

## Research Article

# Sequential Truncation of $R$ -Vine Copula Mixture Model for High-Dimensional Datasets

Fadhah Amer Alanazi 

Prince Sultan University, Riyadh, Saudi Arabia

Correspondence should be addressed to Fadhah Amer Alanazi; [fanazi@psu.edu.sa](mailto:fanazi@psu.edu.sa)

Received 12 May 2021; Accepted 22 July 2021; Published 31 July 2021

Academic Editor: Fernando Bobillo

Copyright © 2021 Fadhah Amer Alanazi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Uncovering hidden mixture dependencies among variables has been investigated in the literature using mixture  $R$ -vine copula models. They provide considerable flexibility for modeling multivariate data. As the dimensions increase, the number of the model parameters that need to be estimated is increased dramatically, which comes along with massive computational times and efforts. This situation becomes even much more complex and complicated in the regular vine copula mixture models. Incorporating the truncation method with a mixture of regular vine models will reduce the computation difficulty for the mixture-based models. In this paper, the tree-by-tree estimation mixture model is joined with the truncation method to reduce computational time and the number of parameters that need to be estimated in the mixture vine copula models. A simulation study and real data applications illustrated the performance of the method. In addition, the real data applications show the effect of the mixture components on the truncation level.

## 1. Introduction

Copula is a statistical tool used to model dependencies' structures among variables independently from their margins. Several forms of copula functions exist, which can deal with a wide range of dependency shapes ranging from independent to non-Gaussian distribution. Elliptical copulas are most commonly used multivariate models, due to their ease of computation. Archimedes copula is another famous class of copula functions. These families are able to control a wide class of dependency structures including heavy tails. For example, the Clayton copula function can capture a lower tail dependence, while the Gumbel copula is an upper tail function. For more copula families, interested readers are referred to Nelsen [1] and Joe [2].

Copula has received interesting attention in many applications. For example, Bárdossy [3], Kazianka and Pilz [4] (geostatistic), and Patton [5] (a review of copula models in economics area) [6] used a copula-based multivariate model to analyse the drought in the Northeast Brazil. As each copula family is corresponding to a specific shape of

dependency, the copula imposes the same dependence structure type among all variables, which may have different shapes of dependency. Assuming the same relationship among all variables may not be the case for most of the real life datasets. Gaussian and  $t$ -student copulas are the most commonly used families in high-dimensions' cases, while other families are almost restricted to bivariate cases. Parameters restriction and limited type of the multivariate copula are two main reasons for leading the copula-based model to be inappropriate for modeling high-dimensions' datasets that exhibit multiple dependency types among variables. Even though mixture copula models show significant results comparing to noncopula mixture models (see, for example, [7, 8]), they still suffer from the same limitations as copula-based models. Therefore, pair copula or regular vine copula model has been established in the literature to address the drawbacks of copula models. Pair copulas are hierarchical models, which model only two variables at a time using bivariate copula functions (pair copula). In vine copula models, the type of bivariate copula does not necessarily need to be identical for all pair of

variables. Therefore, multivariate distribution is still valid even if, for each pair of variables, we determine the copula that best fits the data [9, 10]. This forms the main strength of the vine copula models as the dependence shapes may vary from one pair of variables to another. Since 2009, vine copula models have received raising interests in the literature (see, for example, [11–16]).

Although the individual choice for the best fit bivariate copula is one of the main strengths of vine copula models, identifying the type of each bivariate copula can be a very difficult challenge. For this reason, the mixture model has been incorporated with the copula model as a solution for the identification problem of copula models. Mixture models are commonly used to uncover the complex dependency pattern among variables. The Gaussian mixture is one of the most widely used methods in literature. For example, Yuan et al. [17] introduced the Gaussian mixture regression model for quality prediction in multiphase/multimode processes. Madenova and Madani [18] applied Gaussian mixture model-based clustering for partitioning the Fe ore deposit into the geometallurgical clusters with similar properties. The traditional model, such as  $t$ -student and Gaussian mixture models assume that all the mixture components follow the same parametric distribution form, which is almost not the case in the real applications. The fitting Gaussian mixture model for non-Gaussian data, may result in a poor modeling [19]. In addition, mixture copula models suffer from the same limitation of the copula model in high-dimensional cases. Unlike these models, the mixture vine copula does not require all the mixture components to follow the form. Mixture pair-copula models are one of the main solutions for identifying the best fit bivariate copula types for each copula term. Reducing the misspecification of pair-copula types and uncovering complex hidden dependency among variables are the two main advantages of the mixture pair-copula models (see, for example, [19–22]). Unfortunately, besides the identification problem,  $R$ -vine copula models also suffer from the dramatic increase of the number of the model parameters in high-dimension. For  $n$ -dimension  $R$ -vine copula, one need to estimate  $((n * (n - 1))/2)$ , which becomes huge for large datasets. The mixture regular vine copula even increases the difficulties of the pair-copula models, which can be discussed in two points. First, estimating the mixture components for each pair of variables is not straightforward. Second, the number of the parameters to be estimated increases dramatically with the number of the mixture components and the dimension. To overcome the complexity of the model's parameter estimation and, hence, the model complexity, the truncation vine copula was first introduced by Brechmann et al. [23] and Brechmann and Joe [24]. By the truncation method, all the bivariate copulas at the higher level (at the truncated level) are replaced by independent copulas. Hence, the parameters at these levels do not need to be estimated which result in reducing the computation complexity of the model significantly.

In mixture vine-copula models, Roy and Parui [19] have used fixed truncation levels (at the second tree) based on fixed types of mixture pair-copula components. However,

the truncation levels should be estimated as fixing the truncation levels may result in losing some important information among the variables. That is, after the truncation level, all the pair of variables must (almost) show independence structures; otherwise, the model should not be truncated. In the mixture vine copula models, the mixture components affect the truncation levels. This can be shown in the real data application in Section 6. Hence, in truncation models, the modelers try to (hopefully) reduce the estimated tree in the model. Hence, they need to estimate the optimal tree where the model should be truncated. For the mixture models, to the best of my knowledge, estimating the truncation levels using statistical selection methods has not been investigated yet in the literature, which is the main aim of this work.

The rest of the paper is structured as follows: Section 2 briefly discusses the theoretical background of the copula and pair copula. Section 3 introduced the  $R$ -vine copula mixture model and the expectation maximization algorithm (EM), which is the algorithm used to estimate the model parameters. A truncation method is introduced in Section 4. The truncation method with the  $R$ -vine copula mixture model is illustrated with a simulation and real data applications in Sections 5 and 6, respectively.

## 2. Theoretical Background

The aim of this section is to provide a general summary of the theoretical background of copula and pair-copula models. For more details, the interested readers are referred to the given references.

Copula is a multivariate function that couples the margins distribution to their one-dimension standard uniform margin [1].

*Definition 1* (see [25]). Copula is a multivariate distribution function with standard uniform margins, such that

$$C: [0, 1]^n \longrightarrow [0, 1], \quad n \geq 2. \quad (1)$$

**Theorem 1.** Let  $F$  be a  $n$ -dimensional distribution function with marginal distribution  $F_1, \dots, F_n$ . Then, there exists  $n$ -dimensional copula function  $C$  such that,  $\forall x_1, \dots, x_n \in \mathbb{R}^n$ ,

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (2)$$

If  $F_n$  are continuous, then  $C$  is unique.

One main advantage of copula models is that the modelers are able to model the margins independently from the dependency structures, which is captured via the copula function. Another advantage is the ability of copula families to deal with a wide range of dependency forms including non-Gaussian, Gaussian, and heavy tails. However, the copula function imposes the same type of dependency shapes among all the variables, even in the high-dimensions' cases regardless of the strengths of the type of these dependencies. This forms one main limitation of the copula

models. In addition, identifying the form of the copula function that best fits the data is not an easy step as each copula function associates with a specific shape of dependency. Therefore, most copula models are limited to a bivariate case. The multivariate copula is almost limited to Gaussian and *t*-student. However, these families are inadequate to deal with nonelliptical dependency.

In 2009, Aas et al. [9] established a much more promising method, based on the work of Bedford and Cooke [26], Bedford and Cooke [27], Joe [28], and Kurowicka and Cooke [29], to address the problem of copula models in high dimensions. Their method is known as vine copula, pair-copula construction (PCC), and regular vine (*R*-vine) copula. The PCC method builds a multivariate model using only bivariate copula (pair copula). Therefore, only two variables are modeled at a time. Hence, The PCC-based model provides even much more model flexibility and capability than the copula-based model.

**Definition 2** (tree, see [26]).  $\mathcal{T} = \{N, E\}$  is a tree (an acyclic graph) with  $N$  nodes and  $E$  edges (connect each pairs of  $N$ ).

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

**Definition 3** (vine and regular vine, see Ch. 4 in [29]).  $V$  is a vine on  $n$  elements if

- (i)  $V = (T_1, T_2, \dots, T_{n-1})$ , where  $T_1$  indicates the first tree of the vine and so on.
- (ii)  $T_1$  is a connected tree with its nodes  $N_1 = \{1, 2, \dots, n\}$  and edges  $E_1$ .
- (iii) For  $j = 2, 3, \dots, n - 1$ ,  $T_j$  is a connected tree with nodes  $N_j = E_{j-1}$  and edges set  $E_j$ .

In addition,  $V$  becomes a regular vine on  $n$  elements if

- (iv) For  $j = 2, \dots, n - 1$ , 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  $m_j$ ,  $j = 1$  and  $2$ . 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 [29] defined the *D*- and *C*-vine models as follows:

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.

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 [27]).  $(\mathcal{F}, \mathcal{V}, \mathcal{B})$  is a regular vine copula (*R*-vine copula) specification if

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

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

$\mathcal{B} = \{B_e | i = 1, 2, \dots, n - 1; e \in E_i\}$  is a set of the bivariate copula

Let  $\mathbf{x} = (x_1, \dots, x_n)$  be a vector of random variables,  $e = \{l, m\}$  be an edge,  $i = 1, \dots, n$ , and  $D_e$  be a conditioning set of the edge  $e$ . Bedford and Cooke [27] defined a regular vine dependence as follows:

**Definition 5** (regular vine (*R*-vine) dependence). A joint distribution function  $F$  on  $x$  is said to realize 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  $X_{C_{el}}$  and  $X_{C_{em}}$  given  $X_{D_e}$  is a member of the bivariate copula  $B_e(X_{D_e})$ . The marginal distribution of  $X_i$  is  $F_i$ , for  $i = 1, 2, \dots, n$ .

The bivariate copula of  $X_{C_{el}}$  and  $X_{C_{em}}$  given  $X_{D_e}$  is a conditional bivariate copula is assumed to be independent of conditioning variables (see, [9, 30]).

**Theorem 2** (see [31]). 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

$$f_{(1,2,\dots,n)}(\mathbf{x}) = \prod_{j=1}^n f_j(x_j) \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{C_{el}, C_{em} | D_e}(F_{C_{el} | D_e}(x_{C_{el}} | \mathbf{x}_{D_e}), F_{C_{em} | D_e}(x_{C_{em}} | \mathbf{x}_{D_e})), \tag{3}$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $e = \{l, m\}$ ,  $\mathbf{x}_{D_e}$  denotes conditioning variables 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  $F_i$ ,  $i = 1, 2, \dots, n$ . Moreover,  $c_{C_{el}, C_{em} | D_e}$  stands for the density function of bivariate copulas between edge  $e = \{l, m\}$ .

Continuing to the last theorem, let  $e \in E_i$ ,  $e = \{l, m\}$ ,  $l = \{l_1, l_2\}$ , and  $m = \{m_1, m_2\}$  be the edge that joined  $C_{el}$  and  $C_{em}$ . Joe [28] showed that the conditional marginal distribution,  $F_{C_{el} | D_e}(x_{C_{el}} | \mathbf{x}_{D_e})$  and  $F_{C_{em} | D_e}(x_{C_{em}} | \mathbf{x}_{D_e})$ , can be obtained as follows:

$$F_{C_{ei}D_e}(x_{C_{ei}}|\mathbf{x}_{D_e}) = \frac{\partial C_{C_i|D_i}(F_{C_{i1}|D_i}(x_{C_{i1}}|\mathbf{x}_{D_i}), F_{C_{i2}|D_i}(x_{C_{i2}}|\mathbf{x}_{D_i}))}{\partial F_{C_{i2}|D_i}(x_{C_{i2}}|\mathbf{x}_{D_i})} \\ =: h(F_{C_{i1}|D_i}(x_{C_{i1}}|\mathbf{x}_{D_i}), F_{C_{i2}|D_i}(x_{C_{i2}}|\mathbf{x}_{D_i})), \quad (4)$$

where  $F_{C_{ei}D_e}(x_{C_{ei}}|\mathbf{x}_{D_e})$  and  $F_{C_{em}D_e}(x_{C_{em}}|\mathbf{x}_{D_e})$  are then called transformed variables (see, [9] and [31]).

Both PCC and copula models share the same identification problem, which is even much more harder in PCC than copula models. Furthermore, for  $n$ -dimensional  $R$ -vine copula, there are  $((n * (n - 1))/2)$  parameters to be estimated, which becomes huge for high-dimensions' datasets. This number is, however, very large for mixture models. For example, for mixture models, one needs to estimate (for single parameters)  $(2K - 1)((n * (n - 1))/2)$ , where  $K$  is the number of the mixture components. However, the possible estimated parameters of mixture PCC models is  $(3K - 1)((n * (n - 1))/2)$ . Hence, the number of model parameters strongly depends on the number and the type of the mixture components. For example, for the 31-dimensional dataset and for 2 mixture components, one needs to estimate 2790 parameters. This number highly increases with the dimensions and the number of the mixture components. Therefore, model reduction is necessary to reduce the model complexity of the mixture PCC models. This can be achieved by only modeling a limited number of vine trees instead of the full models, where the higher-order trees are set to independent copulas (see, [31]).

### 3. Mixture $R$ -Vine Models and EM Algorithm

Mixture models facilitate modeling complex hidden correlations among variables by fitting a sum of weighted densities' functions to the underlying problem. A finite mixture pair-copula construction combines the benefits of both the mixture and the vine copula models, in order to provide huge flexibility and modeling capabilities for modeling high-dimensions' datasets. By doing so, the mixture pair-copula models allow fitting different mixture

bivariate copulas for each pair of variables. That is, mixture vine copulas may be defined as a building block of the mixture pair copulas.

**3.1. Finite Mixture Model.** Let  $X_1$  and  $X_2$  be two univariate random variables, with  $N$  observations and continuous cdfs  $F_1$  and  $F_2$ , respectively. Then, their probability integral transformation can be given by  $U_1 = F(X_1)$  and  $U_2 = F_2(X_2)$ , respectively. Assume further that the interesting part is in modeling the dependencies' structure between two random variables,  $X_1$  and  $X_2$ , using mixture bivariate copulas. Hence, the density of the mixture bivariate copulas, which model the bivariate dependence structures between  $X_1$  and  $X_2$ , is given by

$$c(u_1, u_2|\Theta) = \sum_{k=1}^K \pi_k c_k(u_1, u_2|\theta_k), \quad (5)$$

where  $\pi_k$  is an unknown parameter (known as a mixture coefficient or weights) of the  $k^{\text{th}}$  ( $k = 1, \dots, K$ ) component which satisfies the following:

$$0 < \pi_k < 1 \\ \sum_{k=1}^K \pi_k = 1$$

$\Theta$  is the set of all model parameters, while  $\theta_k$  is the vector of all the parameters of the  $k^{\text{th}}$  component. In mixture models, expectation maximization 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.

**3.2. EM Algorithm.** The expectation maximization algorithm (EM) [32] is an estimation method with two steps, the so-called expectation step ( $E$ -step) and the maximization step ( $M$ -step). Suppose that a bivariate data sample  $\mathbf{x} = ((x_{11}, x_{21}), \dots, (x_{1n}, x_{2n}))$  of size  $N$ , ( $n = 1, \dots, N$ ), is available. Suppose further that the data is converted to the uniform distribution using the empirical cumulative distribution function. Then, the pseudosample of the copula is given by  $\mathbf{u} = ((u_{11}, u_{21}), \dots, (u_{1n}, u_{2n}))$ . Then, the log-pseudo likelihood function of  $\Theta$  is given as follows:

$$l(\Theta) = \log \left\{ \prod_{n=1}^N \sum_{k=1}^K \pi_k c_k(u_{1n}, u_{2n}|\theta_k) \right\} = \sum_{n=1}^N \log \sum_{k=1}^K \pi_k c_k(u_{1n}, u_{2n}|\theta_k), \quad (6)$$

where  $\Theta$  is the set of all model parameters, while  $\theta_k$  is the set of all the parameters of the  $k^{\text{th}}$  component. Based on the idea of the EM method, the observed data  $\mathbf{u}$  is treated as incomplete information, and hence, the EM algorithm introduces latent variables  $\mathbf{z}_n = (z_{n1}, z_{n2}, \dots, z_{nk})$  where  $z_{nk} = 1$  if the  $n^{\text{th}}$  observation is drawn from the  $k^{\text{th}}$  component 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 unconditionally distributed from the multinomial distribution such that

$$\mathbf{z}_n \sim \mathcal{M}_K(1; \pi_1, \dots, \pi_k). \quad (7)$$

Consequently, we now have the complete data:

$$\mathbf{u}^c = (\mathbf{u}_n, \mathbf{z}_n). \quad (8)$$

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

$$\ell_c(\Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \log \pi_k + \sum_{n=1}^N \sum_{k=1}^K z_{nk} \log_k c_k(u_{1n}, u_{2n} | \theta_k). \tag{9}$$

The EM algorithm starts with initial values of the unknown parameters  $\Theta^{(0)}$ , and the two steps (*E* and *M*) are

repeated until convergence ( $\ell(\Theta)^{(m+1)} - \ell(\Theta)^{(m)}$ ) is smaller than a prespecified tolerance.

*E*-step: calculate the conditional expectation of the complete data log likelihood,  $\ell_c(\Theta)$  in equation (9), given the observed data and using the current estimate of the parameter  $(\Theta)$ .

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

$$\hat{z}_{nk} = E[z_{nk} | -(\mathbf{u}_1, \mathbf{u}_2), \Theta^{(m)}] = P[z_{nk} = 1 | (\mathbf{u}_1, \mathbf{u}_2), \Theta^{(m)}] = \frac{\pi_k^{(m)} c_k(u_{1n}, u_{2n} | \theta_k^{(m)})}{\sum_{s=1}^K \pi_s^{(m)} c_s(u_{1n}, u_{2n} | \theta_s^{(m)})}. \tag{10}$$

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

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

$$\hat{\pi}_k^{(m+1)} = \frac{\sum_{n=1}^N \hat{z}_{nk}^{(m)}}{N}, \tag{11}$$

while the updated of  $\theta_k^{(m+1)}$  can be obtained by maximizing the following equation using the numerical maximization method:

$$\hat{\theta}_k^{(m+1)} = \max_{\theta} \sum_{n=1}^N \sum_{k=1}^K \hat{z}_{nk}^{(m)} \log c_k(u_{1n}, u_{2n} | \theta_k^{(m)}). \tag{12}$$

### 4. Truncation R-Vine Copula Mixture Model

The flexibility of pair-copula models reduces as the dimensions increase. Truncating *R*-vine models is one main solution that plays a key role to address this problem of pair-copula models. Truncating *R*-vine refers to replacing all the pair copula in higher-order trees to independent copulas. The main idea of the truncation *R*-vine copula mixture model can be presented in the following example.

*Example 1* (truncation *R*-vine copula mixture model). Considering 7-dimensional mixture *R*-vine models, with two mixture components of single parameters' bivariate copulas, as shown in Figure 1.

In this example, a mixture of two bivariate single parameter copulas is fitted to each pairs. Therefore, there are 63 parameters to be estimated for the full models. Assume that this model is truncated at tree 3 ( $k = 3$ ). Hence, we will have a 3-truncated *R*-vine copula mixture model. By doing so, the conditional mixture bivariate copulas at trees 4, 5, and 6 are set to independent copulas. Hence, in this case, only 45 parameters need to be estimated instead of 63 as with the full model. That is because, in 3-truncated *R*-vine copula mixture model, there are only 15 edges, while there are 21 edges for

the full *R*-vine copula mixture model. For very high-dimensions' datasets with a large number of mixture components (say 5), the truncations at the first trees will be very reasonable.

*4.1. Methodology.* Brechmann et al. [23] developed the most widely used truncation method, which truncated the *R*-vine models sequentially, using different goodness model fit, including Akaike information criteria (AIC) of Akaike [33] and Bayesian information criteria (BIC) of Schwarz et al. [34]. In this section, the sequential truncation method of the Brechmann et al. [23] (Algorithm 1) (also see, Algorithm 7 in [35]) is incorporating with *R*-vine copula mixture models using well-known selection criteria. In this paper, the AIC, BIC, and consistent Akaike information criteria (CAIC) of Bozdogan [36] are employed. The formulas of these criteria are given as follows:

$$\begin{aligned} \text{AIC} &= -2 \ln L(\hat{\theta}) + 2P, \\ \text{BIC} &= -2 \ln L(\hat{\theta}) + P(\ln(N)), \\ \text{CAIC} &= -2 \ln L(\hat{\theta}) + P(\ln(N) + 1), \end{aligned} \tag{13}$$

where  $\hat{\theta}$  is the estimation values of the parameters,  $N$  is the number of observation of the modeled variables, and  $P$  is the number of the model parameters.

The truncation of *R*-vine copula mixture models can be summarized in the following steps:

- (1) Select specific number of trees, say the first two trees
- (2) Compute the selection criteria of the model
- (3) Add a new tree to the previous model, in order to obtain a new model
- (4) Compute the selection criteria for the new model
- (5) If the new model shows the poor contribution to the previous model, based on the values of the selection criteria, then truncate the *R*-vine copula mixture model at the previous model
- (6) If the new model shows the significant contribution to the previous model, then iterate Steps 3 : 6.

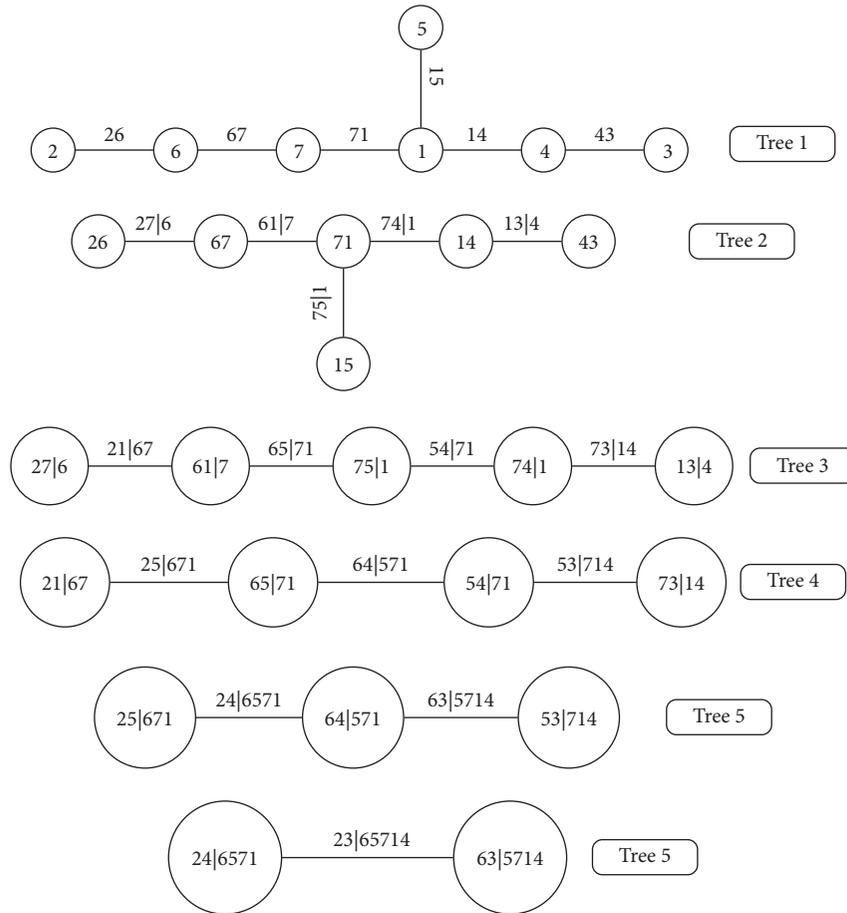


FIGURE 1: 7-dimensional R-vine copula mixture model, 6 trees, 21 edges, and 63 parameters.

**Input:** R-vine tree structures.  
 copula data for  $n$  variables.  
 R-vine dimension:  $n$ .  
 R-vine trees:  $t = 0, \dots, n - 2$ .

**Output:** Truncated R-vine copula mixture model at level  $t$ , or the full R-vine copula mixture model, if there is no possible truncation.

**for**  $t = 0, \dots, n - 2$  **do**  
 Constructed mixture model  $M_{t+1}$  by considering the tree  $t + 1$  and fitting mixture bivariate copula for each pair of variables.  
 Compute BIC for the mixture models  $M_t$  (first model) and mixture model  $M_{t+1}$  (second model).  
**if**  $BIC_1 < BIC_2$  **then**  
     Truncated R-vine copula mixture at level  $t$ .  
**end if**  
**end for**

ALGORITHM 1: Sequential truncation of the R-vine copula mixture model.

For example, consider the R-vine copula mixture model shown in Example 1. At the first step, a small model (only the first two trees) is constructed (first model). Then, the mixture of two components of bivariate copulas is fitted to each pair of variables of this model. Then, the model parameters are estimated. After that, and at the second step,  $BIC_1$  is computed, where  $BIC_1$  refers to the BIC of the first model. Then, a new tree is added to the model. Now, the model is constructed using only three trees (second model). After that,  $BIC_2$  is computed

for the second model. If  $BIC_1 < BIC_2$ , then the model is truncated at the second tree, and the first model is returned. Otherwise, a new model is constructed by adding a new tree, and the steps are iterated until the optimal truncated level is reached.

As mentioned above, the truncation process with mixture dependencies is complex and not straightforward as it is affected by the combination of the bivariate copulas. For example, one type of mixture bivariate copulas may cause the

model to be truncated at one level, while the same model may be truncated at different levels when fitting different mixture components. This potential result is illustrated in Section 6.

### 5. Simulation Study

To illustrate the performance of the sequential mixture truncation method, a simulated data is generated from two-component  $R$ -vine copula mixture model with only two levels (see Figure 2). After that, the true model, three levels, and full two-component 5-dimensional mixture  $R$ -vine models are fitted to the data, respectively. Then, the AIC, BIC, and CAIC are computed for each model. Since the test aims to show the performance of the truncation method, and for a comparison reason, the result of all the fitted  $R$ -vine copula mixture models is reported.

Before reporting the final results, the idea of the simulation study is represented in much more detail using a graph representation. Consider 5-dimensional, two components, and  $R$ -vine copula mixture model. Figures 2–4 present 3 different  $R$ -vine copula mixture models. These models are full  $R$ -vine copula mixture model, 3-levels, and 2-levels truncated  $R$ -vine copula mixture models, respectively. The main difference between these models is the number of trees to be modeled. For example, for the full  $R$ -vine copula mixture models, there are 4 trees, and one needs to estimate the whole model. However, in the case of the truncation models, the conditional mixture bivariate copulas at levels 2 and 3 are replaced by independent copulas ( $\Pi_c$ ). Hence, instead of modeling the whole model, one only needs to estimate the mixture bivariate copulas up to the truncation levels. For very large datasets, say 100 dimensions, this will result in a very huge reduction of the model complexity and the parameters that need to be estimated.

Tables 1–3 summarise the information of three fitted models. The summary includes the mixture type of the bivariate copulas (at each pair) and mixture weights, while Tables 4–6 report the result of the three models.

From Tables 4 and 5, the estimation values of the model parameters at the first trees of all the models are very close to the true values. Hence, the dependencies' structures are described well, and the performance of the EM algorithm is satisfied. In addition, for the 3-levels truncated model, the corresponding parameters of the mixture bivariate copulas at trees 3 and 4 are very close to the independent boundary of each bivariate copula. For example, at tree 3 of 3-levels truncated mixture  $R$ -vine models, the parameters of Frank and Gaussian copulas are  $-0.5$  and  $0.091$ , respectively. In addition, the corresponding Kendall's tau value of these copulas is  $-0.040$  and  $0.061$ , which are very small, indicating that the corresponding variables are almost independent. Again, this illustrates the performance of the EM algorithm to accurately estimate the model's parameters.

After estimating the model parameters and testing the model performance, the three model selection criteria are computed for each model, in order to illustrate the ability of the truncation method to select the most optimal truncation level of the mixture  $R$ -vine models. The values of the selection criteria are shown in Table 7.

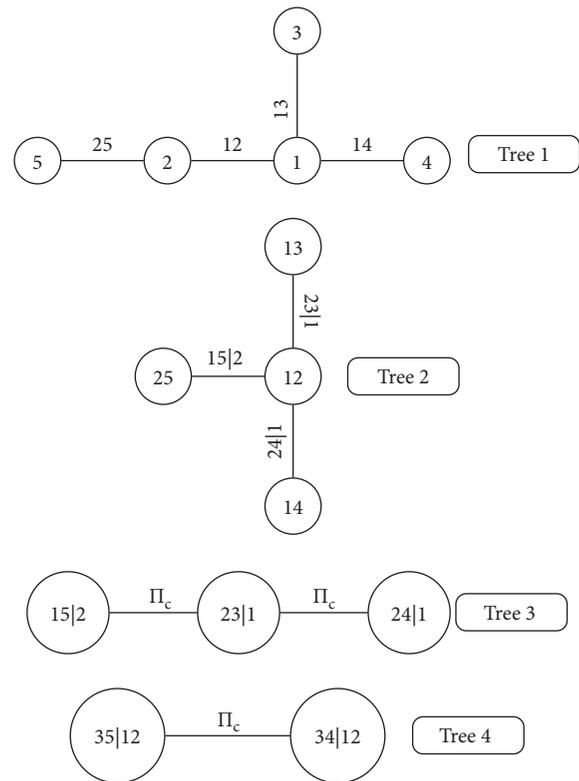


FIGURE 2: 2-levels truncated  $R$ -vine copula mixture model. 5-dimensional, 4 trees, 7 edges (nonindependent mixture of bivariate copulas and 3 edges (independent bivariate copulas)), and 21 parameters.

From Table 7, the truncated  $R$ -vine copula mixture model at level 2 shows the best model fits, while the full model shows the worst model fits. In addition, all the selection criteria selected the true model (the model from where the simulated data has been generated). Comparing the selection methods' values of the truncation  $R$ -vine copula mixture model with the 3-levels truncated model, one can clearly see that the model is truncated correctly. That is, let  $AIC_1, BIC_1,$  and  $CAIC_1$  correspond to the 2-levels truncated  $R$ -vine copula mixture model and  $AIC_2, BIC_2,$  and  $CAIC_2$  are corresponding to the 3-levels truncated model. Then, from the table,  $AIC_1 < AIC_2, BIC_1 < BIC_2,$  and  $CAIC_1 < CAIC_2$ . The same result holds when comparing the true model with the full one. Therefore, the result can be interpreted as evidence of the ability of the truncation method to select the most optimal truncation level of the  $R$ -vine copula mixture model. Hence, the performance of the truncation method with  $R$ -vine copula mixture is illustrated.

### 6. Real Data Applications

This section aims to demonstrate the performance of the sequential truncation method of  $R$ -vine copula mixture models when applied to real datasets. For this reason, two high-dimensional real datasets are tested, namely, Vowel and Ionosphere datasets, which were obtained from the repository [37]. They consist of 990 and 351 observations, respectively. As the

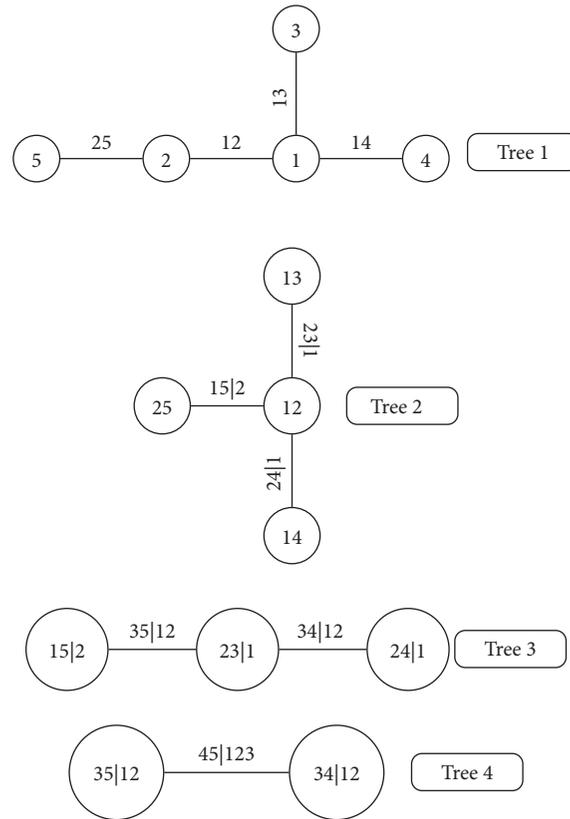


FIGURE 3: Full *R*-vine copula mixture model. 5-dimensional, 4 trees, 10 edges, and 30 parameters.

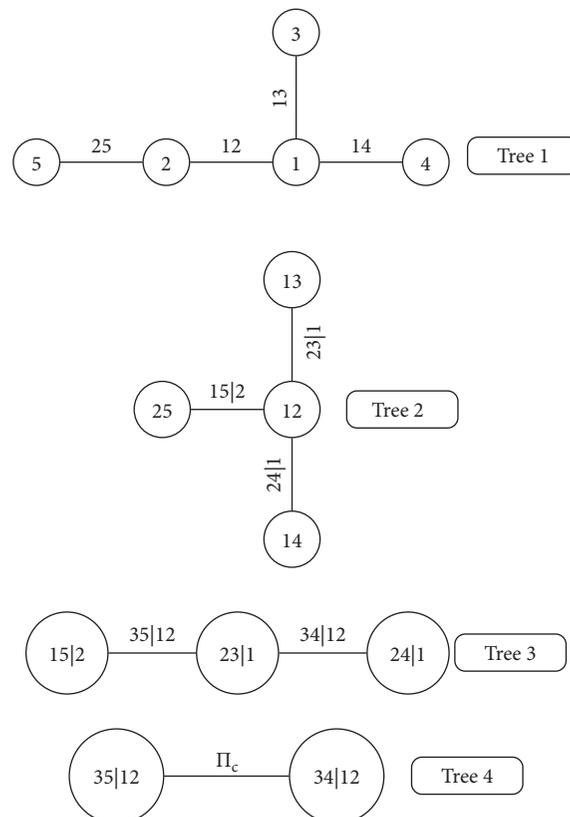


FIGURE 4: 3-levels truncated *R*-vine copula mixture model. 5-dimensional, 4 trees, 9 edges (nonindependent mixture of bivariate copulas and one edge (independent bivariate copulas)), and 27 parameters.

TABLE 1: Summary of the two-component full *R*-vine copula mixture model.

	Edge	w	Type (1st)	$\theta$ ( $\tau$ )	Type (2nd)	$\theta$ ( $\tau$ )
1st tree	2, 5	0.2	C	3 (0.6)	Rot.C90	-4 (-0.67)
	1, 4	0.3	F	4 (0.39)	G	3 (0.67)
	1, 3	0.8	J	5 (0.68)	F	-6 (-0.51)
	1, 2	0.3	F	7 (0.56)	Sur.G	-5 (-0.8)
2nd tree	1, 5; 2	0.6	G	5 (0.8)	Ga	0.5 (0.33)
	2, 4; 1	0.7	F	4 (0.39)	J	3 (0.52)
	2, 3; 1	0.7	Ga	0.3 (0.19)	G	3 (0.67)
3rd tree	3, 5; 1, 2	0.5	F	0.6 (0.07)	Ga	0.8 (0.59)
	3, 4; 2, 1	0.3	Ga	0.5 (0.33)	Ga	0.4 (0.26)
4th tree	4, 5; 3, 1, 2	0.7	G	3 (0.67)	C	1.9 (0.49)

TABLE 2: Summary of 3-levels truncated *R*-vine copula mixture models.

	Edge	w	Type (1st)	$\theta$ ( $\tau$ )	Type (2nd)	$\theta$ ( $\tau$ )
1st tree	2, 5	0.2	C	3 (0.6)	Rot.C90	-4 (-0.67)
	1, 4	0.3	F	4 (0.39)	G	3 (0.67)
	1, 3	0.8	J	5 (0.68)	F	-6 (-0.51)
	1, 2	0.3	F	7 (0.56)	Sur.G	-5 (-0.8)
2nd tree	1, 5; 2	0.6	G	5 (0.8)	Ga	0.5 (0.33)
	2, 4; 1	0.7	F	4 (0.39)	J	3 (0.52)
	2, 3; 1	0.7	Ga	0.3 (0.19)	G	3 (0.67)
3rd tree	3, 5; 1, 2	0.5	F	0.6 (0.07)	Ga	0.8 (0.59)
	3, 4; 2, 1	0.3	Ga	0.5 (0.33)	Ga	0.4 (0.26)
4th tree	4, 5; 3, 1, 2	—	$\prod_c$	0 (0)	—	—

TABLE 3: Summary of 2-levels truncated *R*-vine copula mixture models.

	Edge	w	Type (1st)	$\theta$ ( $\tau$ )	Type (2nd)	$\theta$ ( $\tau$ )
1st tree	2, 5	0.2	C	3 (0.6)	Rot.C90	-4 (-0.67)
	1, 4	0.3	F	4 (0.39)	G	3 (0.67)
	1, 3	0.8	J	5 (0.68)	F	-6 (-0.51)
	1, 2	0.3	F	7 (0.56)	Sur.G	-5 (-0.8)
2nd tree	1, 5; 2	0.6	G	5 (0.8)	Ga	0.5 (0.33)
	2, 4; 1	0.7	F	4 (0.39)	J	3 (0.52)
	2, 3; 1	0.7	Ga	0.3 (0.19)	G	3 (0.67)
3rd tree	3, 5; 1, 2	—	$\prod_c$	0 (0)	—	—
	3, 4; 2, 1	—	$\prod_c$	0 (0)	—	—
4th tree	4, 5; 3, 1, 2	—	$\prod_c$	0 (0)	—	—

TABLE 4: Summary of the two-component full *R*-vine copula mixture model.

Edge	w (SD)	Type (1st)	$\theta$ ( $\tau$ ) (SD)	Type (2nd)	$\theta$ ( $\tau$ ) (SD)
1st tree					
2, 5	0.202 (0.015)	C	3.082 (0.605) (0.356)	Rot.C90	-4.032 (-0.668) (0.039)
1, 4	0.301 (0.063)	F	4.056 (0.390) (0.594)	G	3.004 (0.666) (0.464)
1, 3	0.802 (0.014)	J	5.005 (0.677) (0.183)	F	-6.109 (-0.517) (0.436)
1, 2	0.301 (0.014)	F	7.085 (0.565) (0.575)	Sur.G	-5.006 (-0.8) (0.154)
2nd tree					
1, 5; 2	0.600 (0.030)	G	4.999 (0.799) (0.309)	Ga	0.497 (0.332) (0.039)
2, 4; 1	0.688 (0.062)	F	4.151 (0.398) (0.474)	J	2.912 (0.500) (0.464)
2, 3; 1	0.689 (0.064)	Ga	0.295 (0.191) (0.042)	G	3.006 (0.661) (0.436)
3rd tree					
3, 5; 1, 2	0.592 (0.238)	F	-0.501 (-0.040) (2.169)	Ga	0.091 (0.061) (0.209)
3, 4; 2, 1	0.338 (0.244)	Ga	0.086 (0.058) (0.180)	Ga	-0.076 (-0.051) (0.166)
4th tree					
4, 5; 3, 1, 2	0.518 (0.148)	G	0.985 (0.005) (0.142)	C	0.031 (0.014) (0.098)

TABLE 5: Summary of 3-levels truncated *R*-vine copula mixture model.

Edge	w (SD)	Type (1st)	$\theta$ ( $\tau$ ) (SD)	Type (2nd)	$\theta$ ( $\tau$ ) (SD)
1st tree					
2, 5	0.202 (0.015)	C	3.082 (0.605) (0.356)	Rot.C90	-4.032 (-0.668) (0.039)
1, 4	0.301 (0.063)	F	4.056 (0.390) (0.594)	G	3.004 (0.666) (0.464)
1, 3	0.802 (0.014)	J	5.005 (0.677) (0.183)	F	-6.109 (-0.517) (0.436)
1, 2	0.301 (0.014)	F	7.085 (0.565) (0.575)	Sur.G	-5.006 (-0.8) (0.154)
2nd tree					
1, 5; 2	0.600 (0.030)	G	4.999 (0.799) (0.309)	Ga	0.497 (0.332) (0.039)
2, 4; 1	0.688 (0.062)	F	4.151 (0.398) (0.474)	J	2.912 (0.500) (0.464)
2, 3; 1	0.689 (0.064)	Ga	0.295 (0.191) (0.042)	G	3.006 (0.661) (0.436)
3rd tree					
3, 5; 1, 2	0.592 (0.238)	F	-0.501 (-0.040) (2.169)	Ga	0.091 (0.061) (0.209)
3, 4; 2, 1	0.338 (0.244)	Ga	0.086 (0.058) (0.180)	Ga	-0.076 (-0.051) (0.166)
4th tree					
4, 5; 3, 1, 2	—	$\prod_c$	0 (0)	—	—

TABLE 6: Summary of 2-levels truncated *R*-vine copula mixture models.

Edge	w (SD)	Type (1st)	$\theta$ ( $\tau$ ) (SD)	Type (2nd)	$\theta$ ( $\tau$ ) (SD)
1st tree					
2, 5	0.202 (0.015)	C	3.082 (0.605) (0.356)	Rot.C90	-4.032 (-0.668) (0.039)
1, 4	0.301 (0.063)	F	4.056 (0.390) (0.594)	G	3.004 (0.666) (0.464)
1, 3	0.802 (0.014)	J	5.005 (0.677) (0.183)	F	-6.109 (-0.517) (0.436)
1, 2	0.301 (0.014)	F	7.085 (0.565) (0.575)	Sur.G	-5.006 (-0.8) (0.154)
2nd tree					
1, 5; 2	0.600 (0.030)	G	4.999 (0.799) (0.309)	Ga	0.497 (0.332) (0.039)
2, 4; 1	0.688 (0.062)	F	4.151 (0.398) (0.474)	J	2.912 (0.500) (0.464)
2, 3; 1	0.689 (0.064)	Ga	0.295 (0.191) (0.042)	G	3.006 (0.661) (0.436)
3rd tree					
3, 5; 1, 2	—	$\prod_c$	0 (0)	—	—
3, 4; 2, 1	—	$\prod_c$	0 (0)	—	—
4th tree					
4, 5; 3, 1, 2	—	$\prod_c$	0 (0)	—	—

TABLE 7: Selection criteria values of the traditional, fixed, and new proposed models.

Simulated dataset with 1500 observations			
Model	AIC	BIC	CAIC
Truncated MRVM at level 2	<b>-9103.53</b>	<b>-8991.952</b>	<b>-8970.952</b>
Truncated MRVM at level 3	-9095.752	-8953.039	-8926.179
Full MRVM	-9090.656	-8932.162	-8902.332

TABLE 8: Selection criteria values of the second scenario of the simulation study.

Model	No. of mixture components	Component type
<i>Vowel Dataset</i>		
1	3	Ga, Rot.C (90), and Rot.G (90)
2	4	F, Ga, Rot.C (90), and Rot.G (90)
3	5	C, Rot.C (90), Rot.G (90), G, and Rot.J (180)
<i>Ionosphere Dataset</i>		
1	4	C, Rot.J (90), Rot.G (180), and Rot.C (180)
2	4	C, Rot.J (270), G, and F

TABLE 9: Selection criteria values of the second scenario of the simulation study.

		Vowel dataset								
Model	Criteria	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>
1	AIC	-2251.271	-2627.365	-2893.095	-3175.234	-3347.374	-3571.632	-3680.863	-3731.910	-3754.592
	BIC	-2119.033	-2367.787	-2550.256	-2729.543	-2808.627	-2974.112	-3053.956	-3080.515	-3093.401
	CAIC	-2092.033	-2314.787	-2480.256	-2638.543	-2698.627	-2852.112	-2925.956	-2947.515	-2958.401
2	AIC	-2248.314	-2636.790	-2913.117	-3212.841	-3429.796	-3683.609	-3787.775	-3850.875	-3883.757
	BIC	-2101.383	-2333.132	-2521.301	-2713.275	-2827.378	-3022.419	-3087.403	-3106.424	-3124.613
	CAIC	-2071.383	-2271.132	-2441.301	-2611.275	-2704.378	-2887.419	-2944.403	-2954.424	-2969.613
3	AIC	-2195.317	-2567.953	-2783.818	-3035.184	-3259.367	-3471.338	-3529.098	-3564.625	-3572.847
	BIC	-2063.079	-2283.886	-2343.025	-2471.948	-2607.973	-2746.478	<b>-2760.159</b>	-2736.913	-2730.442
	CAIC	-2036.079	-2225.886	-2253.025	-2356.948	-2474.973	-2598.478	<b>-2603.159</b>	-2567.913	-2558.442

aim of this paper is to incorporate the truncation method with *R*-vine copula mixture models, the focus will be on fixed *R*-vine copula mixture models, in order to avoid extra complexity and model computation. For each dataset, different fixed *R*-vine copula mixture models are used.

Before illustrating the performance of the truncation method on the *R*-vine copula mixture models, full information of the fitted mixture bivariate copulas for each dataset of each model is given in Table 8, where Gaussian (Ga), Rotated Clayton 90 degree (Rot.C (90)), Rotated Gumbel 90 degree (Rot.G (90)), Frank (F), Rotated Joe 180 degree, Rotated Gumbel 180 degree (Rot.G (180)), and Rotated Joe 270 degree (Rot.J (270)) stand for the fitted bivariate copulas and their short names.

The dimensions of these datasets are 10 and 32, respectively. Hence, there are two different full *R*-vine copula mixture models, one with 10-dimensional, 9 trees, and 45 edges, while the second one is 32-dimensional *R*-vine copula mixture with 31 trees and 496 edges. For these models, and unlike nonmixture *R*-vine model, the number of the parameters to be estimated strongly depends on the type and the number of the mixture components. For example, for 4-mixture components of single parameter bivariate copulas,

the second model will contain 3,472 parameters. One can imagine how much the significant reduction of the model complexity will be obtained if the truncation level can be reached at the first levels. Another important point, as mentioned above, is the influences of the mixture components on the truncation levels. These two points are illustrated in Tables 9 and 10.

From Tables 9 and 10, the two main points, mentioned above, are illustrated. First, from Table 9, the truncation level is strongly influenced by the type of the mixture components. For first and second mixture models, there is no possible truncation level, while the third mixture model is truncated at level 7. Hence, the truncation level should not be fixed and need to be estimated, in order to avoid ignoring any possible information. Furthermore, for the third mixture model, and by truncation method, there are 27 parameters that not need to be estimated in comparison with the full model (third model without the truncation level). For the second dataset, both mixture models are truncated at the third levels. Therefore, there are only 609 parameters to be estimated out of 3,472, which provides a very significant reduction of the model computation complexity and effort, which illustrates the second point mentioned above.

TABLE 10: Selection criteria values of the second scenario of the simulation study.

Model	Criteria	Ionosphere dataset										
		T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>31</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>	T <sub>10</sub>	T <sub>final</sub>
1	AIC	-13280.45	-14894.52	-15720.22	-16126.05	-16293.46	-16503.84	-16568.63	-16665.10	-16668.79	-16661.21	-15265.95
	BIC	-12535.317	-13485.335	<b>-13654.703</b>	-13535.459	-13181.666	-12909.453	-12480.060	-12051.462	-11615.025	-11205.917	-5142.970
	CAIC	-12342.317	-13120.335	<b>-13119.703</b>	-12864.459	-12375.666	-11978.453	-11421.060	-10856.462	-10306.025	-9792.917	-2520.970
2	AIC	-13341.64	-14911.16	-16025.68	-16415.01	-16742.75	-16985.35	-17086.39	-17306.91	<b>-17370.22</b>	-17336.46	-16459.97
	BIC	-12627.390	-13494.256	<b>-14056.684</b>	-13893.917	-13700.452	-13437.283	-13048.005	-12832.254	-12517.212	-12047.179	-7047.376
	CAIC	-12442.390	-13127.256	<b>-13546.684</b>	-13240.917	-12912.452	-12518.283	-12002.005	-11673.254	-11260.212	-10677.179	-4609.376

## 7. Conclusion

Modeling only two variables at a time using (mixture) bivariate copulas is one of the main benefits of (mixture) pair-copula models. However, this flexibility is reduced with the dimensions, due to the large number of the model parameters to be estimated. In this paper, the truncation method was incorporated with mixture  $R$ -vine models. Estimating the truncation levels for the mixture  $R$ -vine model is not a straightforward approach as the effect of the mixture components on the result of the models. The performance of the truncation method with the EM algorithm was illustrated. The simulation study showed the ability of the model to accurately estimate the truncation level and the model parameters. The real data study showed the significant reduction of the model computation. In addition, from the real study, the effect of the mixture components on the truncation level was illustrated.

The remaining questions are how would estimating the mixture components, for each pair of variables, affect the optimal truncation level? In addition, how could ordering the variables, based on the mixture components, provide a new way to estimate the mixture components of each pair of variables and how would it affect the truncation level? These questions are left as future works.

## Data Availability

The datasets used to support the findings of this study have been deposited in the Keel repository (<https://sci2s.ugr.es/keel/development.php>).

## Disclosure

The author acknowledges that the manuscript has been submitted as a preprint in Research Square in the below link: <https://www.preprints.org/manuscript/202102.0458/v1>. This work was conducted in School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] R. B. Nelsen, *An Introduction to Copulas*, Springer, New York, NY, USA, 2nd edition, 2006.
- [2] H. Joe, *Dependence Modeling with Copulas*, CRC Press, Boca Raton, FL, USA, 2014.
- [3] A. Bárdossy, "Copula-based geostatistical models for groundwater quality parameters," *Water Resources Research*, vol. 42, no. 11, pp. 1–12, 2006.
- [4] H. Kazianka and J. Pilz, "geoENV VII – geostatistics for environmental applications," in *Spatial Interpolation Using Copula-Based Geostatistical Model* Springer, Berlin, Germany, 2010.
- [5] A. J. Patton, "A review of copula models for economic time series," *Journal of Multivariate Analysis*, vol. 110, pp. 4–18, 2012.
- [6] J. D. Pontes Filho, F. D. A. Souza Filho, E. S. P. R. Martins, and T. M. d. C. Studart, "Copula-based multivariate frequency analysis of the 2012–2018 drought in Northeast Brazil," *Water*, vol. 12, no. 3, p. 834, 2020.
- [7] L. Hu, "Dependence patterns across financial markets: a mixed copula approach," *Applied Financial Economics*, vol. 16, no. 10, pp. 717–729, 2006.
- [8] Q. Zhang and X. Shi, "A mixture copula bayesian network model for multimodal genomic data," *Cancer Informatics*, vol. 16, 2017.
- [9] 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.
- [10] 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.
- [11] C. Czado, "Pair-copula constructions of multivariate copulas," in *Copula Theory and Its Applications* Springer, Berlin, Germany, 2010.
- [12] C. Czado, E. C. Brechmann, and L. Gruber, "Selection of vine copulas," in *Copulae in Mathematical and Quantitative Finance* Springer, Berlin, Germany, 2013.
- [13] 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.
- [14] L. Gruber and C. Czado, "Sequential bayesian model selection of regular vine copulas," *Bayesian Analysis*, vol. 10, no. 4, pp. 937–963, 2015.
- [15] R. Yu, R. Yang, C. Zhang, M. Špoljar, N. Kuczyńska-Kippen, and G. Sang, "A vine copula-based modeling for identification of multivariate water pollution risk in an interconnected river system network," *Water*, vol. 12, no. 10, p. 2741, 2020.
- [16] C. Trucíos, A. K. Tiwari, and F. Alqahtani, "Value-at-risk and expected shortfall in cryptocurrencies' portfolio: a vine copula-based approach," *Applied Economics*, vol. 52, no. 24, pp. 2580–2593, 2020.
- [17] X. Yuan, Z. Ge, and Z. Song, "Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 138, pp. 97–109, 2014.
- [18] Y. Madenova and N. Madani, "Application of Gaussian mixture model and geostatistical co-simulation for resource modeling of geometallurgical variables," *Natural Resources Research*, vol. 30, no. 2, pp. 1199–1228, 2021.
- [19] A. Roy and S. K. Parui, "Pair-copula based mixture models and their application in clustering," *Pattern Recognition*, vol. 47, no. 4, pp. 1689–1697, 2014.
- [20] 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.
- [21] G. N. F. Weiß and M. Scheffer, "Mixture pair-copula-constructions," *Journal of Banking & Finance*, vol. 54, pp. 175–191, 2015.
- [22] F. A. Alanazi, "A mixture of regular vines for multiple dependencies," *Journal of Probability and Statistics*, vol. 2021, Article ID 5559518, 15 pages, 2021.
- [23] E. C. Brechmann, C. Czado, and K. Aas, "Truncated regular vines in high dimensions with application to financial data," *Canadian Journal of Statistics*, vol. 40, no. 1, pp. 68–85, 2012.
- [24] E. C. Brechmann and H. Joe, "Truncation of vine copulas using fit indices," *Journal of Multivariate Analysis*, vol. 138, pp. 19–33, 2015.
- [25] B. Schweizer and S. Abe, *Probabilistic Metric Spaces*, Courier Corporation, North Chelmsford, MA, USA, 2011.

- [26] 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.
- [27] 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.
- [28] H. Joe, "Families of  $m$ -variate distributions with given margins and  $m(m-1)/2$  bivariate dependence parameters," in *Institute of Mathematical Statistics Lecture Notes-Monograph Series* Institute of Mathematical Statistics, Hayward, CA, USA, 1996.
- [29] D. Kurowicka and R. M. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*, John Wiley & Sons, Hoboken, NJ, USA, 2006.
- [30] I. Hobæk 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.
- [31] 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, pp. 52–69, 2013.
- [32] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [33] H. Akaike, "Information theory and an extension of the likelihood ratio principle," in *Proceedings of the Second International Symposium of Information Theory*, Budapest, Hungary, September 1973.
- [34] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [35] E. Brechmann, "Truncated and simplified regular vines and their applications," Diploma Thesis, Faculty of Mathematics, Technische Universität München, München, Germany, 2010.
- [36] 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.
- [37] J. Alcalá-Fdez, A. Fernández, J. Luengo et al., "Keel data-mining 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.