Research Article

A Mixture of Regular Vines for Multiple Dependencies

Fadhah Amer Alanazi [ID]

General Science Department, Prince Sultan University, Riyadh, Saudi Arabia

Correspondence should be addressed to Fadhah Amer Alanazi; fanazi@psu.edu.sa

To uncover complex hidden dependency structures among variables, researchers have used a mixture of vine copula constructions. To date, these have been limited to a subclass of regular vine models, the so-called drawable vine, fitting only one type of bivariate copula for all variable pairs. However, the variation of complex hidden correlations from one pair of variables to another is more likely to be present in many real datasets. Single-type bivariate copulas are unable to deal with such a problem. In addition, the regular vine copula model is much more capable and flexible than its subclasses. Hence, to fully uncover and describe complex hidden dependency structures among variables and provide even further flexibility to the mixture of regular vine models, a mixture of regular vine models, with a mixed choice of bivariate copulas, is proposed in this paper. The model was applied to simulated and real data to illustrate its performance. The proposed model shows significant performance over the mixture of R-vine densities with a single copula family fitted to all pairs.

1. Introduction

Real data often exhibit complex multivariate mixture dependency structures among variables. These dependencies may vary from one pair of variables to another. The variation in the dependency structures adds extra complexity for modelling and capturing these types of relationships. The Gaussian mixture model is commonly used to model data with complex dependency structures due to its ease of implementation. However, the main limitation of this model is that mixture Gaussian dependencies are assumed between all variables. Furthermore, all univariate marginal distributions are assumed to be Gaussian, which may not be the case for many real datasets. Hence, this type of model may not provide an adequate fit to the data, resulting in inaccurate estimates of quantities derived from the fitted model. Copula theory, however, provides two main strengths that overcome these limitations. First, in copula models, due to Sklar's theorem (Sklar [1]), margins are modelled separately from the dependency structures. Hence, margins do not need to be from the same type of distribution. Second, there is a wide range of copula families to describe various types of dependencies, including heavy tails. These two main advantages make copula a popular model in many areas. In a mixture context, for example, Zhang and Shi [2] introduced a Bayesian network with a copula model to study cancer data. Gunawan et al. [3] studied Bayesian estimation of copulas in high-dimensional cases of discrete and mixed margins. In finance, for example, a mixture of three copulas, Gaussian, Gumbel, and Gumbel survival, was studied by Hu [4]. Khaled and Kohn [5] investigated the properties of the mixture copula using different copula classes. Marbac et al. [6] introduced a mixture of Gaussian copulas.

Regardless of the flexibility of the copula model, it imposes the same dependency structures among variables. Furthermore, multivariate copula types are mostly limited to elliptical copula families. Also, extending copula functions to high-dimensional cases is known as a difficult problem [7].

To overcome the limitations of copula models, Aas et al. [7] introduced the so-called pair-copula construction (PCC) model. PCC decomposes a multivariate density into a cascade of bivariate copula functions. Therefore, only two variables are modelled at a time using bivariate copula families, possibly belonging to different types. These flexibilities have led PCC to receive great interest in the literature of different fields, for example, in geostatistics (Gräler and

Pebesma [8]; Erhardt et al. [9]) and in finance (Min and Czado [10, 11]).

In the mixture context, several researchers have incorporated pair-copula models with a finite mixture model. For example, Kim et al. [12] formed a mixture of drawable vine (D-vine) density models using multiple D-vine constructions that provide a high-flexibility mixture model and facilitate the comprehensive study of complex multivariate models. The authors illustrated the improvement of the model fit to capture hidden dependency structures. Similarly, a mixture of D-vine copulas was introduced by Zheng et al. [13] for chemical process monitoring. In their models, only one type of copula family was fitted to all pairs of variables.

A mixture canonical vine (C-vine) model was introduced by Sun et al. [14]. They showed that the performance of the C-vine copula mixture model was superior to other methods, including K-means and Gaussian mixture models (GMMs). Similarly, Qiu et al. [15] introduced C-vine and D-vine mixture copula models to analyze the dependency structure of the multiwind power output. Evkaya et al. [16] incorporated C-vine into the D-vine copula model (CD-vine) using fixed pair-copula for each density to model the dependency pattern among variables based on their temporal order.

Despite the strengths of the mixture PCC models, only C- and D-vine mixture-based models have been introduced in the literature due to their ease of computation. The regular vine (R-vine) model, however, is a general class that is formed as a mix of C- and D-vines. R-vine gives more flexibility for tree structures than C- and D-vines since it does not have a specific role in ordering the variables. Dißmann et al. [17] stated that R-vine copula-based models provide much more modelling capabilities. Besides, the authors provided a comparison example with several scenarios, including mixed R-vine (different copula types for each pair of variables). The authors concluded that overall, their example demonstrated the usefulness of the R-vine model with different choices of copula for each copula term.

Nowadays, R-vine models have been receiving increasing interest in the literature. For example, Dißmann et al. [17] provided a novel method for selecting and estimating R-vine copula models sequentially (tree by tree) using array representation. The array representation goes back to Morales Napoles et al. [18], who aimed to count the number of R-vine tree structures. The spatial R-vine model was introduced by Erhardt et al. [19]. In finance, two Bayesian R-vine models were developed by Gruber and Czado [20, 21]. In chemical process monitoring, Zhou and Li [22] applied the R-vine copula model to describe high-dimensional interdependence among complex variables.

In a mixture context, however, R-vine has not been investigated in the literature. Introducing mixture vine copulas into conventional vine pair-copula construction is a challenging task that opens up areas for future research regarding the selection of such models [21]. From the previous studies of the mixture PCC models and the R-vine models, a mixture of R-vine models may provide further flexibility for modelling highly complex hidden dependencies among the variables. Furthermore, specifying different bivariate copula types for each copula term improves the model fit for capturing multivariate dependencies among the variables and hence reduces the misspecification of the dependency structures. Therefore, a mixture of R-vine density models with mixed copula types will introduce further improvement and flexibility to the model. Although Kim et al. [12] mentioned, as a valuable extension of their work, fitting different bivariate copula types for all copula terms of a mixture of D-vines, this has not been investigated in the literature.

Therefore, inspired by and building on the works of Dißmann et al. [17] and Kim et al. [12], this paper develops the first mixture of the R-vine density model, with a mixed choice of bivariate copulas (copula family is specified individually for each pair in each density), in the literature. The new proposed model aims to introduce higher flexibility to mixture vine copula models, providing a way to fully captured mixed hidden complex correlations among variables.

The rest of the paper is organised as follows: Section 2 reviews a brief theoretical background of the copula and the regular vine copula. Section 3 discusses the selection strategy of the mixture of R-vines; the model describes the expectation-maximisation algorithm (a method used to estimate the model parameters) and provides an example of a two-component mixture R-vine density model stored in an array representation. Section 4 illustrates the performance of the proposed model in simulated and real datasets. Moreover, the new mixture model is compared with a mixture R-vine model where a single copula family is fitted to all pairs of the variables.

2. Theoretical Background

2.1. Copula. Generally, a copula function can be defined as a function that "links or joins" multivariate distribution functions to their univariate uniform marginal distribution functions [23].

Definition 1. Copulas, which were introduced by Sklar [1], are multivariate cumulative distribution functions with uniform marginal distributions on (0, 1) such that (Schweizer and Sklar [24])

$$C: [0, 1]^n \longrightarrow [0, 1], \quad n \geq 2. \tag{1}$$

Theorem 1 (see Sklar [1] in Joe [25]). *If F is an n-variate distribution function with univariate margins $F_1, F_2, \ldots \ldots, F_n$, then there exists an n-copula function C such that*

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \tag{2}$$

If the margins are continuous, then the copula

$$C(u_1, u_2, \ldots, u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_n^{-1}(u_n)) \tag{3}$$

is unique, where F^{-1} is the inverse function of the marginal functions and $u \in [0,1]^n$.

The joint density function can be obtained by taking the partial derivatives of (2):

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n) = {} & c_{12,\ldots,n}(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \\
& \times f_1(x_1) \times f_2(x_2), \ldots, f_n(x_n),
\end{aligned}
\tag{4}
$$

where $c_{12,\ldots,n}(\cdot)$ is a copula density function.

From (4), the joint density function can be factorised into its dependence structures and univariate margins. Hence, from Sklar's theorem (Sklar [1]), margins can be modelled separately from the dependence structures. This forms the main advantage of copula models. Another advantage of copulas is that they are invariant under strictly increasing transformations of the random variables, while the margins may be changed [26]. In addition to these advantages, there are various copula families, such as elliptical copula (Gaussian and t-student) and Archimedean copula (e.g., Frank, Clayton, and Gumbel), which can model a wide range of non-Gaussian dependence structures including tail dependencies. For more families, see Joe [25]. These advantages make copula-based models commonly used models in different areas such as finance (e.g., Embrechts et al. [27] and Cherubini et al. [28]), spatial statistics (e.g., Bárdossy [29] and Kazianka and Pilz [30, 31]), hydrogen production (e.g., Qiu et al. [15]), and climate study (e.g., Khan et al. [32]).

Despite the advantages of the copula-based model, it imposes the same dependency structures between all variables in high-dimensional cases. Also, constructing a higher-dimensional copula is known as a difficult problem (Aas et al. [7]). To overcome these limitations, a novel multivariate model using only bivariate copula families has been developed in the literature. This new model is known as a vine copula or a pair-copula-based model.

The following section introduces a brief theoretical background of the vine copula model. For more details, the interested readers are referred to Aas et al. [7] and Dißmann [33].

2.2. Pair-Copula.
A vine copula model decomposes multivariate copulas into bivariate copulas (pair-copula) to build high-dimension models using only bivariate copulas. In doing so, the pair-copula models provide a flexible way to model multiple complex high-order dependence structures using only bivariate copulas, possibly belonging to different copula families. The pair-copula model originated with Joe [34] and was later investigated and named as a regular vine model by Bedford and Cooke [35, 36]. Kurowicka and Cooke [37] further developed this model. Later, the full inference of the C-vine and D-vine models was introduced by Aas et al. [7].

In C-vine, all variables are modelled with respect to a particular variable, while in D-vine, variables are ordered sequentially.

Definition 2 (tree; see Bedford and Cooke [35]). $\mathcal{T} = \{N, E\}$ is a tree (an acyclic graph) with N nodes and E edges (connect each pair of N).

The degree of the node is the total number of edges connected to this node.

Definition 3 (vine, regular vine; see Ch. 4 of Kurowicka and Cooke [37]). V is a vine on d elements if

   (i) $V = (T_1, T_2, \ldots, T_{d-1})$, where T_1 indicates the first tree of the vine, and so on.

   (ii) T_1 is a connected tree with nodes $N_1 = \{1, 2, \ldots, d\}$ and edges E_1.

   (iii) For $j = 2, 3, \ldots, d-1$, $j = 2, 3, \ldots, d-1$ is a connected tree with nodes $D_j = E_{j-1}$.

   In addition, V becomes a regular vine on d elements if

   (iv) For $j = 2, \ldots, d-1$, if $l = \{l_1, l_2\}$ and $m = \{m_1, m_2\}$ in T_j are two nodes connected by an edge in T_j, then exactly one of l_j is equal to m_j, $j = 1, 2$. This condition is known as the proximity condition.

Under the proximity condition, two nodes in tree (T_{j+1}) are only connected by an edge if they were sharing a common node in the previous tree (T_j). Kurowicka and Cooke [37] defined the D- and C-vine models as follows:

   (i) If every node at the first tree of a regular vine is connected at maximum with two nodes, then the regular vine is called D-vine.

   (ii) If at each tree T_j of a regular vine, there is one particular node that is connected to all other nodes, then the regular vine is called a C-vine. At the first tree, this node is called a root node.

Definition 4 (regular vine (R-vine) specification; see Bedford and Cooke [36]). $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ is a regular vine copula (R-vine copula) specification if

   (i) $\mathcal{F} = (F_1, F_2, \ldots, F_n)$ is a vector of continuous invertible distribution functions

   (ii) $\mathcal{V}$ is an n-dimensional regular vine (R-vine)

   (iii) $\mathcal{B} = \{B_e | i = 1, 2, \ldots, n-1; e \in E_i\}$ is a set of bivariate copulas

Let $x = (x_1, \ldots, x_n)$ be a vector of random variables, $e = \{l, m\}$ an edge, and $i = 1, \ldots, n$ and D_e a conditioning set of edge e. Bedford and Cooke [36] defined regular vine dependence as follows.

Definition 5 (regular vine (R-vine) dependence). A joint distribution function F on x is said to realise a regular vine copula specification $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ or exhibit regular vine dependence if for each $e \in E_i$, the bivariate copula of X_{Cel} and X_{Cem} given X_{De} is a member of the bivariate copula $B_e(X_{De})$. The marginal distribution of X_i is F_i, for $i = 1, 2, \ldots, n$.

The bivariate copula of X_{Cel} and X_{Cem} given X_{De} is a conditional bivariate copula which is assumed to be

independent of conditioning variables (see Aas et al. [7] and Haff et al. [38]).

Theorem 2 (see [17]). *Let $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ be an n-dimensional regular vine specification. Then, there is a unique distribution function F that realizes $(\mathcal{F}, \mathcal{V}, \mathcal{B})$. Its density is*

$$f_{(1,2,\ldots,n)}(x) = \prod_{j=1}^{n} f_j(x_j) \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{Cel,Cem|D_e}\big(F_{Cel|D_e}$$

$$\big(x_{Cel}|\mathbf{x}_{D_e}\big), F_{C_{em}|D_e}\big(x_{C_{em}}|\mathbf{x}_{D_e}\big)\big), \tag{5}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $e = \{l, m\}$, $\mathbf{x}_{De}$ denotes a conditioning variable in a conditioning set D_e, i.e., $\mathbf{x}_{D_e} = \{x_i | i \in D_e\}$, and f_i is the density of F_i, $i = 1, 2, .., n$. Moreover, $c_{C_{el}, C_{em}|D_e}$ stands for the density function of bivariate copulas between edge $e = \{l, m\}$.

Continuing with the last theorem, let $e \in E_i$, $e = \{l, m\}$, $l = \{l_1, l_2\}$, $m = \{m_1, m_2\}$, be the edge that joined C_{el} and C_{em}. Joe [34] showed that the conditional marginal distribution, $F_{Cel|D_e}(x_{Cel}|\mathbf{x}_{D_e})$ and $F_{C_{em}|D_e}(x_{C_{em}}|\mathbf{x}_{D_e})$, can be obtained as follows:

$$F_{Cel|D_e}\big(x_{Cel}|\mathbf{x}_{D_e}\big) = \frac{\partial C_{C_l|D_l}\big(F_{C_{l l_1}|D_l}\big(x_{C_{l l_1}}|\mathbf{x}_{D_l}\big), F_{C_{l l_2}|D_l}\big(x_{C_{l l_2}}|\mathbf{x}_{D_l}\big)\big)}{\partial F_{C_{l l_2} D_l}\big(x_{C_{l l_2}}|\mathbf{x}_{D_l}\big)},$$

$$=: h\big(F_{C_{l l_1}|D_l}\big(x_{C_{l l_1}}|\mathbf{x}_{D_l}\big), F_{C_{l l_2}|D_l}\big(x_{C_{l l_2}}|\mathbf{x}_{D_l}\big)\big). \tag{6}$$

$F_{CelD_e}(x_{Cel}|\mathbf{x}_{D_e})$ and $F_{C_{em}D_e}(x_{C_{em}}|\mathbf{x}_{D_e})$ are then called transformed variables (see Aas et al. [7] and Dißmann et al. [17]).

Regular vine is a general case of vine copula that includes both C- and D-vines. The number of R-vine tree structures is considerable in comparison with C- and D-vines. Morales Napoles et al. [18] showed that, for n variables, there are $(n!/2) \cdot 2\binom{n-2}{2}$ possible R-vine tree structures. The following presents one possible example of a 5-dimensional R-vine copula model (following the details given in Aas et al. [7]).

2.2.1. Example of a 5-Dimensional R-Vine Copula. Figure 1 shows the tree structure of a 5-dimensional R-vine copula (Aas et al. [7]).

In this example, we have 5 variables, 4 trees, and 10 edges. At the first tree, all pairs of variables are modelled via unconditional bivariate copula families. For the upper levels, the dependencies are captured via conditional bivariate copulas. In this case, there are two sets of variables: the conditioning and conditioned variables. For example, $c_{13|2}$ is the density of the conditional bivariate copula family that models the dependency structure between the first and third variables given the second variable. In this case, the first and third variables are called conditioned variables, while the second variable is called a conditioning variable. In general, all the variables that are present before | are called conditioning variables, while the variables after | are called conditioned variables. The joint density of this 5-dimensional R-vine is as follows:

$$f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot f_5(x_5),$$

$$c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \cdot c_{34}\{F_3(x_3), F_4(x_4)\} \cdot c_{35}\{F_3(x_3), F_5(x_5)\},$$

$$c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{24|3}\{F(x_2|x_3), F(x_4|x_3)\} \cdot c_{45|3}\{F(x_4|x_3), F(x_5|x_3)\}, \tag{7}$$

$$c_{14|23}\{F(x_1|x_2,x_3), F(x_4|x_2,x_3)\} \cdot c_{25|34}\{F(x_2|x_3,x_4), F(x_5|x_3,x_4)\}$$

$$\cdot c_{15|234}\{F(x_1|x_2,x_3,x_4), F(x_5|x_2,x_3,x_4)\}.$$

3. A Mixture of R-Vine Densities

Mixture models provide a flexible way to capture multivariate hidden dependency structures among variables. A finite mixture model is a model applied to data that are assumed to be generated from a finite number of unknown distributions. The density in the finite mixture model is a weighted sum of a finite number of densities.

Kim et al. [12] formed a mixture of D-vine models with a single copula family fitted to all copula pairs. This section extends the work of Kim et al. [12] from a mixture of D-vines to a mixture of R-vine model with an individual choice of copula types for each pair in each density of the model.

3.1. Finite Mixture Model. Suppose that an n-dimensional random vector $\mathbf{x} = (\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kn})$ is generated from a k-component mixture of R-vine density model. Hence, the density of $\mathbf{x}$ is given by

$$f(\mathbf{x}|\Theta) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{x}|\theta_k), \tag{8}$$

where π_k is an unknown parameter (known as a mixture coefficient or weights) of the kth component that satisfies the following:

$$0 < \pi_k < 1,$$

$$\sum_{k=1}^{K} \pi_k = 1. \tag{9}$$

Θ is the set of all model parameters, while θ_k is the set of all the parameters of the kth component. In mixture models,
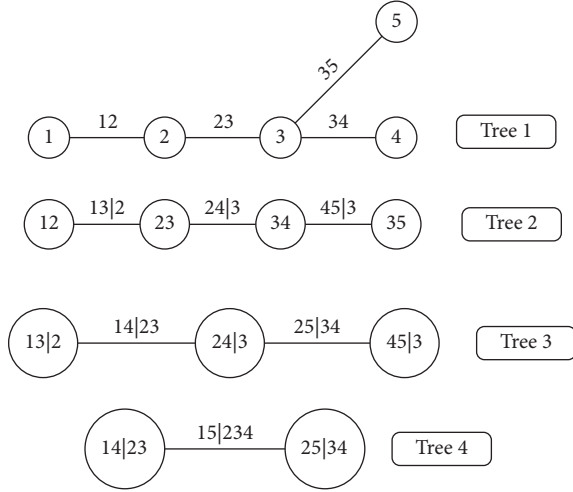
Figure 1: 5-dimensional R-vine copula, 4 trees and 10 edges.

expectation-maximisation algorithm (EM algorithm) is a commonly used method to estimate the model parameters. Further details of this method will be introduced in the next section.

3.1.1. EM Algorithm. Expectation-maximisation algorithm (EM) (Dempster et al. [39]) is an estimation method with two steps: the so-called expectation step (E-step) and the maximisation step (M-step).

Suppose that a dataset of size N is drawn independently from the k-component mixture of R-vine density model given in equation (8). Suppose further that the data are converted into uniform distribution using an empirical cumulative distribution function. Then, the log pseudo-likelihood function of Θ is given as follows:

$$l(\Theta) = \log\left\{\prod_{n=1}^{N}\sum_{k=1}^{K}\pi_k f(x_n|\theta_k)\right\} = \sum_{n=1}^{N}\log\sum_{k=1}^{k}\pi_k f(x_n|\theta_k),\tag{10}$$

where Θ is the set of all model parameters and θ_k is the set of all the parameters of the kth component. EM algorithm introduces a latent variable $z_n = (z_{n1}, z_{n2}, \ldots, z_{nk})$, where $z_{nk} = 1$ if $\mathbf{x_n}$ is drawn from the kth component (the kth component of the mixture model) and $z_{nk} = 0$, otherwise. In other words, z_{nk} indicates from which mixture component each observation was drawn. These latent variables are assumed to be independent and identically distributed from the multinomial distribution such that

$$z_{nk} \sim \mathcal{M}(k, \pi_1, \ldots, \pi_k), \quad k = 1, \ldots, K.\tag{11}$$

Consequently, we now have the complete data:

$$\mathbf{x_c} = (\mathbf{x_n}, \mathbf{z_n}), \quad \mathbf{z} = (z_1, \ldots, z_k).\tag{12}$$

Then, the complete-data log likelihood function, $\ell_c(\Theta)$, is given as follows:

$$\ell_c(\Theta) = \sum_{n=1}^{N}\sum_{k=1}^{K}z_{nk}\log \pi_k + \sum_{n=1}^{N}\sum_{k=1}^{K}z_{nk}\log f(\mathbf{x}_n|\theta_k).\tag{13}$$

EM algorithm starts with initial values of the unknown parameters $\Theta^{(0)}$, and the two steps (E and M) are repeated until the convergence $(\ell(\Theta)^{(i+1)} - \ell(\Theta)^{(i)})$ is smaller than a prespecified tolerance.

E-step: calculate the conditional expectation of the complete-data log likelihood, $\ell_c(\Theta)$ in equation (14), given the observed data and using the current estimate of the parameters (Θ).

Suppose that we are at iteration $i + 1$. Then, the conditional expectation of z_{nk} is calculated as follows:

$$\widehat{z}_{nk} = E\left[z_{nk}|\mathbf{x}, \Theta^{(i)}\right] = P\left[z_{nk} = 1|\mathbf{x}, \Theta^{(i)}\right] = \frac{\pi_k^{(i)} f\left(\mathbf{x_n}|\theta_k^{(i)}\right)}{\sum_{s=1}^{K}\pi_s^{(i)} f\left(\mathbf{x_n}|\theta_k^{(i)}\right)}.\tag{14}$$

M-step: maximize the complete-data log likelihood, $\ell_c(\Theta)$ (from E-step), with respect to (Θ) in order to produce a new estimate of the model parameters $(\Theta^{(n+1)})$. In this step, the estimation of each component parameter is computed independently, i.e., $\pi_1^{(i+1)}, \ldots, \pi_k^{(i+1)}$ and $\theta_1^{(i+1)}, \ldots, \theta_k^{(i+1)}$.

The new estimate of $\pi_k^{(i+1)}$ can be obtained as follows:

$$\widehat{\pi}_k^{(i+1)} = \frac{\sum_{n=1}^{N}\widehat{z}_{nk}^{(i)}}{N},\tag{15}$$

while the updated $\theta_k^{(i+1)}$ can be obtained by maximising the following equation using the numerical maximisation method:

$$\widehat{\theta}_k^{(i+1)} = \max_{\theta}\sum_{n=1}^{N}\sum_{k=1}^{K}\widehat{z}_{nk}^{(i)}, \quad \log f\left(\mathbf{x_n}|\theta_k^{(i)}\right).\tag{16}$$

3.2. Model Selection of the Mixture of R-Vine Densities. Statistical inference algorithms for computing the log-likelihood functions and simulating strategy from an R-vine distribution using the lower triangular arrays of R-vine copula models were introduced by Dißmann et al. [17]. Algorithm 2.1 from Dißmann et al. [17] computes a given R-vine specification density. In addition, the authors showed how to calculate the log likelihood of the evaluated density to be used for estimating the pair-copula parameters using, for example, maximum likelihood. Their method introduced high flexibility for R-vine copula models.

Inspired by and based on Algorithm 2.1 from Dißmann et al. [17] and the method of Kim et al. [12], the proposed mixture of R-vine density model with mixed copula families is presented in this section. I termed it a Mixture of R-vine Density Model with Mixed Families (MRDMMF). An example of constructing a two-component mixture of R-vine model, in an array representation, is given in Section 3.3.

3.2.1. Full Inferences and the Selection Strategy of the R-Vine Mixture Model. The full inference of pair-copula models (see Aas et al. [7]), in general, requires the following four main steps:

   (i) Obtain the normalised ranks of the original data

   (ii) Order the variables (tree structure)

(iii) Select the appropriate bivariate copula type for each copula term (possibly belonging to different copula types)

(iv) Estimate the parameters of the bivariate copula families

For the first step and to avoid misspecifying the margins, it is useful to transform the margins nonparametrically using the empirical cumulative distribution function.

A comparison study between semiparametric and parametric methods for estimating copulas was investigated by Kim et al. [40]. The authors illustrated, by simulation studies, that the performance of the pseudo-maximum likelihood (PML) function (see Genest et al. [41] and Ch. 5 of Cherubini et al. [28]) for estimating copula parameters is significant in comparison to the full maximum likelihood (ML) and the inference function for margins (IFM) (Joe [42, 43]) methods when the marginal distributions are unknown, which is almost always the case in practice.

For the second step, it is well known that there are different possible ways to order the variables, and one needs to select the most appropriate way. The only way that guarantees, with no doubt, that the chosen order is the best is by testing all these possible structures, which is infeasible, especially in high-dimensional cases. One existing method of selecting the tree structure is based on the largest values of Kendall's tau (see, for example, Aas and Berg [44] and Kim et al. [12]).

Dißmann et al. [17] introduced a tree-by-tree selection strategy, or a sequential estimation, using the maximum spanning tree. In their model, at the first tree, the empirical Kendall's tau is computed for all possible pairs of variables. Then, the tree that maximises the sum of absolute empirical Kendall's tau is selected. Finally, the pair-copula families are chosen for each pair, and corresponding parameters are estimated. Aas et al. [7] ordered the variables based on the strongest tail dependencies.

Having selected the appropriate order of the variables, identifying the bivariate copula type that best fits the data is the most crucial part of the pair-copula models. In the literature, several methods have been introduced to solve this challenging issue. For example, Aas et al. [7] used a scatter plot to determine the shape of each copula family. Other methods involve selection criteria, such as the Akaike information criteron (AIC) from Akaike [45] (see, for example, Dißmann et al. [17]). After that, the final step of the pair-copula models is estimating the copula parameters.

The finite mixture of the R-vine density model, however, requires estimating not only the pair-copula type but also the number of mixture components. For k components' mixture of R-vine model, there are $k*(d*(d-1)/2)$ bivariate copula families to be specified and estimated. These families do not have to be of the same type. Thus, constructing mixture R-vine density models with mixed copula types is challenging. The difficulty lies in specifying the appropriate bivariate copula type for each pair in each density (component). Testing all possible bivariate copula types in each mixture component is highly infeasible in practice. However, even though the scatter plot, in the mixture context, is

almost not straightforward enough to provide exact information on the involved pair-copula types, due to the mixture dependencies, it can still give some information on the possible types of the dependence structures. In addition to the scatter plot point, Dißmann et al. [17] said that the bivariate copula types determined at the first tree have a significant influence on the model fit. From these two points, a method for specifying bivariate copula types in the mixture context is introduced based on the scatter plot method of Aas et al. [7]. Following the structure of Aas et al. [7], the first step is plotting the original data. Then, based on the possible information extracted from the scatter plot of each pair, several possible mixture models are constructed (first tree only). After that, the selection criteria are computed to determine the most appropriate model. For this step, three commonly used selection criteria are employed, namely, Akaike information criterion (AIC) of Akaike [45]; Bayesian information criterion (BIC) of Schwarz et al. [46]; and the consistent Akaike information criteria (CAIC) of Bozdogan [47]. The formulas of these criteria are given as follows:

$$
\begin{aligned}
\texttt{AIC} &= -2\ln \texttt{L}\left(\widehat{\theta}\right) + 2\texttt{P}, \\
\texttt{BIC} &= -2\ln \texttt{L}\left(\widehat{\theta}\right) + \texttt{P}(\ln(N)), \\
\texttt{CAIC} &= -2\ln \texttt{L}\left(\widehat{\theta}\right) + \texttt{P}(\ln(N) + 1),
\end{aligned}
\tag{17}
$$

where $\widehat{\theta}$ is the estimation value of the parameters, P is the number of estimated parameter in the model, and N is the number of the observations.

After that, the remaining trees of the selected model are determined following the same step as Aas et al.'s [7] method. However, instead of estimating the model parameters sequentially, the model parameters are estimated jointly using the EM algorithm.

To end this section, the main idea of specifying the bivariate copula type in the mixture R-vine density model can be summarised as follows:

(1) Plot the original data. From the plot, construct different possible mixtures of R-vine density models. In this step, only the first tree of each density of each model is constructed.

(2) Fit all the constructed mixture of R-vine density models to the data.

(3) Compute the values of AIC, BIC, and CAIC. Then, the model with the smallest values is selected as the most appropriate model.

(4) Then, for the second tree and based on the selection families of the first tree of each density of the selected model, a method from Aas et al. [7] is used to determine the bivariate copula types for each pair.

(5) After that, the model parameters are estimated jointly using the EM algorithm.

(6) For the remaining trees T = 3, . . . , d, Steps 4 and 5 are repeated until the last tree is reached.

The following steps summarise the full inference of the MRDMMF:

(i) Obtain the normalised ranks of the observed data

(ii) Construct different mixture of R-vine density models with mixed copula families

(iii) Estimate copula parameters using the EM algorithm

(iv) Fit all models to the data, and compute the AIC, BIC, and CAIC

(v) The model with the smallest values of the selection criteria is selected as the best-fit model to the data

Since the primary focus of this paper is to introduce a new mixture model, the types of bivariate copula families used in the simulation studies are prespecified as the most commonly used copula families. In addition, the number of mixture components in both the simulation and the real datasets is fixed at two.

3.3. A Mixture of R-Vine Densities in an Array Representation.
This section explains by example the representation array of the mixture of R-vine density model. In this model, each array of the pair-copula families is treated as a mixture component.

Example 1. (example of a two-component mixture of R-vine densities in an array representation).

To store a two-component mixture of R-vine density model of one-parameter copula families in an array representation, one need to have five arrays: one array for the tree structure, two arrays for storing the involved bivariate copula types (each array stands for a single mixture component), and another two arrays for their corresponding parameters. In this example, S stands for the R-vine tree structure (numbers refer to the variables), while BC_i and BCp_i are the mixture components (numbers refer to the bivariate copula) and the corresponding parameters, respectively.

$$S = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 1 & 3 & 3 & 3 & 0 \\ 3 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$BC_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & 0 \\ 4 & 5 & 6 & 0 & 0 \\ 4 & 5 & 6 & 3 & 0 \end{pmatrix},$$

$$BC_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 5 & 6 & 0 & 0 & 0 \\ 3 & 4 & 3 & 0 & 0 \\ 5 & 3 & 3 & 5 & 0 \end{pmatrix},$$

$$BCp_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 3.5 & 0 & 0 & 0 & 0 \\ 3.5 & 3 & 0 & 0 & 0 \\ 3 & 3.9 & 3.5 & 0 & 0 \\ 2.9 & 3 & 2.5 & 1.8 & 0 \end{pmatrix},$$

$$BCp_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 2.5 & 0 & 0 & 0 & 0 \\ 3 & 3.9 & 0 & 0 & 0 \\ 1.5 & 2.5 & 3 & 0 & 0 \\ 3.5 & 2 & 1.5 & 3.5 & 0 \end{pmatrix}. \tag{18}$$

Figures 2 and 3 show the construction and the contour plots of the given mixture of R-vine model, respectively (these plots are generated via contour.RVineMatrix and plot (RVM) functions from the VineCopula package [48] in R [49]).

4. Numerical Application

This section aims to illustrate the efficiency of the model fit of the MRDMMF. Therefore, the simulation studies were designed with two main scenarios. The first scenario tests the performance of the proposed model in order to evaluate its ability to estimate the true complex multivariate dependency structures. The second scenario contains two main parts. The first part compares the performance of MRDMMF with one of the mixtures of R-vine model where a single copula family is fitted to all pairs of variables. I termed the latter model the Single-Family Mixture of R-vine Density Model (SFMRDM). This comparison aims to express the flexibility and the usefulness of MRDMMF over the SFMRDM. The second part aims to study the effect of misidentifying the number of mixture components on the model fit when MRDMMF is fitted to a nonmixture dataset (a similar scenario was applied by Kim et al. [12]).

For both scenarios, two samples of size 300 and 1500 were simulated using the RVineSim function from the VineCopula package in R (Schepsmeier et al. [48]). These datasets were repeated 100 times each. The simulated data for the first scenario are based on the mixture of R-vine density model given in Example 1 (the mixture weight is $\pi_1 = 0.6$), while the data for the second scenario are a nonmixture dataset based on the first density of the mixture model.

4.1. Simulated Data Application

4.1.1. First Scenario. In this scenario, the two-component MRDMMF model, given in Example 1, is fitted to the simulated datasets. To evaluate its performance, the difference between the true and the mean of the estimated parameter values of the simulated data is computed. Figure 4 shows the scatter and contour pair plots of the simulated dataset (size 300). Table 1 gives a full description (or information) of the simulated mixture of the R-vine density
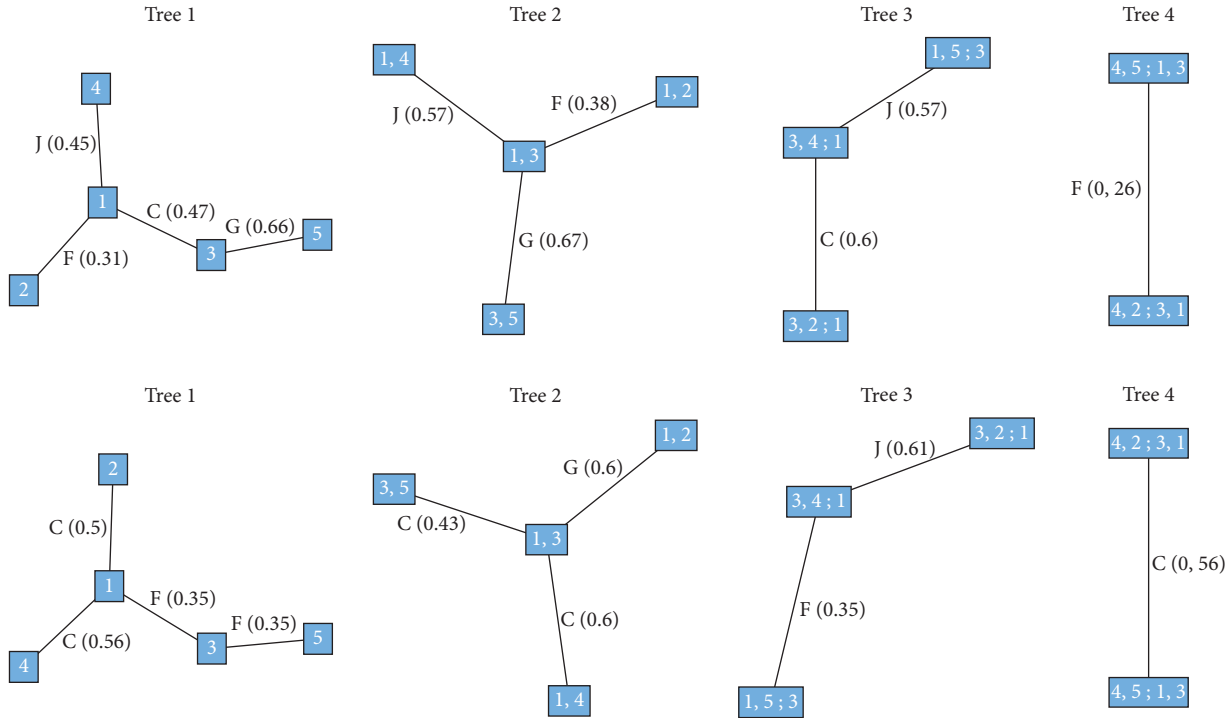
FIGURE 2: Construction plots of the mixture components. Top: first component. Bottom: second component. F: Frank, J: Joe, G: Gumbel, and C: Clayton. The values inside the parentheses are the Kendall tau's values.
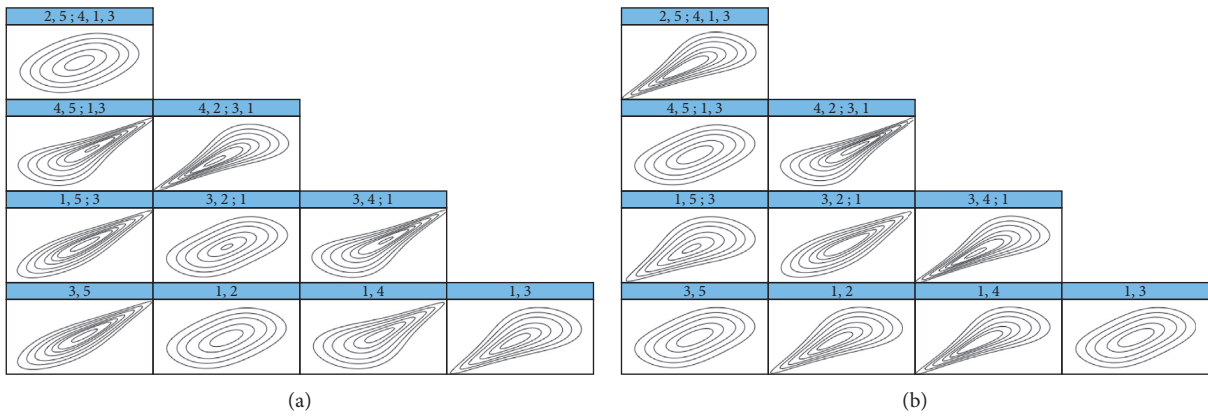

FIGURE 3: Contour plot of the R-vine model. (a) Contour plot of the first component. (b) Contour plot of the second component.

model. This description includes information of the type of the pair-copula families fitted for each pair, its parameters, and the corresponding Kendall's tau (τ) values. For convenience, the short names of the bivariate copula types are used. Hence, the involved bivariate copula families, with their short names, are Frank (F), Clayton (C), Gumbel (G), and Joe (J). The estimation result is summarised in Table 2.

Comparing the true simulation model (provided in Table 1) and the estimated model (provided in Table 2) shows that the estimated correlation parameters and the corresponding Kendall's tau values are very close to the true values. Hence, the performance of the EM algorithm is satisfied, and the underlying dependencies are well captured.

4.1.2. Second Scenario (First Part). This section aims to express the significant performance and the improvement of the model fit of the MRDMMF over the SFMRDM. For the single mixture model, Frank copula is specified as a pair-copula type for all pairs in the mixture of R-vine densities. After that, all models are fitted to the data, and the values of different selection criteria (AIC, BIC, and CAIC) are computed for each model. Their values are shown in Table 3.

From Table 3, the values of all selection criteria of both datasets support the MRDMMF over SFMRDM. The results also show a very poor fit of the SFMRDM. Hence, MRDMMF improves the model fit for modelling high multivariate complex dependence structures.
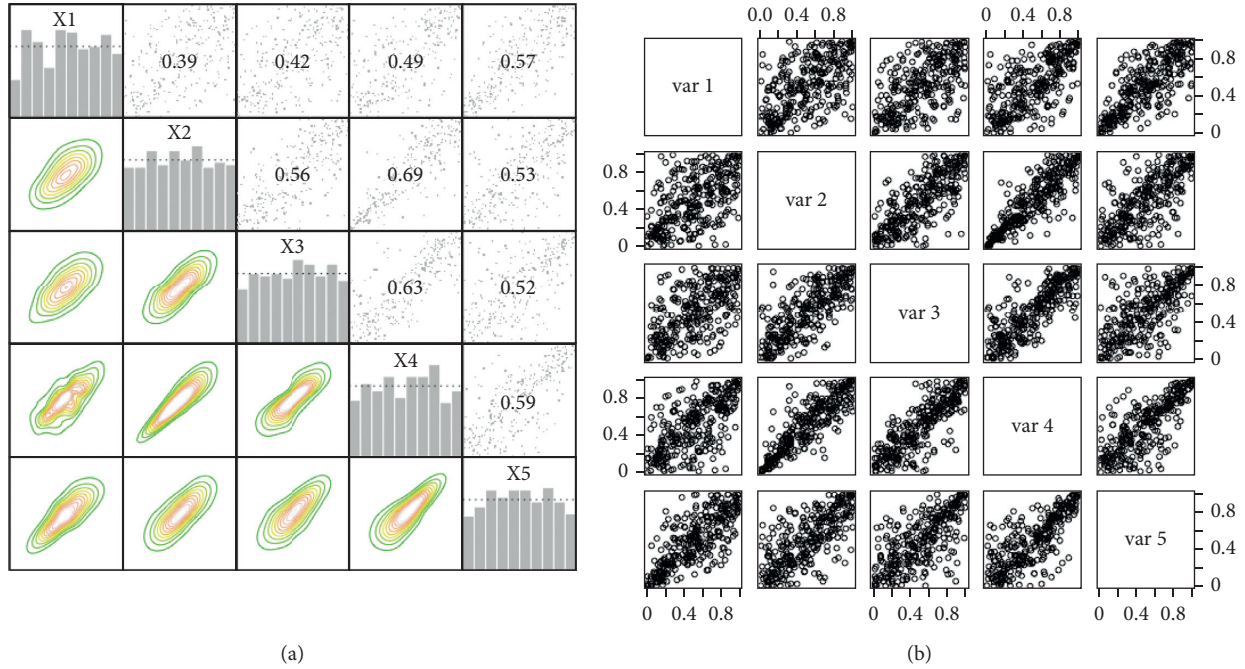
(a)

(b)

FIGURE 4: (a) The contour (lower triangular), scatter plot (upper triangular), and the correlation computed by Kendall's tau (the diagonal). (b) The scatter pair plot of the simulated data based on the mixture of R-vine densities.

TABLE 1: Summary of the mixture of R-vine density model (simulated data of size 300).

	Edge	Type (1st component)	$\theta\ (\tau)$	Type (2nd component)	$\theta\ (\tau)$
1st tree					
	3, 5	G	2.9 (0.66)	F	3.5 (0.35)
	1, 2	F	3 (0.31)	C	2 (0.5)
	1, 4	J	2.5 (0.45)	C	2.5 (0.56)
	1, 3	C	1.8 (0.47)	F	3.5 (0.36)
2nd tree					
	1, 5; 3	G	3 (0.67)	C	1.5 (0.43)
	3, 2; 1	F	3.9 (0.38)	G	2.5 (0.6)
	3, 4; 1	J	3.5 (0.57)	C	3 (0.6)
3rd tree					
	4, 5; 1, 3	J	3.5 (0.57)	F	3.5 (0.43)
	4, 2; 3, 1	C	3 (0.6)	J	3.9 (0.6)
4th tree					
	2, 5; 4, 1, 3	F	2.5 (0.26)	C	2.5 (0.56)

4.1.3. Second Scenario (Second Part). The previous section demonstrated the usefulness of the mixture of R-vine model with a mix choice of pair-copula type in each density over the single-family model. In this part, the effect of mis-specifying the number of mixture components is investigated. For this case, the mixture of R-vine model is fitted to the nonmixture simulated datasets. The performance of the model is shown in Tables 4 and 5.

From the result, it is evident that the mixture weight of the first component of the model is very high (it almost equals one), while the weight of the second component is minimal (almost zero). This is very acceptable since the data are nonmixture. For the estimation of the parameters, the results show that the estimated parameters are very close to the true values. Hence, the results support and demonstrate the performance of the proposed model.

4.2. Real Data Application

4.2.1. Real Datasets. This section aims to illustrate the performance of the proposed model with real datasets. For (only) an evaluation test, I applied the newly proposed mixture of R-vine density model (MRDMMF) to two different datasets, namely, Newthyroid and Glass datasets, from the [50] repository. These datasets consist of 215 and 214 observations, respectively. Figure 5 shows the scatter pair plots of both original data.

TABLE 2: Summary of the estimation results of the mixture of R-vine density model (simulation data of size 300), where SD stands for the standard deviation.

	Edge	Type (1st component)	θ (τ) (SD)	Type (2nd component)	θ (τ) (SD)
1st tree					
	3, 5	G	2.895 (0.655) (0.061)	F	3.494 (0.348) (0.262)
	1, 2	F	3.013 (0.308) (0.250)	C	1.981 (0.497) (0.086)
	1, 4	J	2.510 (0.450) (0.056)	C	2.480 (0.553) (0.088)
	1, 3	C	1.799 (0.473) (0.063)	F	3.496 (0.349) (0.143)
2nd tree					
	1, 5; 3	G	3.013 (0.670) (0.164)	C	1.494 (0.427) (0.150)
	3, 2; 1	F	3.915 (0.381) (0.403)	G	2.509 (0.601) (0.098)
	3, 4; 1	J	3.528 (0.573) (0.186)	C	3.022 (0.602) (0.188)
3rd tree					
	4, 5; 1, 3	J	3.569 (0.575) (0.342)	F	3.600 (0.355) (0.616)
	4, 2; 3, 1	C	3.031 (0.601) (0.300)	J	3.980 (0.610) (0.369)
4th tree					
	2, 5; 4, 1, 3	F	2.520 (0.262) (0.508)	C	2.629 (0.565) (0.382)

TABLE 3: Selection criteria values of the second scenario of the simulation study.

Model	AIC	BIC	CAIC	π_1
Simulation data (size 300)				
MRDMMF	−2237.701	−2165.128	−2164.128	0.598
SFMRDM (Frank)	−1413.754	−1410.05	−1409.05	0.532
Simulation data (size 1500)				
MRDMMF	−11161.95	−11156.64	−11155.64	0.599
SFMRDM (Frank)	−6898.332	−6786.755	−6765.755	0.57

TABLE 4: Estimated values of the mixture weight of the simulation study (second scenario).

Sample size	Model	π_1
$N = 300$	MRDMMF	0.982
$N = 1500$	MRDMMF	0.994

TABLE 5: Summary of the estimation results of the mixture of R-vine density model fitted to nonmixture simulated datasets for the simulated data of size 300 and 1500.

	Edge	Type	θ (τ)	SD
1st tree				
	3, 5	G	2.911 (0.656)	0.046
	1, 2	F	3.022 (0.309)	0.0172
	1, 4	J	2.499 (0.449)	0.040
	1, 3	C	1.802 (0.474)	0.042
2nd tree				
	1, 5; 3	G	3.020 (0.668)	0.115
	3, 2; 1	F	3.975 (0.386)	0.246
	3, 4; 1	J	3.531 (0.574)	0.119
3rd tree				
	4, 5; 1, 3	J	3.568 (0.577)	0.226
	4, 2; 3, 1	C	3.103 (0.607)	0.245
4th tree				
	2, 5; 4, 1, 3	F	2.638 (0.274)	0.401
1st tree				
	3, 5	G	2.899 (0.655)	0.020
	1, 2	F	3.027 (0.310)	0.054
	1, 4	J	2.501 (0.449)	0.014
	1, 3	C	1.8 (0.474)	0.017
2nd tree				
	1, 5; 3	G	3.018 (0.669)	0.043
	3, 2; 1	F	3.909 (0.381)	0.117
	3, 4; 1	J	3.512 (0.572)	0.058
3rd tree				
	4, 5; 1, 3	J	3.517 (0.573)	0.083
	4, 2; 3, 1	C	3.006 (0.600)	0.099
4th tree				
	2, 5; 4, 1, 3	F	2.508 (0.263)	0.158

Following the steps of the mixture of R-vine model, the observations of both datasets are converted into pseudo-observations in order to obtain the copula data. For this step, the pobs function from the R package copula (Hofert et al. [51]; Jun Yan [52]; Ivan Kojadinovic and Jun Yan [53]; and Marius Hofert and Martin Mächler [54]) is used. After that, the most appropriate order of the variables is selected based on the highest values of the absolute Kendall's tau. The tree structures of both datasets are shown in Figures 6 and 7.

After defining the R-vine tree structure and following the R-vine representation array (see Dißmann [33]), one can define the R-vine matrix corresponding to the selected R-vine structure. Please note that the R-vine matrix is not unique as the R-vine structure can be represented by different R-vine matrices (Dißmann et al. [17]). RVineStructureSelect function from the (Schepsmeier et al. [48]) package was used to find the structure of the Glass dataset (to save effort and time) using Kendall's tau as the edge weight, while the structure of the Newthyriod dataset was built manually based on the highest values of Kendall's tau.

Having determined the order of the variables, the next step is constructing a different possible mixture of R-vine

density models (first tree only) based on the possible information from the scatter plot of the original data.

Figure 8 shows contour and scatter pair plots of the real datasets. The first dataset clearly shows that some of the bivariate dependencies, not perfectly (due to hidden mixture dependencies) but almost, reflect one type of copula family. For example, the dependency structure between x_2 and x_4 is an upper tail dependence. For the second dataset, the situation is less clear. However, one can still get some
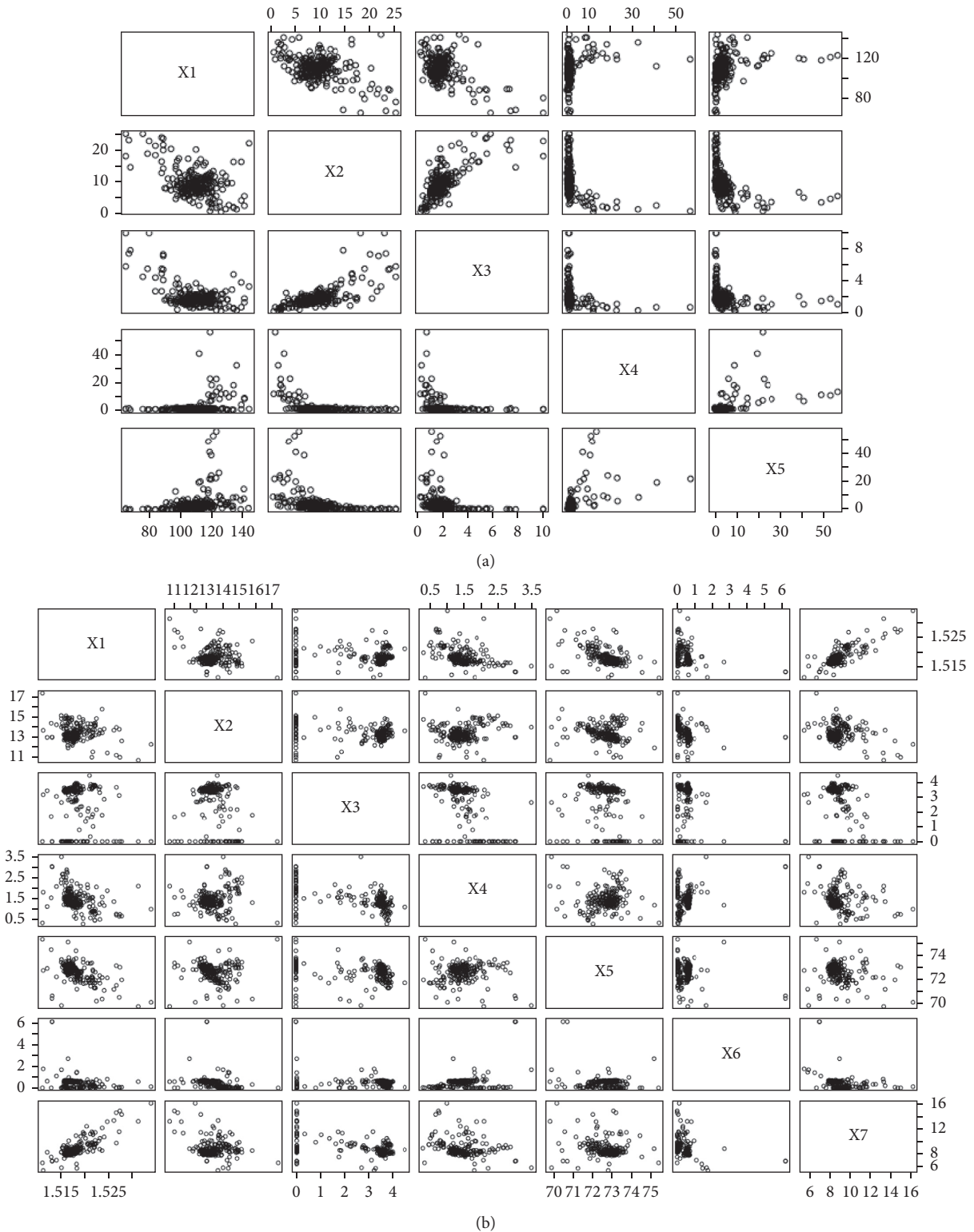
(a)

(b)

FIGURE 5: Scatter pair plots of the real datasets. (a) Newthyroid dataset. (b) Glass dataset.

information. For example, from the plot, some pairs of variables exhibit tail dependencies. Again, this does not show the exact type of the appropriate bivariate copula family. However, extending the number of components of this model may improve the model fit since increasing the

number of components gives extra flexibility to fit several bivariate copula types.

Following the selection strategy given in Section 3, the mixture of R-vine model is constructed with mixed copula types. For comparison, the four most commonly used copula
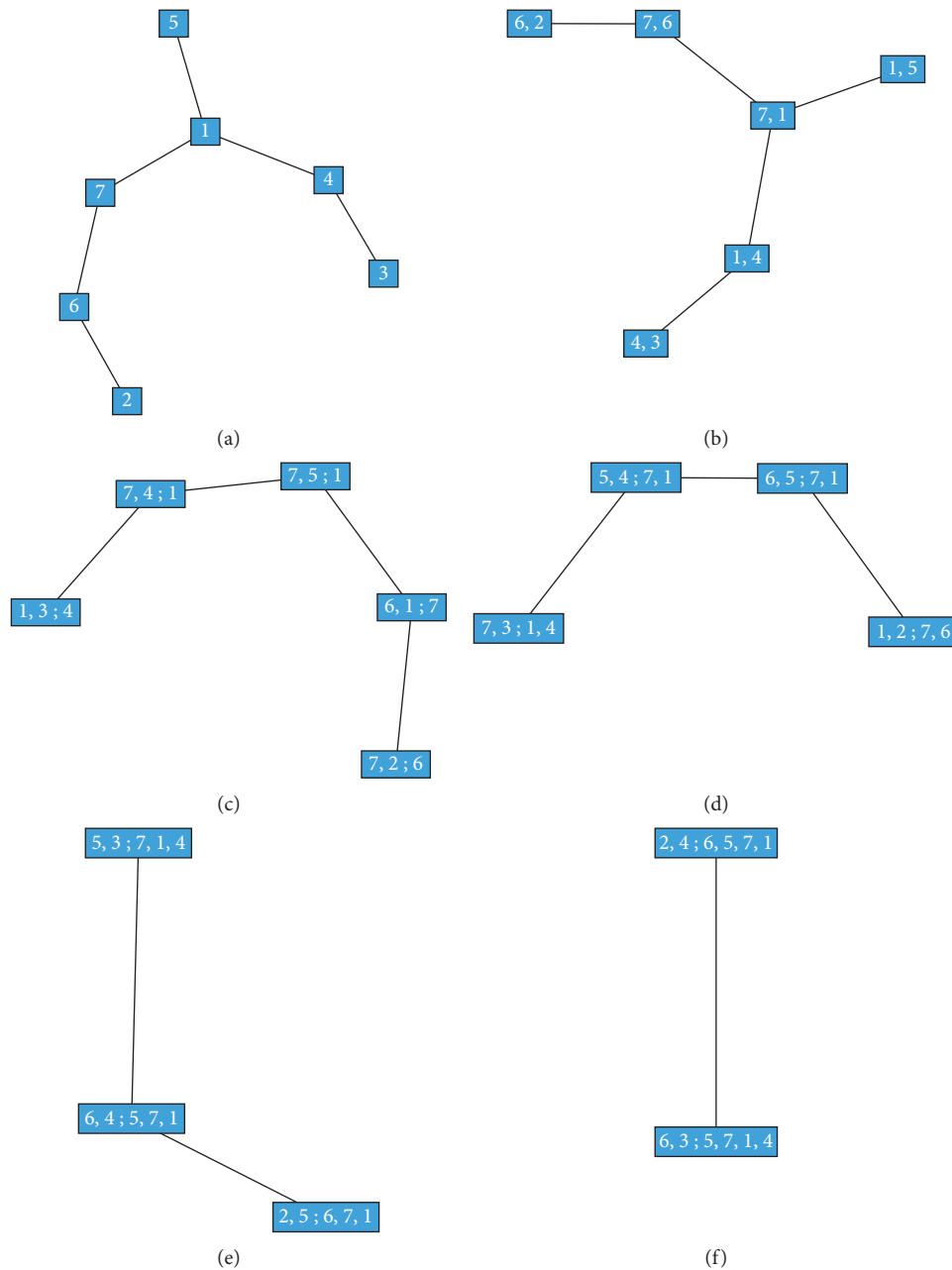
Figure 6: The R-vine tree structure of the Newthyroid data.

types, namely, Frank, Gaussian, Gumbel, and Joe copulas, are fitted to the real datasets as single mixture of R-vine density models. Then, for each dataset, the best-fit model is selected based on the smallest values of the selection criteria. The results are given in Table 6.

Table 6 shows the values of selection criteria and the estimated values of mixture weights. Please note that several starting values were used (the first step) for the EM algorithm. The starting value that returns the largest log likelihood is chosen. After that, the estimation values of the chosen model are again used as starting values for the same model (the second step) in order to double-check for any improvement. For this case, two main improvements were

noticed. First, the number of iterations for the second step is less than half as many iterations as the first. Second, the log-likelihood values barely increased.

From Table 6 and for both datasets, all the selection criteria selected the MRDMMF (the values are shown in bold text). For the SFMRDM, they all showed a poor model fit. However, SFMRDM (Gumbel) and SFMRDM (Joe) were the worst. In these two models, almost half of the pairs of variables are assumed to be independent. This is because the estimates of the copulas parameters (at these pairs) were either 1 (this means independent for these two families) or 0 (out of boundary). This provides evidence that the mixture of one type of copula family is unable to control multiple-
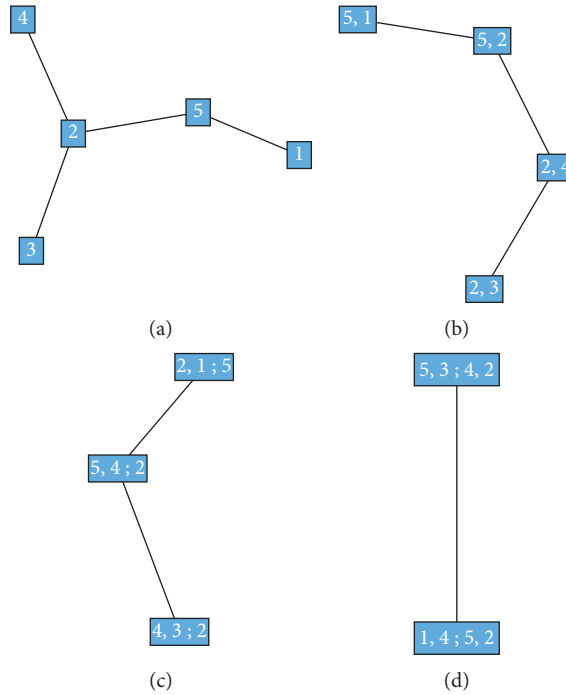
(a)

(b)

(c)

(d)

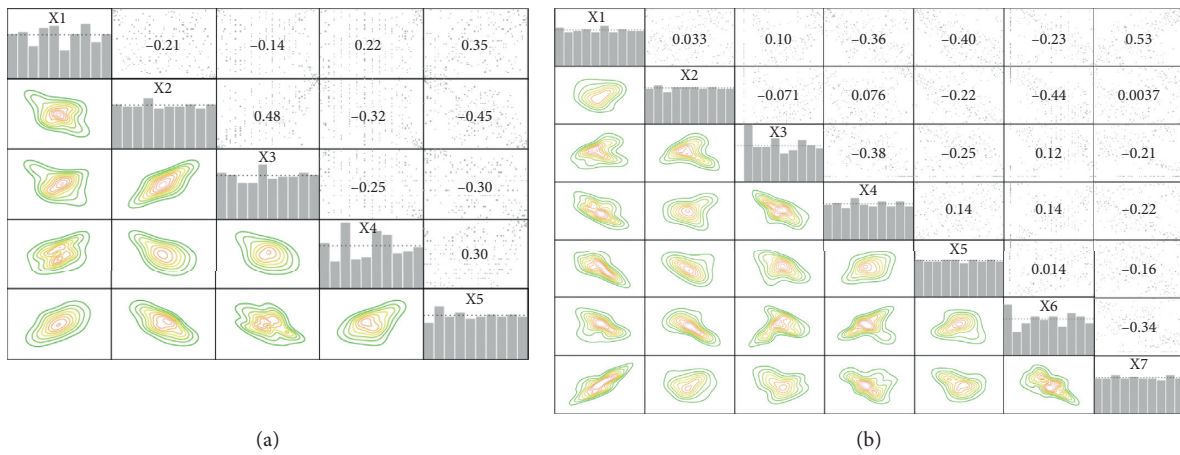FIGURE 7: The R-vine tree structure of the Glass data.

(a)

(b)

FIGURE 8: From left to right: the contour plots (lower triangular) with the scatter plots (upper triangular) and the correlations computed by Kendall's tau (the diagonal) (generated by the pairs.copuladata function from the VineCopula package).

TABLE 6: The values of the selection model criteria and the mixture weights for each model.

Model	AIC	BIC	CAIC	π_1
Newthyriod dataset				
MRDMMF	**−464.5833**	**−390.4292**	**−368.4292**	0.91
SFMRDM (Frank)	−369.6757	−298.8924	−277.8924	0.35
SFMRDM (Gaussian)	−409.8928	−339.1094	−318.1094	0.31
SFMRDM (Gumbel)	−258.6438	−224.9374	−214.0374	0.95
SFMRDM (Joe)	−226.323	−192.6167	−182.6167	0.91
Glass dataset				
MRDMMF	**−1086.1324**	**−907.7352**	**−854.7352**	0.85
SFMRDM (Frank)	−859.3391	−714.6021	−671.6021	0.66
SFMRDM (Gaussian)	−883.504	−738.767	−695.767	0.43
SFMRDM (Gumbel)	−433.4244	−326.7389	−341.7389	0.76
SFMRDM (Joe)	−395.5369	−331.5834	−312.5834	0.78

mixture highly complex dependency structures that may vary from one pair of variables to another. This comparison demonstrates the significant performance of the mixture of R-vine density model with mixed bivariate copula-type single-family mixture of R-vine density model.

5. Conclusion

In this paper, I have introduced a new mixture of R-vine density model with a mixed choice for pair-copula families. This new method provides much more flexibility for modelling complex mixture dependency structures among variables in high-dimensional cases. As with general pair-copula models, the main challenging part of the proposed model is specifying the bivariate copula types. For this, the tree-by-tree scatter plot method was used to construct the mixture components of the model. To illustrate the model performance, the proposed model was fitted to simulated and real datasets. The simulation studies showed well estimation of the true model parameters. In addition, in the simulation and the real application studies, both models showed significant performance of the proposed model over the single-family mixture of R-vine density model. In this study, the number of mixture components is fixed. Fixing the mixture component limits the number and the types of copula families. Hence, estimating the number of mixture components will improve the model fit. This will be done in future work.

Data Availability

All the real data are publicly available through an online database.

Additional Points

This work was conducted in School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

References

[1] A. Sklar, "Fonctions de répartition á n dimensions et leurs marges," *IEEE Transactions on Power Systems*, vol. 8, pp. 229–231, 1959.

[2] Q. Zhang and X. Shi, "A mixture copula bayesian network model for multimodal genomic data," *Cancer Informatics*, vol. 16, 2017.

[3] D. Gunawan, M.-N. Tran, K. Suzuki, J. Dick, and R. Kohn, "Computationally efficient bayesian estimation of high dimensional copulas with discrete and mixed margins," 2016.

[4] L. Hu, "Dependence patterns across financial markets: a mixed copula approach," *Applied Financial Economics*, vol. 16, no. 10, pp. 717–729, 2006.

[5] M. A. Khaled and R. Kohn, "The approximation properties of copulas by mixtures," 2017.

[6] M. Marbac, C. Biernacki, and V. Vandewalle, "Model-based clustering of Gaussian copulas for mixed data," *Communications in Statistics-Theory and Methods*, vol. 46, no. 23, pp. 11635–11656, 2017.

[7] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, 2009.

[8] B. Gräler and E. Pebesma, "The pair-copula construction for spatial data: a new approach to model spatial dependency," *Procedia Environmental Sciences*, vol. 7, pp. 206–211, 2011.

[9] T. M. Erhardt, C. Czado, and U. Schepsmeier, "R-vine models for spatial time series with an application to daily mean temperature," *Biometrics*, vol. 71, no. 2, pp. 323–332, 2015.

[10] A. Min and C. Czado, "Bayesian inference for multivariate copulas using pair-copula constructions," *Journal of Financial Econometrics*, vol. 8, no. 4, pp. 511–546, 2010.

[11] A. Min and C. Czado, "Bayesian model selection for d-vine pair-copula constructions," *Canadian Journal of Statistics*, vol. 39, no. 2, pp. 239–258, 2011.

[12] D. Kim, J.-M. Kim, S.-M. Liao, and Y.-S. Jung, "Mixture of d-vine copulas for modeling dependence," *Computational Statistics & Data Analysis*, vol. 64, pp. 1–19, 2013.

[13] W. Zheng, X. Ren, N. Zhou, D. Jiang, and S. Li, "Mixture of d-vine copulas for chemical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 169, pp. 19–34, 2017.

[14] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Transactions on Power Systems*, vol. 23, 2016.

[15] Y. Qiu, Q. Li, Y. Pan, H. Yang, and W. Chen, "A scenario generation method based on the mixture vine copula and its application in the power system with wind/hydrogen production," *International Journal of Hydrogen Energy*, vol. 44, no. 11, pp. 5162–5170, 2019.

[16] O. O. Evkaya, C. Yozgatlıgil, and A. S. Selcuk-Kestel, "Cd-vine model for capturing complex dependence," *Journal of Applied Statistics*, vol. 10, pp. 1–15, 2020.

[17] J. Dißmann, E. C. Brechmann, C. Czado, and D. Kurowicka, "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics & Data Analysis*, vol. 59, no. 1, pp. 52–69, 2013.

[18] O. Morales Napoles, R. M. Cooke, and D. Kurowicka, "About the Number of Vines and Regular Vines on N Nodes," 2010.

[19] T. M. Erhardt, C. Czado, and U. Schepsmeier, "Spatial composite likelihood inference using local c-vines," *Journal of Multivariate Analysis*, vol. 138, pp. 74–88, 2015.

[20] L. Gruber and C. Czado, "Sequential bayesian model selection of regular vine copulas," *Bayesian Analysis*, vol. 10, no. 4, pp. 937–963, 2015.

[21] L. Gruber and C. Czado, "Bayesian model selection of regular vine copulas," 2015.

[22] N. Zhou and S. Li, "Nonlinear and non-Gaussian process monitoring based on simplified r-vine copula," *Industrial & Engineering Chemistry Research*, vol. 57, no. 22, pp. 7566–7582, 2018.

[23] R. B. Nelsen, *An Introduction to Copulas*, Springer, New York, NY, USA, 2nd edition, 2006.

[24] B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*, Courier Corporation, London, UK, 2011.

[25] H. Joe, *Dependence Modeling with Copulas*, CRC Press, London, UK, 2014.

[26] B. Schweizer and E. F. Wolff, "On nonparametric measures of dependence for random variables," 1981.

[27] P. Embrechts, A. J. McNeil, and D. Straumann, "Correlation and dependence in risk management: properties and pitfalls," *Risk Management*, vol. 1, pp. 176–223, 2002.

[28] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*, John Wiley & Sons, London, UK, 2004.

[29] A. Bárdossy, "Copula-based geostatistical models for groundwater quality parameters," *Water Resources Research*, vol. 42, no. 11, pp. 1–12, 2006.

[30] H. Kazianka and J. Pilz, "Copula-based geostatistical modeling of continuous and discrete data including covariates," *Stochastic Environmental Research and Risk Assessment*, vol. 24, no. 5, pp. 661–673, 2010.

[31] H. Kazianka and J. Pilz, "Bayesian spatial modeling and interpolation using copulas," *Computers & Geosciences*, vol. 37, no. 3, pp. 310–319, 2011.

[32] F. Khan, G. Spöck, and J. Pilz, "A novel approach for modelling pattern and spatial dependence structures between climate variables by combining mixture models with copula models," *International Journal of Climatology*, vol. 40, no. 2, pp. 1049–1066, 2020.

[33] J. Dißmann, *Statistical Inference for Regular Vines and Application*, Technische Universitat Miinchen, London, UK, 2010.

[34] H. Joe, "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters," *Institute of Mathematical Statistics Lecture Notes-Monograph Series*, vol. 23, pp. 120–141, 1996.

[35] T. Bedford and R. M. Cooke, "Probability density decomposition for conditionally dependent random variables modeled by vines," *Annals of Mathematics and Artificial Intelligence*, vol. 32, no. 1/4, pp. 245–268, 2001.

[36] T. Bedford and R. M. Cooke, "Vines: a new graphical model for dependent random variables," *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002.

[37] D. Kurowicka and R. M. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*, John Wiley & Sons, London, UK, 2006.

[38] I. H. Haff, K. Aas, and A. Frigessi, "On the simplified pair-copula construction-simply useful or too simplistic?" *Journal of Multivariate Analysis*, vol. 101, no. 5, pp. 1296–1310, 2010.

[39] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via theEMAlgorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[40] G. Kim, M. J. Silvapulle, and P. Silvapulle, "Comparison of semiparametric and parametric methods for estimating copulas," *Computational Statistics & Data Analysis*, vol. 51, no. 6, pp. 2836–2850, 2007.

[41] C. Genest, K. Ghoudi, and L.-P. Rivest, "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, vol. 82, no. 3, pp. 543–552, 1995.

[42] H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, New York, NY, USA, 1997.

[43] H. Joe, "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, vol. 94, no. 2, pp. 401–419, 2005.

[44] K. Aas and D. Berg, "Models for construction of multivariate dependence–a comparison study," *The European Journal of Finance*, vol. 15, no. 7-8, pp. 639–659, 2009.

[45] H. Akaike, "Information theory and an extension of the likelihood ratio principle," 1973.

[46] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[47] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.

[48] U. Schepsmeier, J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler, and T. Erhardt, "VineCopula: Statistical Inference of Vine Copulas," 2017.

[49] R Development Core Team, "R: A Language and Environment for Statistical Computing," 2008.

[50] J. Alcalá-Fdez, A. Fernández, J. Luengo et al., "Keel datamining software tool: data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, pp. 255–287, 2011.

[51] M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan, "Copula: multivariate dependence with copulas," 2017.

[52] J. Yan, "Enjoy the joy of copulas: with a package copula," *Journal of Statistical Software*, vol. 21, no. 4, pp. 1–21, 2007.

[53] I. Kojadinovic and J. Yan, "Modeling multivariate distributions with continuous margins using the copula R package," *Journal of Statistical Software*, vol. 34, no. 9, pp. 1–20, 2010.

[54] M. Hofert and M. Mächler, "Nested archimedean copulas meet R: the nacopula package," *Journal of Statistical Software*, vol. 39, no. 9, pp. 1–20, 2011.