show that flexible vines are in representing joint distributions.

#### Parametric inference for a given vine

Aas *et al* (2009) develop a maximum likelihood procedure to estimate parameters in copulas for D- and C-vines.

The parametric inference uses Theorem 24.11 to build likelihood function of the data. Starting values of the parameters needed in the numerical maximisation of the log-likelihood are determined sequentially as follows:

- estimate the parameters of the copulas in the first tree from the original data;
- evaluate the conditional distribution functions for the second tree at observed values using the copula parameters from the first tree and the conditional distributions of copulas assigned in the first tree;
- estimate the parameters of the copulas in the second tree using the observations from (b);
- repeat for trees  $T_3, \dots, T_{n-1}$ .

Parameters estimated in the sequential procedure are taken as the starting point for the optimisation of all parameters together.

In the above procedure it is assumed that conditional copulas do not depend on the conditioning variables. Moreover, in practice only a few copula families are considered as possible building blocks of the vine. The assumption of constant conditional copulas and considering only a few types of bivariate families in fitting a vine to data cause the phenomenon that some types of vines can fit the data better than others. To find the best vine structure we would in principle have to estimate all possible vines. In dimensions higher than seven or eight this is infeasible, as the number of

'consideration of?'

Changes to sentence OK?

vines grows rapidly with dimension (Morales Napoles 2010). Moreover, because of sequential estimation, estimates for parameters of conditional copulas in higher-order trees are less reliable. For higher-dimensional cases some simplifying assumptions for fitting vines to data will have to be made.

**Sampling**

The joint copula specified by a vine can be sampled. The cumulative strategy to sample such a distribution will be illustrated by the distribution specified by the D-vine in Figure 24.1,  $D(1, 2, 3, 4)$ : sample four independent variables distributed uniformly on the interval  $[0, 1]$ ,  $U_1, U_2, U_3, U_4$  and calculate values of correlated variables  $X_1, X_2, X_3, X_4$  as follows.

Word added – OK?

Change OK?

1.  $x_1 = u_1$ .
2.  $x_2 = F_{r_{12};x_1}^{-1}(u_2)$ .
3.  $x_3 = F_{r_{23};x_2}^{-1}(F_{r_{13}|2;F_{r_{12};x_1}}^{-1}(u_3))$ .
4.  $x_4 = F_{r_{34};x_3}^{-1}(F_{r_{24}|3;F_{r_{23};x_2}}^{-1}(x_2)(F_{r_{14}|23;F_{r_{13}|2;F_{r_{23};x_2}}^{-1}(x_3)(F_{r_{12};x_1}}^{-1}(u_4)))$ .

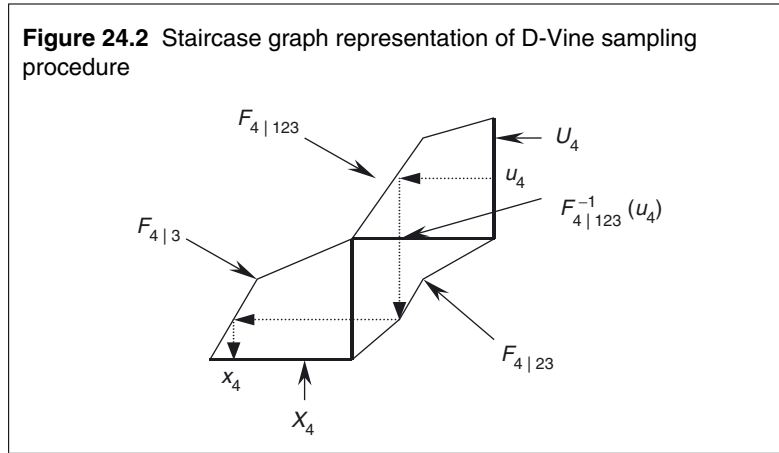
$F_{r_{ij|k};X_i}(X_j)$  denotes the cumulative distribution function for  $X_j$ , applied to  $X_j$ , given  $X_i$  under the conditional copula with rank correlation  $r_{ij|k}$ . Notice that the D-vine sampling procedure uses conditional and inverse conditional distribution functions. A more general form of the above procedure simply refers to conditional cumulative distribution functions

$$\left. \begin{aligned} x_1 &= u_1 \\ x_2 &= F_{2|1;x_1}^{-1}(u_2) \\ x_3 &= F_{3|2;x_2}^{-1}(F_{3|12;F_{1|2}(x_1)}^{-1}(u_3)) \\ x_4 &= F_{4|3;x_3}^{-1}(F_{4|23;F_{2|3}(x_2)}^{-1}(F_{4|123;F_{1|23}(x_1)}^{-1}(u_4))) \end{aligned} \right\} \quad (24.5)$$

Figure 24.2 depicts the sampling of  $X_4$  in the D-vine in Figure 24.1 with a “staircase graph”. Following the dotted arrows, we start by sampling  $U_4$  (realisation  $u_4$ ) and use this with the copula for the conditional rank correlation of  $\{1, 4\}$  given  $\{2, 3\}$  to find the argument of  $F_{4|23}^{-1}$ , etc. Notice that, for the D-vine, values of  $F_{2|3}$  and  $F_{1|23}$  that are used to conditionalise copulas with correlations  $r_{24|3}$  and  $r_{14|23}$  to obtain  $F_{4|23}$  and  $F_{4|123}$ , respectively, have to be calculated.

The staircase graph shows that if any of the cumulative conditional distributions in Figure 24.2 are uniform, then the corresponding abscissa and ordinates can be identified. This corresponds to

**Figure 24.2** Staircase graph representation of D-Vine sampling procedure



noting that the inverse cumulative function in Equation 24.5 is the identity, and this in turn corresponds to a conditional rank correlation being zero and the corresponding variables being conditionally independent. Notice that the conditional rank correlations can be chosen arbitrarily in the interval  $[-1, 1]$ ; they need not satisfy any algebraic constraint.

### CONTINUOUS NON-PARAMETRIC BAYESIAN BELIEF NETS

Bayesian belief nets (BBNs) are directed acyclic graphs that together with conditional probability functions represent high-dimensional uncertainty distributions (Cowell *et al* 1999; Jensen 1996, 2001; Pearl 1988). The nodes represent variables, which can be discrete or continuous, and the arcs represent causal/influential or functional relationships between variables.

Continuous BBNs (Pearl 1988; Shachter and Kenley 1989) 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. They require means, conditional variances and partial regression coefficients which can be specified in an algebraically independent manner. The restriction to joint normal is severe. We cannot simply invoke the theory of linear least squares predictors as applied to arbitrary joint distributions. Suppose  $(X_1, \dots, X_{k-1})$  are the ancestors of  $X_k$  in an ordered BBN.<sup>4</sup> We could interpret the “influence” of  $X_j$  on  $X_k$  as the partial regression of  $X_k$  on  $X_j$  given  $1, \dots, j-1, j+1, \dots, k-1$ . If  $j$  is not a parent of

$k$ , then  $j$  and  $k$  are conditionally independent given the parents of  $k$ ; however, it is not generally the case that the partial regression of  $k$  on  $j$ , given the parents, is necessarily zero (Kurowicka and Cooke 2000). This means that the partial regression coefficients for distributions other than the normal distribution do not reflect the conditional independence structure of the BBN.

Change OK?

Changes to sentence OK?

Kurowicka and Cooke (2006a) advanced a vine-based distribution-free approach to continuous BBNs. Starting with an arbitrary BBN whose nodes have continuous invertible distributions, they associated each arc with a (conditional) parent-child rank correlation according to a protocol presented below. They specified nested sets of high-dimensional joint distributions using the vine-copula approach, where any copula with invertible conditional cumulative distribution functions may be used so long as the chosen copula represents (conditional) independence as zero (conditional) correlation. The conditional rank correlations (like the partial regression coefficients) are algebraically independent, and there are tested protocols for their use in structured expert judgement (Goossens *et al* 1997; Kraan 2002). The vine generates a sampling algorithm which satisfies the conditional independence relations implied by the BBN.

Change to past tense in this paragraph to match the first sentence relating to the citation – OK?

The vine-based approach is quite general, and of course this comes at a price: these BBNs must be evaluated by Monte Carlo simulation. However, if the joint normal copula is used, then updating and conditionalising can be done analytically (Hanea *et al* 2006; Kurowicka and Cooke 2004).

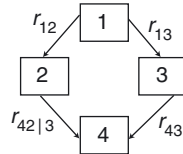
Publication details available for Kurowicka and Cooke (2004)?

We associate nodes of a BBN with univariate random variables  $\{1, \dots, n\}$  having uniform distributions on  $(0, 1)$ . The protocol to associate the arcs, or “influences”, with (conditional) rank correlations is as follows.

1. Construct a sampling order for the nodes, that is, an ordering such that all ancestors of node  $i$  appear before  $i$  in the ordering. A sampling order begins with a source node and ends with a sink node. Of course, the sampling order is not, in general, unique. Index the nodes according to the sampling order  $1, \dots, n$ .
2. Factorise the joint in the standard way following the sampling order. With sampling order  $1, 2, \dots, n$ , write

$$P(1, \dots, n) = P(1)P(2 | 1)P(3 | 21) \dots P(n | n-1, n-2, \dots, 1)$$

**Figure 24.3** A simple BBN on four variables with assigned to arcs (conditional) rank correlations



Change OK?

3. Underscore those nodes in each condition that are not parents of the conditioned variable and thus are not necessary in sampling the conditioned variable. This uses some of the conditional independence relations in the belief net. Hence, if in sampling 2, . . . , n variable 1 is not necessary (ie, there is no influence from 1 to any other variable) then

$$P(1, \dots, n) = P(1)P(2 \mid \underline{1})P(3 \mid \underline{21}) \dots P(n \mid n-1, n-2, \dots, \underline{1}). \quad (24.6)$$

The underscored nodes could be omitted thereby yielding the familiar factorisation of the BBN as a product of conditional probabilities, with each node conditionalised on its parents (for source nodes the set of parents is empty).

I have assumed that the equation number refers to both equations. OK?

4. For each node  $i$  with parents (non-underscored variables)  $i_1, \dots, i_{p(i)}$  in Equation 24.6, associate the arc  $i_{p(i)-k} \rightarrow i$  with the conditional rank correlation

$$\left. \begin{aligned} &r(i, i_{p(i)}) \quad \text{if } k = 0 \\ &r(i, i_{p(i)-k} \mid i_{p(i)}, \dots, i_{p(i)-k+1}) \quad \text{if } 1 \leq k \leq p(i) - 1 \end{aligned} \right\} \quad (24.7)$$

Commas needed between terms here?

where the assignment is vacuous if  $\{i_1 \dots i_{p(i)}\} = \emptyset$ , hence if this node does not have any parents. This way every arc in the BBN is associated with a conditional rank correlation between parent and child. These correlations can be found from data or assessed by experts.

In Figure 24.3 a simple BBN with assigned to arcs (conditional) rank correlations is shown.

Changes to sentence OK?

According to the protocol above we have associated arcs pointing to the variable 4 with two rank correlations, one unconditional ( $r_{43}$ ) and the other conditional ( $r_{42|3}$ ).

The following theorem (Hanea *et al* 2006; Kurowicka and Cooke 2006a) shows that assignments of (conditional) rank correlations in

the algorithm above uniquely determine the joint distribution and are algebraically independent.

**Theorem 24.13.** Given a BBN with  $n$  nodes and continuous invertible univariate margins, the specification of conditional rank correlations (Equation 24.7),  $i = 1, \dots, n$  and a copula realising all correlations  $[-1, 1]$  for which correlation 0 entails independence uniquely determines the joint distribution. This joint distribution satisfies the characteristic factorisation (Equation 24.6) and the conditional rank correlations in Equation 24.7 are algebraically independent.

### Sampling

Assuming that variables in the BBN on Figure 24.3 are uniform, the sampling procedure for the joint distribution represented by this BBN is

1.  $x_1 = u_1,$
2.  $x_2 = F_{r_{12};x_1}^{-1}(u_2),$
3.  $x_3 = F_{r_{23};x_2}^{-1}(u_3),$
4.  $x_4 = F_{r_{34};x_3}^{-1}(F_{r_{24};F_{r_{23};x_3}(x_2)}^{-1}(u_4)).$

Notice that  $F_{r_{23};x_3}(x_2)$  is in general not known analytically and has to be calculated from the joint distribution of variables 1, 2 and 3 by integrating out the variable 1. Numerical integration on every sample is time consuming. This problem disappears if the normal copula is used to realise all conditional rank correlations. These computational difficulties when sampling BBNs restrict the choice of copula to the normal copula. The choice of normal copula also allows analytical conditioning, which is possibly the biggest advantage of this model.

Change OK?

### Partial correlations on BBN

Instead of (conditional) rank correlations, partial correlations can be assigned to arcs of a BBN following the same algorithm as for rank correlations. Then a factorisation of the determinant similar to that for regular vines holds (Hanea *et al* 2010).

**Theorem 24.14.** For any partial correlation BBN specification

$$D = \prod (1 - \rho_{ij;D_{ij}}^2)$$

where  $\rho_{ij;D_{ij}}$  is the partial correlation associated with the arc between node  $i$  and node  $j$ ,  $D_{ij}$  is the conditioning set for the arc between node  $i$  and node  $j$ , and the product is taken over all arcs in the BBN.

Change OK?

'by'?

Continuous non-parametric BBNs are in some sense very similar to vines. They both specify the dependence structure of a joint distribution via specification of the joint copula. The joint copula is described with (conditional) copulas assigned to the edges or arcs. Kurowicka and Joe (2010) compared vines and BBNs. The most important difference between BBNs and vines appears to be the difference in conditional independence statements that they can represent.

### MODEL LEARNING

Words added – OK?

An approach to model learning inspired by Whittaker (1990) was developed in Kurowicka and Cooke (2006a) for vines, based on the factorisation of the determinant in Theorem 24.10. The same approach can be followed in order to learn a BBN model and is based on the factorisation of the determinant in Theorem 24.14. We propose a strategy to choose a regular vine or a BBN structure which captures the mutual information in a small number of conditional bivariate terms. We will approximate the mutual information by  $-\frac{1}{2} \ln(D_r)$ , where  $D_r$  is the determinant of the rank correlation matrix.

### BBNs

Changes to sentence OK?

Since BBNs are currently used only with normal copula then the first step of BBN model building consists of validating the joint normal copula. If this step is successful we may learn a graph structure by removing arcs associated with small conditional rank correlation (which in case of normal copula correspond to small partial correlations). This procedure is detailed in Hanea *et al* (2010) and it uses the following three determinants.

**DER:** the determinant of the empirical rank correlation matrix.

**DNR:** the determinant of the rank correlation matrix obtained by transforming the marginals to standard normals, and then transforming the product moment correlations to rank correlations using Pearson's transformation

$$r = \frac{6}{\pi} \arcsin\left(\frac{1}{2}\rho\right)$$



where  $r$  and  $\rho$  denote rank and Pearson's correlations for the joint normal, respectively.<sup>5</sup> DNR will generally differ from DER because DNR assumes the normal copula. A statistical test for the suitability of DNR for representing DER uses the sampling distribution of DNR and checks whether DER is within the 90% central confidence band of DNR.

**DBBN:** the determinant of the rank correlation matrix for normal copula with correlation structure of the BBN.

The procedure of constructing a BBN is as follows.

- (a) Construct a skeletal BBN by adding arcs to capture known causal or temporal relations.
- (b) If DNR is within the 90% central confidence band of the determinant of the skeletal BBN (DBBN), then stop; otherwise continue with step (c).
- (c) Find the pair of variables such that the arc between them is not in the graph and their rank correlation is greater than the rank correlation of any other pair not in the BBN.
- (d) Add an arc between them and recompute DBBN together with its 90% central confidence band. If DNR is within the 90% central confidence band of DBBN, then stop; otherwise, continue adding arcs.

Changes to sentence  
OK?

### Vines

Strategies for choosing a regular vine structure which captures the mutual information in a small number of conditional bivariate terms will be described.

#### Majorisation

We now write Theorem 24.10 as

$$\text{MI}(f) \approx -\frac{1}{2} \ln(D_r) = \sum_{e \in E(\mathcal{V})} b_{e_1, e_2; D(e)} \quad (24.8)$$

where  $D(e)$  is the conditioning set for the node in  $\mathcal{V}$  with conditioned set  $\{e_1, e_2\}$ ,  $b_{e_1, e_2; D(e)} = -\frac{1}{2} \ln(1 - r_{e_1, e_2; D(e)}^2)$  and  $r_{e_1, e_2; D(e)}$  denotes partial rank correlation. The terms  $b_{e_1, e_2; K(e)}$  will depend on the regular vine which we choose to represent the second-order rank structure. However, the sum of these terms must satisfy Equation 24.8.

We seek a regular vine for which the terms  $b_{e_1, e_2; K(e)}$  in Equation 24.8 are as “spread out” as possible. In other words, we wish to capture the total dependence  $MI(f)$  in a small number of terms, with the remaining terms being close to zero. This concept is made precise with the notion of majorisation (Marshall and Olkin 1979).

**Definition 24.15.** Let  $x, y \in \mathbb{R}^n$  be such that

$$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

Then  $x$  majorises  $y$  if for all  $k, k = 1, \dots, n$

$$\sum_{j=1}^k x_{(j)} \leq \sum_{j=1}^k y_{(j)} \tag{24.9}$$

where  $x_{(j)}$  is the increasing arrangement of the components of  $x$ , and similarly for  $y$ .

In view of Equation 24.8, the model inference problem may be cast as the problem of finding a regular vine whose terms  $b_{e_1, e_2; K(e)}$  are non-dominated in the sense of majorisation. In that case, setting the smallest mutual informations equal to zero will change the overall mutual information as little as possible. Pairs of variables whose (conditional) mutual information is zero, are (conditionally) independent. Finding non-dominated solutions may be difficult, but a necessary condition for non-dominance can be found by maximising any Schur convex function.

‘data?’ amounts of information’, ‘information values?’ (Information is usually a plural noun.)

**Definition 24.16.** A function  $f: \mathbb{R}^k \rightarrow \mathbb{R}$  is Schur convex if  $f(x) \geq f(y)$  whenever  $x$  majorises  $y$ .

Schur convex functions have been studied extensively. A sufficient condition for Schur convexity is given by Marshall and Olkin (1979).

**Proposition 24.17.** If  $f: \mathbb{R}^k \rightarrow \mathbb{R}$  may be written as  $f(x) = \sum f_i(x_i)$  with  $f_i$  convex, then  $f$  is Schur convex.

**Vine inference strategy.** The following strategy for model inference suggests itself:

1. choose a Schur convex function  $f: \mathbb{R}^{(n(n-1))/2} \rightarrow \mathbb{R}$ ;
2. find a regular vine  $\mathcal{V}$  whose vector  $b_{e_1, e_2; D(e)}$  maximises  $f$ ;
3. set the mutual informations in  $\mathcal{V}$  equal to zero for which the terms  $b_{e_1, e_2; D(e)}$  are smallest;

‘data points’, ‘information values?’

4. associate copulas with the nodes in the vine, such that the non-zero mutual information values are preserved.

In our application to wind data we use as  $f$  the negative entropy function and inspect vectors of  $b_{e_1, e_2; D(e)}$  for all vines on seven variables (more than 2 million). This procedure will not be possible for high-dimensional problems.

Two greedy strategies described below do not require the generation of all possible vines; instead we generate the “best vine” sequentially to satisfy required constraints.

Changes to sentence OK?

#### *Most dependence in lowest order trees*

A regular vine on  $n$  variables can be generated in different ways. One way is to follow its definition: choose the first tree; for  $j = 2, \dots, n-1$  build  $T_j$  by connecting two edges in  $T_{j-1}$  if they share a common node. This way of generating a vine can be used for the following strategy.

1. Choose the first tree with maximum sum of absolute values of rank correlations assigned to it. This can be achieved with maximum spanning tree algorithm.<sup>6</sup>
2. Check which edges can be connected in the second tree. (Construct a line graph for the first tree by connecting by an edge all edges that share a common node in the first tree. A line graph is usually not a tree.)
3. Calculate partial correlations on the line graph in step 2.
4. Choose a tree with maximum absolute values of partial correlations on the line graph calculated in step 3.
5. Repeat for trees  $T_3, \dots, T_{n-1}$ .

This strategy ensures that the first tree captures the most dependence, but higher trees may not be optimal due to the regularity condition that has to be satisfied when generating a vine. The first tree accounts for only  $n-1$  out of  $\binom{n}{2}$  correlations in the correlation matrix. Hence, fixing the lowest tree first may not lead to the optimal strategy. On the other hand, estimates of lower-order partial correlations are less affected by estimation error.

Change OK?

*Least dependence in highest order trees*

We can also start building the vine from the top edge (edge in tree  $n - 1$ ) and progress to the lower trees, ensuring that the regularity condition is satisfied.<sup>7</sup> This way of generating a vine can be used for the following strategy.

1. Choose the edge in tree  $n - 1$  with minimum absolute value of partial correlation given all remaining variables (these partial correlations can be found in the normalised inverse of the correlation matrix). If we have chosen  $n$  and  $n - 1$  then the constraint set of this edge is  $\{n, n - 1, 1, \dots, n - 2\}$ . Constraint sets of its two  $m$ -children are  $\{n, 1, \dots, n - 2\}$  and  $\{n - 1, 1, \dots, n - 2\}$ .
2. We must choose variables for the conditioned sets with variables  $n$  and  $n - 1$  in  $T_{n-2}$  such that regularity will be satisfied. We call them partners of  $n$  and  $n - 1$  and denote them as  $pt(n)$  and  $pt(n - 1)$ , respectively. The partners must be chosen such that absolute values of partial correlations

$$\rho_{n,pt(n);\{1,\dots,n-2\}\setminus pt(n)} \quad \text{and} \quad \rho_{n-1,pt(n-1);\{1,\dots,n-2\}\setminus pt(n-1)}$$

are minimal.

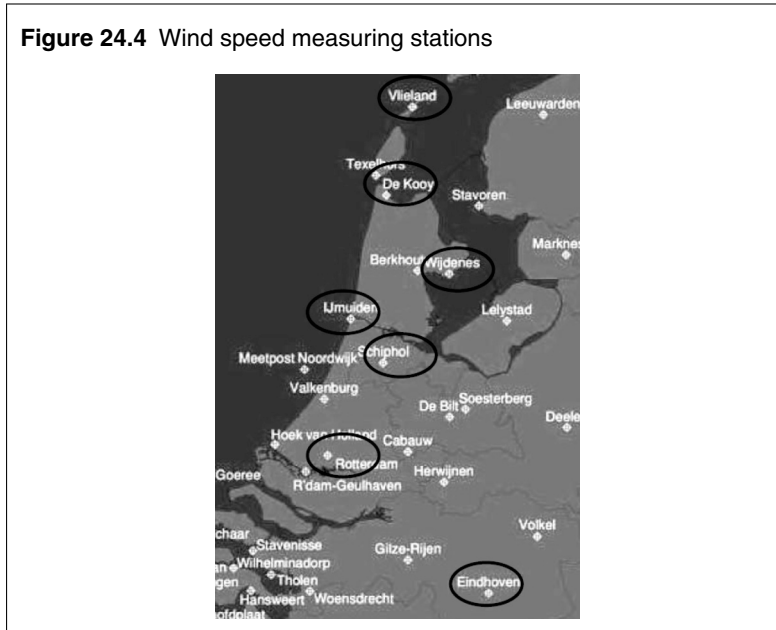
3. Repeat for trees  $T_{n-2}, \dots, T_1$ .

*Truncated model*

So far we have chosen a vine structure with few different heuristics. The procedure of testing which copulas on the vine can be set to the independent copula is as follows:

- (a) choose a copula assigned to the edges of the vine with the smallest positive absolute value of the correlation;
- (b) set it to the independent copula and refit the model;
- (c) sample the copula obtained in (b) and calculate the determinant of the rank correlation matrix;
- (d) repeat (c) and construct 90% central confidence band of this determinant;
- (e) if the empirical rank determinant is outside the 90% central confidence band of vine rank determinant, then reset this copula and stop; otherwise go to step (a).

We can additionally require that first copulas that are set to the independent copula are in trees with the most conditioning (higher-order trees).

**Figure 24.4** Wind speed measuring stations**MODEL LEARNING: RESULTS FOR WIND SPEED DATA**

To illustrate these ideas, we use two years of hourly wind speed data at seven locations in the Netherlands (indicated in Figure 24.4), provided by the Royal Netherlands Meteorological Institute (KNMI). Predicting wind speeds on this timescale is of interest for wind farms entering the spot electricity markets. Our goal is to model the dependence structure of wind speed at these locations. At low wind speeds, measurements are jittery and values are binned to reduce jitter, causing granularity in the data.

The Spearman rank correlation matrix of this data given in Table 24.1 indicates quite high correlations between wind speed in all stations. The determinant of this matrix DER is equal to  $2.9917 \times 10^{-4}$ .

Figure 24.5 shows a rank scatter plot for variables Schip and Ijmu. We can observe the granular nature of this data specially for low wind speed values. Correlations at high wind speeds are stronger than at low wind speeds, indicating upper tail dependence in wind speeds at different locations.

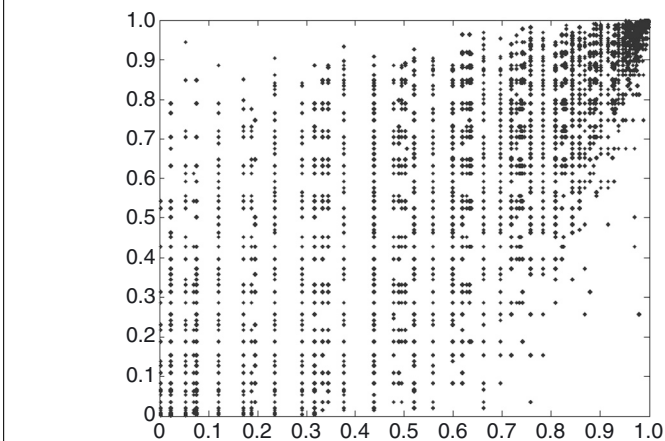
We select unconditional bivariate copulas for each pair of variables by choosing from the following four families of bivariate copulas: Normal (N), Frank's (F), Student's  $t$  (T) and Gumbel (G). The

Matrix changed to table environment to fit page size - OK?

**Table 24.1** Hourly wind speed data for the seven locations in the Netherlands given in Figure 24.4

	Rott	Eind	Schip	Ijmu	Deko	Vlie	Wijd
Rott	1.0000	0.8030	0.8566	0.8050	0.7791	0.7094	0.7680
Eind	0.8030	1.0000	0.7991	0.6901	0.7046	0.6147	0.6870
Schip	0.8566	0.7991	1.0000	0.8233	0.8342	0.7308	0.8054
Ijmu	0.8050	0.6901	0.8233	1.0000	0.8410	0.8211	0.7679
Deko	0.7791	0.7046	0.8342	0.8410	1.0000	0.8764	0.8054
Vlie	0.7094	0.6147	0.7308	0.8211	0.8764	1.0000	0.7458
Wijd	0.7680	0.6870	0.8054	0.7679	0.8054	0.7458	1.0000

**Figure 24.5** Scatter plot of ranks for variables Schip and Ijmu



Gumbel copula is the best for almost all bivariate margins. This is not surprising in light of the upper tail dependence noted in Figure 24.5. We see the results for each bivariate margin in the matrices below

*	2.6863	3.2433	2.6878	2.5029	2.1452	2.4269
	*	2.6915	2.0636	2.1223	1.8148	2.0504
		*	2.8545	2.9610	2.2483	0.8327(17.45)
			*	2.9903	2.8217	2.4110
				*	3.4462	0.8297(17.85)
					*	0.7748(17.20)
						*

$$\begin{bmatrix} * & G & G & G & G & G & G \\ & * & G & G & G & G & G \\ & & * & G & G & G & T \\ & & & * & G & G & G \\ & & & & * & G & T \\ & & & & & * & T \\ & & & & & & * \end{bmatrix}$$

The first matrix shows parameters of copulas estimated form the data. In the second matrix the type of copula that gives the maximum likelihood is presented. Since the Student *t*-copula has two parameters, the degrees of freedom parameter are printed in the first matrix in parentheses. We see that the best copula for Schip and Ijmu (variables 3 and 4 in the rank correlation matrix) is Gumbel with parameter 2.8545.

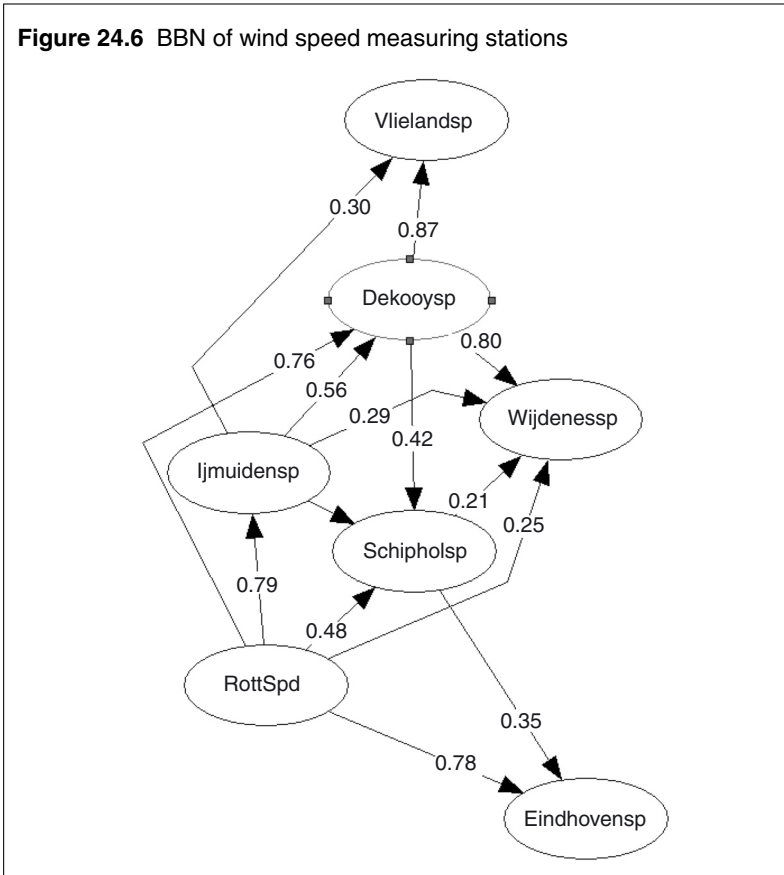
We will first try to model these data with non-parametric continuous BBNs and then test different strategies of choosing the best vine structure.

### Non-parametric continuous BBNs

All calculations in this subsection are done with software UNINET. We first test the assumption of normal copula using the re-sampling technique presented in the previous section (see page 624). The 90% confidence bound of sampling distribution for DNR was  $[4.4444 \times 10^{-4}, 5.2782 \times 10^{-4}]$ . The DER is  $2.9917 \times 10^{-4}$  and falls outside of this interval, meaning that normal copula does not adequately represent dependence of this data set. This is not surprising, due to the upper tail dependence in this data. The fully connected BBN with normal copula leads to the DNR equal to  $4.8206 \times 10^{-4}$ .

In the further analysis we will neglect the differences in DNR and DER, and try to build a BBN structure that will capture DNR in a small number of conditional bivariate terms. For this purpose, the algorithm presented in the previous section (see page 624) is used. We have also included knowledge of the prevailing wind direction in the Netherlands for directionality of arcs in the BBN. Figure 24.6 shows the BBN with 14 arcs and with (conditional) rank correlations assigned to these arcs. Saving 90% confidence bounds for DBBN are  $[4.544 \times 10^{-4}, 5.4557 \times 10^{-4}]$  which contain DNR. Hence, the dependence of DNR is captured using 14 out of 21 correlations.

**Figure 24.6** BBN of wind speed measuring stations



The log-likelihood for the data with normal copula is equal to 70,461.

**Vines**

*Majorisation*

We use as a Schur convex function the negative entropy function. Hence, we calculate

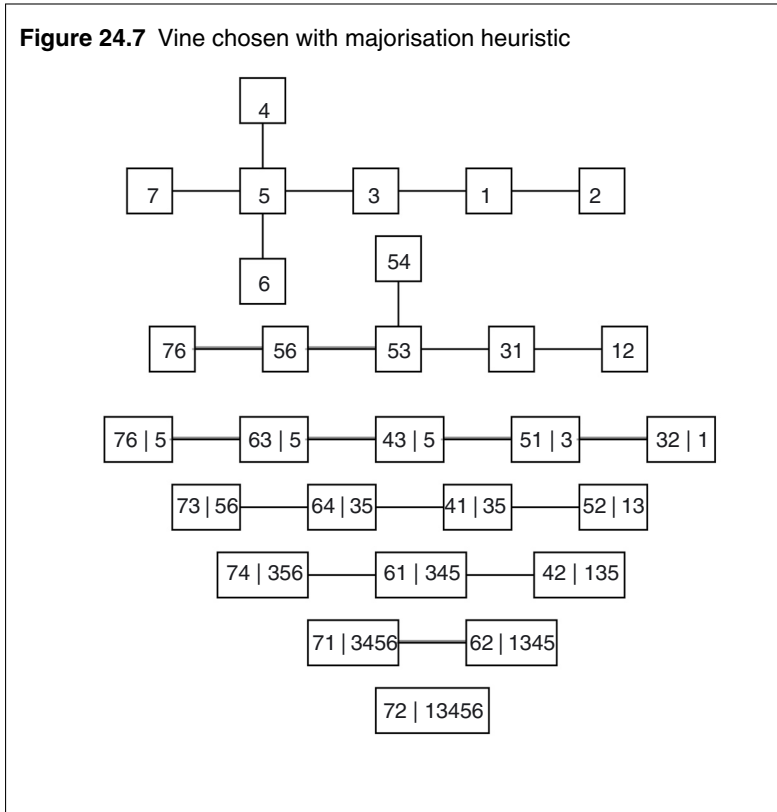
$$\text{Entropy}(\mathcal{V}) = \sum_{e \in E(\mathcal{V})} b_{e_1, e_2; D_e} \ln(b_{e_1, e_2; D_e})$$

for all 2,580,480 possible vines and choose one for which entropy is the smallest. To generate all possible vines, the algorithm presented in Kurowicka and Joe (2010) was used. The “best vine” according to the majorisation strategy ( $\mathcal{V}_M$ ) is shown in Figure 24.7. To show results of inference for different types of vines we can use the matrix

I have assumed that the subscript ‘M’ is a label for ‘majorisation’ and set it upright according to book label style. OK?



Figure 24.7 Vine chosen with majorisation heuristic



representation of a regular vine which contains all information about conditioning and conditioned sets on the vine. The conditioning and conditioned sets of edges of the vine in matrix  $A$  are coded as follows:

$$\{A(i,j), A(j,j) \mid A(1,j) \cdots A(i-1,j)\}, \quad i = 1, \dots, n, \quad j = i+1, \dots, n$$

The matrix representation  $\mathcal{V}_M$  can be compared with the conditioned and conditioning sets of edges in each tree shown in Figure 24.7

$$\mathcal{V}_M = \begin{bmatrix} 5 & 5 & 5 & 3 & 5 & 1 & 5 \\ & 3 & 3 & 5 & 3 & 3 & 6 \\ & & 4 & 4 & 4 & 5 & 3 \\ & & & 1 & 1 & 4 & 4 \\ & & & & 6 & 6 & 1 \\ & & & & & 2 & 2 \\ & & & & & & 7 \end{bmatrix}$$



**Table 24.2** 5th, 50th and 95th percentile of the distribution of the rank determinant for majorised vine with parameters obtained sequentially  $\mathcal{V}_M^{SF}$  and optimised together  $\mathcal{V}_M^{MLL}$

Vine/rank	LL	5%	50%	95%
$\mathcal{V}_M^{SF}$	74,350	$2.8366 \times 10^{-4}$	$3.1906 \times 10^{-4}$	$3.5078 \times 10^{-4}$
$\mathcal{V}_M^{MLL}$	74,543	$4.5528 \times 10^{-4}$	$5.0320 \times 10^{-4}$	$5.5357 \times 10^{-4}$

param<sub>M</sub><sup>MLL</sup>

*	2.6783	2.8101	2.9224
	*	0.3995(10.80)	0.2200
		*	0.2829(19.55)
=			*
		3.2192	2.4327
		0.0209(19.70)	0.3767(12.78)
		1.2766	0.0217
		-0.0037	-0.0785
		*	-0.3893
			*
			0.9527
			0.0386
			*
			0.8116(17.36)
			0.1446
			0.3955
			0.1166(18.36)

As with the BBN model, we can use the rank determinant test to see if the dependence structure of the data is represented well by the vine copula obtained with sequential fitting procedure or with fully optimised parameters. We find the sampling distribution of the rank determinant for majorised vine with parameters obtained sequentially and optimised together. The 5th, 50th and 95th percentile of these distributions are presented in Table 24.2.

The empirical rank determinant is contained in the 90% confidence bound of  $\mathcal{V}_M^{SF}$  and is strongly outside 90% bounds  $\mathcal{V}_M^{MLL}$ . In the penultimate section (see page 640) we discuss briefly the issue of sequential fitting.

Due to computational issues we have followed a somewhat different heuristic than that presented in the previous section (see page 624) to select  $\alpha$  as should be set to the independent cop-

Changes to sentence OK?

Table 24.3 is not cited in the text. Please suggest where this should be mentioned. Also variable 'e' OK in that table?

Change OK?

Change OK?

Clarify sentence?

ula, basing the decision on values of partial rank correlations calculated for the vine. The partial rank correlation matrix  $PR_M$  for the vine coded in the matrix  $\mathcal{V}_M$  is equal to

Change to text and equation layout to fit page size – OK?

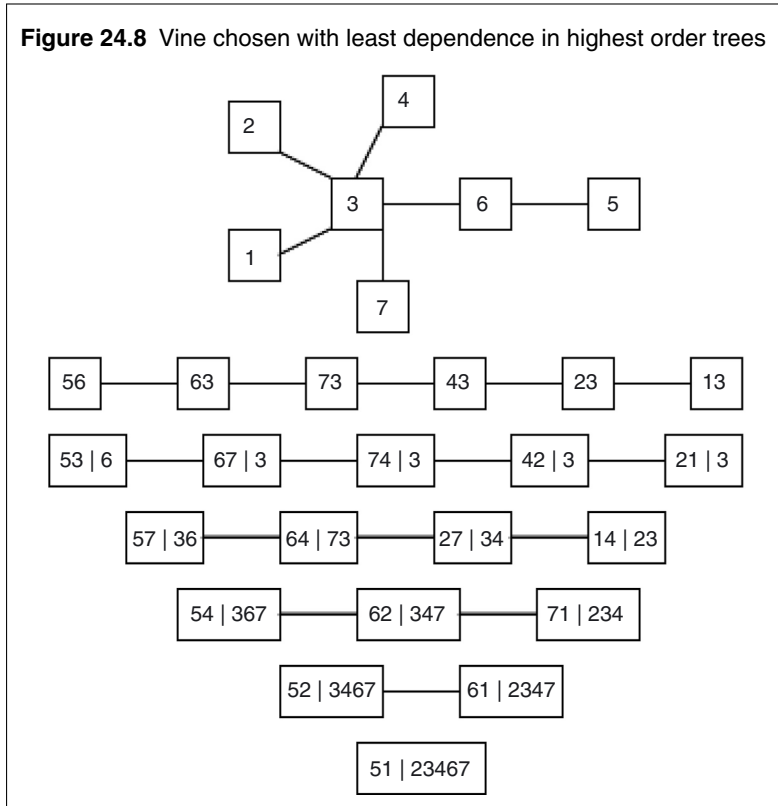
$$\begin{bmatrix} * & 0.8342 & 0.8410 & 0.8566 & 0.8764 & 0.8030 & 0.8054 \\ & * & 0.4078 & 0.2270 & -0.0010 & 0.3618 & 0.1397 \\ & & * & 0.2700 & 0.3538 & 0.0311 & 0.4127 \\ & & & * & 0.0166 & -0.0618 & 0.0912 \\ & & & & * & -0.0350 & 0.1518 \\ & & & & & * & 0.0269 \\ & & & & & & * \end{bmatrix}$$

Changes to sentence OK?

We see seven partial correlations smaller than 0.1 in absolute value. Assigning the independent copula to edges with small partial rank correlations, we found the set of parameters of the remaining (non-independent) copulas that maximise log-likelihood of the data. The estimates obtained  $param_M^*$  and  $types_M^*$  (where  $I$  denotes the independent copula for which the parameter is 0) can be compared with  $param_M^{MLL}$  and  $types_M^{MLL}$ . The log-likelihood, as expected, become smaller (74,269) and the 5th, 50th and 95th percentile of the rank determinant distribution become  $[4.7461 \times 10^{-4}, 5.323 \times 10^{-4}, 5.8028 \times 10^{-4}]$

$$param_M^* = \begin{bmatrix} * & 2.6671 & 2.8074 & 2.9293 \\ & * & 0.3970(11.29) & 0.2212 \\ & & * & 0.2634(19.53) \\ & & & * \\ & & & & * \\ & 3.2051 & 2.4347 & 0.8115(16.88) \\ & 0 & 0.3781(12.69) & 0.1364 \\ & 1.2798 & 0 & 0.3974 \\ & 0 & 0 & 0 \\ & * & 0 & 0.9593 \\ & & * & 0 \\ & & & * \end{bmatrix}$$

**Figure 24.8** Vine chosen with least dependence in highest order trees



$$\text{types}_M^* = \begin{bmatrix} * & G & G & G & G & G & T \\ & * & T & N & I & T & N \\ & & * & T & G & I & N \\ & & & * & I & I & I \\ & & & & * & I & F \\ & & & & & * & I \\ & & & & & & * \end{bmatrix}$$

*Most dependence in first trees*

Interestingly the “most dependence in lowest trees” heuristic leads to the same vine as the majorisation heuristic in the preceding section. This naturally does not have to be the case in general.

*Least dependence in highest order trees*

Following the algorithm presented earlier (see page 628) we obtain vine shown in Figure 24.8. It can be coded in the matrix form

Change OK?

denoted as

$$\mathcal{V}_T = \begin{bmatrix} 3 & 3 & 3 & 3 & 3 & 3 & 6 \\ & 4 & 4 & 4 & 7 & 2 & 3 \\ & & 7 & 7 & 4 & 4 & 7 \\ & & & 2 & 2 & 7 & 4 \\ & & & & 6 & 6 & 2 \\ & & & & & 1 & 1 \\ & & & & & & 5 \end{bmatrix}$$

$\mathcal{V}_T$  is neither a D-vine nor a C-vine. We started with the sequential fitting procedure, which led to the choice of types of the copulas assigned to the edges  $\mathcal{V}_T$  and the initial values of parameters of these copulas. The log-likelihood for the data with  $\mathcal{V}_T$  fitted sequentially was equal to 74,446 and the 90% interval of rank determinant was  $[3.0285 \times 10^{-4}, 3.6775 \times 10^{-4}]$ . Then, the parameters of copulas were optimised together and the following parameter values (types<sup>MLL</sup>) (param<sup>MLL</sup><sub>T</sub>) were found to maximise likelihood (equal to 74,582) for the data. The rank determinant distribution for this vine with parameters optimised together had the following 5th, 50th and 95th quantiles  $[4.6830 \times 10^{-4}, 5.1676 \times 10^{-4}, 5.6454 \times 10^{-4}]$

$$\text{param}_T^{\text{MLL}} = \begin{bmatrix} * & 2.6054 & 0.8067(17.25) & 2.4450 \\ & * & 0.3300 & 0.0760(19.66) \\ & & * & 0.0973 \\ & & & * \\ & 2.1352 & 2.9048 & 3.1961 \\ & 0.3920 & 0.3558(12.79) & 0.5761(13.76) \\ & 1.4863 & 0.3497(17.90) & 0.2185 \\ & -0.0220 & 1.0991 & 0.1490(18.50) \\ & * & -0.0085 & 0.4399 \\ & & * & -0.0125 \\ & & & * \end{bmatrix}$$



**Table 24.3** 5th, 50th and 95th percentile of the distribution of the rank determinant for majorised vine with normal copulas with parameters obtained sequentially  $\mathcal{V}_M^{\text{NSF}}$  and optimised together  $\mathcal{V}_M^{\text{NMLL}}$

Vine	LL	Det
$\mathcal{V}_M^{\text{NSF}}$	70,223	$1.9199 \times 10^{-4}$
$\mathcal{V}_M^{\text{NMLL}}$	70,463	$3.0456 \times 10^{-4}$

Vine	5%	50%	95%
$\mathcal{V}_M^{\text{NSF}}$	$2.7046 \times 10^{-4}$	$3.0295e \times 10^{-4}$	$3.3330e \times 10^{-4}$
$\mathcal{V}_M^{\text{NMLL}}$	$4.2812 \times 10^{-4}$	$4.7508 \times 10^{-4}$	$5.2415 \times 10^{-4}$

Column “Det” contains determinants of the product moment correlation matrix corresponding the the vine specification or the product of one minus squared parameters on the vine. Column “LL” contains log-likelihood values.

the rank determinant for  $\mathcal{V}_T$  with 6 independent copulas were [ $4.7791 \times 10^{-4}, 5.2712 \times 10^{-4}, 5.8028 \times 10^{-4}$ ].

**SEQUENTIAL FITTING: NORMAL**

Parameters for the  $\mathcal{V}_M$  using sequential fitting procedure are given by  $\text{param}_M^N$  equal to

Change to text and equation layout to fit page width - OK?

$$\begin{bmatrix} * & 0.8460 & 0.8526 & 0.8672 & 0.8859 & 0.8163 & 0.8187 \\ & * & 0.4003 & 0.2370 & 0.0204 & 0.3925 & 0.1615 \\ & & * & 0.2883 & 0.3464 & 0.0416 & 0.4039 \\ & & & * & 0.0041 & -0.0616 & 0.1008 \\ & & & & * & -0.0480 & 0.1587 \\ & & & & & * & 0.0379 \\ & & & & & & * \end{bmatrix}$$

Changes to sentence OK?

The log-likelihood is 70,223, which is much smaller than that calculated earlier (see page 631 onwards). Parameters of normal copulas are (conditional) product moment correlations which are equal to partial correlations. Hence, we can calculate the product moment correlation matrix of the normal copula represented on  $\mathcal{V}_M$  with parameters obtained with sequential fitting. The determinant of this product moment correlation matrix is  $1.9199 \times 10^{-4}$ . Optimising all parameters together leads to a new set of parameters and a higher



log-likelihood for the data, equal to 70,463. The determinant of the product moment correlation matrix for the vine with parameters optimised together is  $3.0456 \times 10^{-4}$ .

In all cases examined so far, the sequential procedure for choosing optimal conditional bivariate copula parameters leads to greater joint dependence, as measured by the rank correlation determinant, than the procedure of optimising these parameters all together. Since the latter procedure leads to greater likelihood than the sequential procedure, it appears that the sequential procedure overestimates the second-order dependence structure. Optimising all parameters together smooths the distribution out.

## CONCLUSIONS

BBNs and regular vines pose some of the same challenges in model learning, and therefore can be treated with similar tools. Choosing an appropriate model requires separating the problems of estimating marginal distributions from the problem of identifying the dependence structure which connects the margins. Regular vines enable us to represent a joint distribution in terms of (conditional) bivariate margins, and hence allow us to split the margin/dependence structure in a convenient way: the margins are one dimensional and the dependence structure is that of a regular vine. Non-parametric continuous BBNs offer a different graphical structure for capturing the dependence, and the proof of their adequacy in this regard is spun off of results for regular vines. In either case, judging model adequacy requires a suitable scalar measure of multivariate dependence. The mutual information is the appropriate concept, and the determinant of the rank correlation matrix is proposed as a suitable proxy. This can be used to assess the adequacy of the dependence structure.

Of course, rank correlations do not determine the full dependence structure; for this we must choose conditional bivariate copulas. This entails choosing a copula class for each conditional bivariate distribution in the regular vine or BBN, and optimally choosing its parameter values. Owing to the complexity of these problems, heuristics must be used. Three heuristics are discussed, namely

- majorisation,
- most dependence in lowest trees, and
- least dependence in highest trees.

Changes to sentence OK?

'an application of'? Please clarify phrase.

The choice of heuristic is ultimately governed by convenience. Optimising parameter values can be done either sequentially or all at once. The application to Dutch wind speed data shows that the sequential procedures result in greater dependence than the all-at-once optimisation, regardless which heuristic is used. The all-at-once approach will break down for large data sets, leaving us only with sequential procedures. Understanding and correcting the overestimation of dependence of sequential procedures is an important task for the future.

- 1 The degree of a node is the number of edges attached to it.
- 2 Equivalently, we can formulate this proposition for edges of  $\mathcal{V}$ .
- 3 See Bedford and Cooke (2002).
- 4  $Y$  is an ancestor of  $X$  with respect to an ordering of the variables which preserves the parent-child relations, that is, an ordering such that parents occur before their children in the ordering.
- 5 See, for example, Kurowicka and Cooke (2006a).
- 6 See, for example, Harary (1969).
- 7 For details see Kurowicka and Joe (2010).

#### REFERENCES

- Aas, K., C.zado, A. Frigessi and H. Bakken, 2009 "Pair-Copula Constructions of Multiple Dependence", *Insurance: Mathematics and Economics* 44(2), pp. 182–98.
- Bedford, T. J., and R. M. Cooke, 2001, "Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines", *Annals of Mathematics and Artificial Intelligence* 32, pp. 245–68.
- Bedford, T. J., and R. M. Cooke, 2002, "Vines: A New Graphical Model for Dependent Random Variables", *Annals of Statistics* 30(4), pp. 1031–68.
- Cooke, R. M., 1997, "Markov and Entropy Properties of Tree and Vines-Dependent Variables", in *Proceedings of the ASA Section of Bayesian Statistical Science*, URL: <http://ssor.twi.tudelft.nl/-risk/>.
- Cooke, R. M., O. Morales and D. Kurowicka, 2007, "Vines in Overview", in *Proceedings of Third Brazilian Conference On Statistical Modelling in Insurance and Finance, Maresias*, URL: <http://ssor.twi.tudelft.nl/-risk/>.
- Cowell, R. G., A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter, 1999, *Probabilistic Networks and Expert Systems*, Statistics for Engineering and Information Sciences (Springer).
- Goossens, L. H. J., J. D. Harrison, B. C. P. Kraan, R. M. Cooke, F. T. Harper and S. C. Hora, 1997, "Probabilistic Accident Consequence Uncertainty Analysis: Uncertainty Assessment for Internal Dosimetry", Technical Report, NUREG/CR-6571, EUR 16773, SAND98-0119, Volume 1.
- Hanea, A., D. Kurowicka and R. M. Cooke, 2006, "Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets", *Quality and Reliability Engineering International* 22, pp. 709–29.

- Hanea, A., D. Kurowicka and R. M. Cooke**, 2010, "Mining and Visualising Ordinal Data with Non-Parametric Continuous BBNs", *Computational Statistics and Data Analysis* 54, pp. 668–87.
- Harary, F.**, 1969, *Graph Theory* (Addison-Wesley).
- Jensen, F. V.**, 1996, *An Introduction to Bayesian Networks* (London: Taylor and Francis).
- Jensen, F. V.**, 2001, *Bayesian Networks and Decision Graphs* (Springer).
- Joe, H.**, 1993, "Multivariate Dependence Measures and Data Analysis", *Computational Statistics and Data Analysis* 16, pp. 279–97.
- Joe, H.**, 1999, "Relative Entropy Measures of Multivariate Dependence", *Journal of the American Statistical Association* 84(405), pp. 157–64.
- Kendall, M. G., and A. Stuart**, 1961, *The Advanced Theory of Statistics* (London: Charles Griffin & Co).
- Kraan, B. C. P.**, 2002, "Probabilistic Inversion in Uncertainty Analysis and Related Topics", PhD Dissertation, TU Delft.
- Kurowicka, D., and R. M. Cooke**, 2000, "Conditional and Partial Correlation for Graphical Uncertainty Models", in *Recent Advances in Reliability Theory*, pp. 259–76 (Boston, MA: Birkhäuser).
- Kurowicka, D., and R. M. Cooke**, 2003, "A Parametrization of Positive Definite Matrices in Terms of Partial Correlation Vines", *Linear Algebra and Its Applications* 372, pp. 225–51.
- Kurowicka, D., and R. M. Cooke**, 2004, "Distribution-Free Continuous Bayesian Belief Nets", in *Mathematical Methods in Reliability*.
- Kurowicka, D., and R. M. Cooke**, 2006a, *Uncertainty Analysis with High Dimensional Dependence Modelling* (Chichester: John Wiley & Sons).
- Kurowicka, D., and R. M. Cooke**, 2006b, "Completion Problem with Partial Correlation Vines", *Linear Algebra and Its Applications* 418(1), pp. 188–200.
- Kurowicka, D., and R. M. Cooke**, 2007, "Sampling Algorithms for Generating Joint Uniform Distributions Using the Vine-Copula Method", *Computational Statistics and Data Analysis* 51(6), pp. 2889–906.
- Kurowicka, D., and H. Joe (eds)**, 2010, *Dependence Modeling: Vine Copula Handbook* (Singapore: World Scientific).
- Kurowicka, D., R. M. Cooke and U. Callies**, 2006, "Vines Inference", *Brazilian Journal of Probability and Statistics* 20, pp. 103–20.
- Marshall, A. W., and I. Olkin**, 1979, *Inequalities: Theory of Majorization and Its Applications* (San Diego: Academic Press).
- Morales Napoles, O., D. Kurowicka and A. Roelen**, 2007, "Elicitation Procedures for Conditional and Unconditional Rank Correlations", *Reliability Engineering and Systems Safety* 95(5), pp. 699–710.
- Morales Napoles, O.**, 2010, "About the Number of Vines and Regular Vines on  $n$  Nodes", in *Bayesian Belief Nets and Vines in Aviation Safety and Other Applications*, PhD Thesis, Chapter 2, Delft Institute of Applied Mathematics, Delft University of Technology.
- Pearl, J.**, 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Mateo: Morgan Kaufman).
- Rosenblat, M.**, 1952, "Remarks on a Multivariate Transformation", *Annals of Mathematical Statistics* 27, pp. 470–72.

**Shachter, R. D., and C. R. Kenley**, 1989, "Gaussian Influence Diagrams", *Management Science* 35(5), pp. 527–50.

**Whittaker, J.**, 1990, *Graphical Models in Applied Multivariate Statistics* (Chichester: John Wiley & Sons).

**Yule, G. U., and M. G. Kendall**, 1965, *An Introduction to the Theory of Statistics*, Fourteenth Edition (Belmont, CA: Charles Griffin & Co).