Constructing a multivariate distribution function with a vine copula: towards multivariate luminosity and mass functions

Tsutomu T. Takeuchi [1,2]⋆ and Kai T. Kono[1]

[1]*Division of Particle and Astrophysical Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan*
[2]*The Research Center for Statistical Machine Learning, the Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

ABSTRACT

The need for a method to construct multidimensional distribution function is increasing recently, in the era of huge multiwavelength surveys. We have proposed a systematic method to build a bivariate luminosity or mass function of galaxies by using a copula. It allows us to construct a distribution function when only its marginal distributions are known, and we have to estimate the dependence structure from data. A typical example is the situation that we have univariate luminosity functions at some wavelengths for a survey, but the joint distribution is unknown. Main limitation of the copula method is that it is not easy to extend a joint function to higher dimensions ($d > 2$), except some special cases like multidimensional Gaussian. Even if we find such a multivariate analytic function in some fortunate case, it would often be inflexible and impractical. In this work, we show a systematic method to extend the copula method to unlimitedly higher dimensions by a vine copula. This is based on the pair-copula decomposition of a general multivariate distribution. We show how the vine copula construction is flexible and extendable. We also present an example of the construction of a stellar mass–atomic gas–molecular gas three-dimensional mass function. We demonstrate the maximum likelihood estimation of the best functional form for this function, as well as a proper model selection via vine copula.

Key words: ISM: atoms – ISM: molecules – galaxies: evolution – galaxies: formation – galaxies: luminosity function, mass function – galaxies: star formation.

1 INTRODUCTION

Galaxies evolve in various aspects. Individual galaxies change their physical properties through merging of their host dark matter haloes, merging of galaxies, star formation, chemical evolution, infall of matter from the large-scale structure, etc. This aspect of galaxy evolution is, say, a life history of galaxies. The collection of the individual life history of galaxies, combined with the evolving cosmological condition, drives the collective evolution of galaxies. To describe this 'sociological' galaxy evolution in a statistical sense, the luminosity function (LF) and/or mass function (MF) of galaxies play a fundamental role (e.g. Binggeli, Sandage & Tammann 1988; Takeuchi, Yoshikawa, & Ishii 2000; Blanton et al. 2001; de Lapparent et al. 2003; Willmer et al. 2006; Johnston 2011; Moffett et al. 2016; Koprowski et al. 2017; Lake et al. 2017; López-Sanjuan et al. 2017; Wright et al. 2017; Bhatawdekar et al. 2019). Even though the LF (MF) is a result of highly complicated and entangled physical processes, still it is the first statistic to be examined from observations.

Now, studies of galaxy evolution is facing the time for drastic change by multiband large surveys. Indeed, all of the modern large surveys are performed at multiband. Connecting the LFs (MFs) obtained at different wavelengths is expected to provide us with a new insight to the fundamental physics to drive galaxy evolution (e.g. Mashian, Oesch & Loeb 2016; Vallini et al. 2016; Caplar, Lilly & Trakhtenbrot 2018; Dutta, Khandai & Dey 2020, and references therein). However, it is not easy to determine the corresponding multivariate function from its marginal distributions, if the distribution is not multivariate Gaussian. As widely known, galaxy LFs are fairly well described by the Schechter function (Schechter 1976) (stellar and gas components) or double-power-law-type function (e.g. Saunders et al. 2000; Takeuchi, Yoshikawa, & Ishii 2003) (dust, radio continuum, and X-ray emission), both of which are far different from the Gaussian distribution. In such a case, there exist infinitely many distributions with the same marginals even if the correlation structure is specified. In astronomical applications, a multivariate distribution has been constructed based on a primary-selection wavelength (e.g. Chołoniewski 1985; Mobasher, Sharples & Ellis 1993; Chapman et al. 2003; Schafer 2007; Calette et al. 2018; Rodríguez-Puebla et al. 2020). A thorough and comprehensive discussion on this method is found in, for example, Rodríguez-Puebla et al. (2020). Though these works are well designed in their own purposes, we often want to have a multivariate PDF estimation method without a specific primary selection in modern astrophysical analysis. Thus, a general method to construct a multivariate distribution function with pre-defined marginal distributions and dependence structure has long been desired.

⋆ E-mail: takeuchi.tsutomu@g.mbox.nagoya-u.ac.jp

Such a function has been commonly used to analyse two covariate random variables, particularly extensively in econometrics and mathematical finance. This is the so-called 'copula'.[1] In a bivariate context, copulae are obviously useful to define non-parametric measures of dependence for pairs of random variables (e.g. Johnson & Kotz 1977; Trivedi & Zimmer 2005; for a recent review, see Lin et al. 2014). In astrophysics, however, copulae started to attract researchers' attention relatively recently (e.g. Benabed et al. 2009; Jiang et al. 2009; Koen 2009; Scherrer et al. 2010; Takeuchi 2010). After a decade since then, the copula method is getting gradually known to the astronomical community: now it is applied to bivariate luminosity function of galaxies (e.g. Takeuchi et al. 2013; Andreani et al. 2014; Gunawardhana et al. 2015; Andreani et al. 2018; Yuan et al. 2018), completeness problems in galaxy surveys (e.g. Johnston, Teodoro & Hendry 2012), cosmology with gravitational lensing (e.g. Sato, Ichiki & Takeuchi 2010, 2011; Lin & Kilbinger 2015; Simon & Schneider 2017), time series analysis of bivariate sequence (e.g. Jo 2019) and many other astrophysical applications (e.g. Jiang, Yeh & Hung 2015; Koen & Bere 2017; Jo et al. 2019).

When we have introduced the copula method to the galactic astrophysics and cosmology in Takeuchi (2010), practical application of the copula was restricted to the bivariate problems. This is because of the fact that the copula method was not easy to extend to higher dimensions ($d > 2$), except some special cases like Gaussian. Further, even if we find such a multivariate analytic function in some very fortunate case, it would probably be very inflexible and impractical, for example, to a realistic statistical estimation in galaxy surveys. Actually, however, a method to improve the copula method and resolve the difficulty to multivariate extension was introduced just some years before Takeuchi (2010). This is based on the decomposition of a general multivariate distribution: a multivariate probability density function can be factorized into a bivariate copulae and univariate density functions. Since we have a rich theoretical method of bivariate copulae, this means that we can extend our methodology to any higher dimension problems (e.g. Aas et al. 2009, and references therein). However, since this decomposition is not unique, we need to sort it out to have a systematic procedure. For this purpose, we introduce the concept of vine copula, invented in the field of graphical modelling (Bedford & Cooke 2002). Since the work of Aas et al. (2009), vine copulae have been applied to vastly wide range of fields: financial risk management (e.g. Sriboonchitta et al. 2014; Aas 2016; Allen, McAleer & Singh 2017; Nagler, Bumann & Czado 2019), insurance (e.g. Peters, Dong & Kohn 2014; Mejdoub & Ben Arab 2018; Shi & Yang 2018), weather forecast and engineering (e.g. Alidoost, Su & Stein 2019; Kloubert 2020; Torabi et al. 2020), multivariate time series analysis (e.g. Almeida, Czado & Manner 2016; Jäger MOrales 2017; Acar, Czado & Lysy 2019), spatio-temporal analysis (e.g. Gräler 2011, 2014; Callau Poduje & Haberlandt 2018), and technological applications (e.g. Xu D. et al. 2017; Xu, Hua & Xu 2017; Khuntia, Rueda & van der Meijden 2019), among many others.

In this work, we show a systematic method to extend the copula to unlimitedly higher dimensions by a vine copula method.[2]

This paper is organized as follows: in Section 2, we briefly review the basics of copula. Then we introduce the central concept of this work, vines, and formulate the systematic construction of a multivariate copula with vines. In Section 3, we make use of these copulae to construct an MMF of galaxies. We discuss some implications and further applications in Section 3. Section 4 is devoted to summary and conclusions.

Throughout this paper, we adopt a cosmological model $(h, \Omega_{M0}, \Omega_{\Lambda0}) = (0.7, 0.3, 0.7)$ $[h \equiv H_0/100\,(\mathrm{km\,s}^{-1})\,\mathrm{Mpc}^{-1}]$.

2 FORMULATION

2.1 Copula

First, we briefly review the concept of copula. In short, copulae are functions that relate joint multivariate distribution functions (DFs) to their one-dimensional marginal DFs.[3] Using a copula C, any multivariate DF, G, can be expressed with margins $F_1, F_2, \ldots, F_d$ as

$$G(x_1, \ldots, x_d) = C[F(x_1), \ldots, F_d(x_d)] . \tag{1}$$

This is guaranteed by Sklar's theorem (Sklar 1959). Especially, if $F_1, \ldots, F_d$ are continuous, then C is unique. A comprehensive proof of Sklar's theorem is found in e.g. Nelsen (2006). This theorem gives a basis that *any* multivariate DF with given margins is expressed with a form of equation (1). If we want a more familiar form, a PDF of $G(x_1, \ldots, x_d)$, $g(x_1, \ldots, x_d)$, is written as

$$g(x_1, \ldots, x_n) = \frac{\partial^d C[F_1(x_1), \ldots, F_d(x_d)]}{\partial x_1 \ldots \partial x_d} f_1(x_1) \ldots f_d(x_d) \equiv c[F_1(x_1), \ldots, F_d(x_d)] f_1(x_1) \ldots f_d(x_d), \tag{2}$$

where $f_1(x_1), \ldots, f_d(x_d)$ are PDFs of $F_1(x_1), \ldots, F_d(x_d)$, respectively. Here, a function $c[F_1(x_1), \ldots, F_d(x_d)]$ is referred to as the copula density of C. For more detailed (but not too rigorous) definitions, readers are guided to Takeuchi (2010, hereafterT10).

The most important statistical aspect of bivariate DFs is their dependence properties between variables. Since the dependence can never be given by the marginals of a DF, this is the most non-trivial information that a bivariate DF provides. Since any bivariate DFs are described by equation (1), all the information on the dependence is carried by their copulae.

For practical data analysis, a measure of dependence is useful for the interpretation of a result. The Pearson's product–moment correlation coefficient ρ is the most frequently used dependence measure for physical scientists (and others). For a while in this paragraph, we focus on

---

[1]In Takeuchi (2010), we used 'copulas' as its plural form. However, since this terminology is a Latin feminine noun, we use 'copulae' instead in this paper.

[2]After the submission of this manuscript, a similar work by Vio, Nagler & Andreani (2020) appeared on arXiv. Readers are also guided to their article as a different astrophysical application of the vine copula.

[3]As in , the DF stands for a cumulative distribution function in statistical terminology. To avoid confusion, we use a term 'probability density function (PDF)' to refer to a distribution function commonly used in physics. In this paper (and statistical literature in general), we distinguish a DF and PDF by an upper and lower case, respectively [e.g. $F(x)$ stands for a certain DF, and $f(x)$ is its PDF].

the bivariate PDF since we consider correlation measures. The bivariate PDF of x_1 and x_2, $g(x_1, x_2)$, is written as

$$g(x_1, x_2) = \frac{\partial^2 C[F_1(x_1), F_2(x_2)]}{\partial x_1 \partial x_2} f_1(x_1) f_2(x_2) = c[F_1(x_1), F_2(x_2)] f_1(x_1) f_2(x_2). \tag{3}$$

Then the correlation coefficient ρ is expressed as

$$\rho = \frac{\int (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) g(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2}{\sqrt{\int (x_1 - \bar{x}_1)^2 f_1(x_1) \mathrm{d}x_1 \int (x_2 - \bar{x}_2)^2 f_2(x_2) \mathrm{d}x_2}} = \frac{\int (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) c[(F_1(x_1), F_2(x_2)] f_1(x_1) f_2(x_2) \mathrm{d}x_1 \mathrm{d}x_2}{\sqrt{\int (x_1 - \bar{x}_1)^2 f_1(x_1) \mathrm{d}x_1 \int (x_2 - \bar{x}_2)^2 f_2(x_2) \mathrm{d}x_2}} . \tag{4}$$

We observe that equation (4) depends not only on the dependence of two variables (copula part) but also on its marginals $f_1(x_1)$, $f_2(x_2)$, i.e. *the linear correlation coefficient ρ does not measure the dependence purely.* Then, sometimes a genuine measure of dependence, e.g. Spearman's ρ_S or Kendall's τ would be more appropriate. Spearman's rank correlation is a non-parametric version of Pearson's correlation using a rank of data. The population version of Spearman's ρ_S is expressed by copula as

$$\rho_S = 12 \int_0^1 \int_0^1 u_1 u_2 \mathrm{d}C(u_1, u_2) - 3 = 12 \int_0^1 \int_0^1 C(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 - 3 . \tag{5}$$

Kendall's τ is also expressed in a simple form in terms of copula as

$$\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) \mathrm{d}C(u_1, u_2) - 1 = 4 \int_0^1 \int_0^1 C(u_1, u_2) c(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 - 1 . \tag{6}$$

The derivation of these equations are found in . In equations (5) and (6), are *independent of the distributions F_1, F_2, or G and depend only on the dependence structure*, i.e. a copula. This is the reason why the two dependence measures are almost always used in the context of copulae in the literature.

What we have from surveys are usually multivariate data sets, and we do not know the functional form of a multivariate DF from which the data are sampled. Namely, there is infinite degrees of freedom for a set of copulae to choose. For a bivariate case, it might be still possible to restrict a class of functions for a copula (e.g. Takeuchi et al. 2013; Andreani et al. 2014; Gunawardhana et al. 2015; Andreani et al. 2018). However, it would be almost impossible to have an intuition to choose an appropriate family of a single multivariate copulae/copula density for a certain survey data. To make the problem practically more accessible, we need a systematic construction method of a multivariate copula from a lower-dimensional information. We introduce such a method in the following.

2.2 Vine copula

Here, we introduce a vine copula as a systematic method to factorize a multivariate PDF as above. A vine is a concept originally introduced in the field of graphical modelling (Bedford & Cooke 2002).

2.2.1 Factorization of a PDF

As we mentioned in Section 1, this is based on the decomposition of a general multivariate distribution. Let $f(x_1, \ldots, x_d)$ be a joint PDF of a set of d-dimensional vector stochastic variable $\vec{X} = (X_1, \ldots, X_d)$. First, recall the formula of conditional probability

$$f(A, B) = f(B|A) f(A) , \tag{7}$$

where A and B are events. If we apply this formula to the above PDF, we have

$$\begin{aligned} f(x_1, \ldots, x_d) &= f_{2\ldots d|1}(x_2, \ldots, x_d | x_1) f_1(x_1) \\ &= f_{3\ldots d|12}(x_3, \ldots, x_d | x_1, x_2) f_{2|1}(x_2 | x_1) f_1(x_1) \\ & \vdots \\ &= f_{d|123\ldots d-1}(x_d | x_1, \ldots, x_{d-1}) \cdots f_{2|1}(x_2 | x_1) f_1(x_1), \end{aligned} \tag{8}$$

which is known as the chain rule. We start from this well-known mathematical formula. By using a bivariate copula density
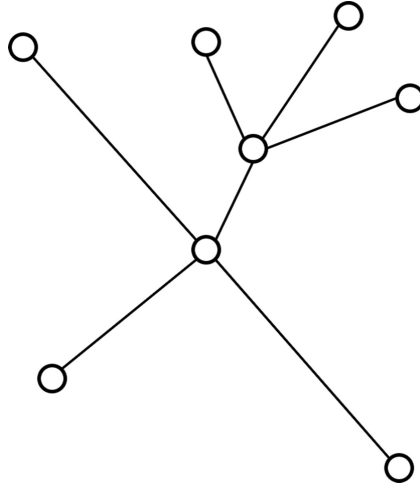
$$f_{12}(x_1, x_2) = c[F_1(x_1), F_2(x_2)] f_1(x_1) f_2(x_2), \tag{9}$$

the conditional probability can be expressed as

$$\begin{aligned} f_{2|1}(x_2 | x_1) &= \frac{f_{12}(x_1, x_2)}{f_1(x_1)} \\ &= c_{12}[F_1(x_1), F_2(x_2)] f_2(x_2). \end{aligned} \tag{10}$$

Similarly, for $f_{23|1}$,

$$f_{23|1}(x_2, x_3 | x_1) = c_{23|1}\left[F_{2|1}(x_2 | x_1), F_{3|1}(x_3 | x_1)\right] f_{2|1}(x_2 | x_1) f_{3|1}(x_3 | x_1) . \tag{11}$$

Figure 1. An example of a tree graph.

Since

$$\frac{f_{23|1}(x_2, x_3|x_1)}{f_{2|1}(x_2|x_1)} = f_{3|12}(x_3|x_1, x_2) \,, \tag{12}$$

we obtain

$$\begin{aligned} f_{3|12}(x_3|x_1, x_2) &= c_{23|1} \left[F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1) \right] f_{3|1}(x_3|x_1) \\ &= c_{23|1} \left[F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1) \right] c_{13} \left[F_1(x_1), F_3(x_1) \right] f_3(x_3) \,. \end{aligned} \tag{13}$$

We can generalize the formula. Set $\vec{v} \equiv (v_1, \ldots, v_k) = x_{i_1}, \ldots, x_{i_k}$. If we define $\vec{v}_{-j} \equiv (v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k)$, the following formula holds

$$f(x|\vec{v}) = c_{x\vec{v}_j|\vec{v}_{-j}} \left[F(x)|\vec{v}_{-j}), F_{v_j}(v_j|\vec{v}_{-j}) \right] f(x|\vec{v}_{-j}) \,. \tag{14}$$

This is a purely mathematical, direct result of the formula of conditional probability.

We should note that such a decomposition is *not unique* if we consider a permutation of the labels of variables, and when d is large, the number of representations increases dramatically. Hence, we need a systematic procedure to choose which pair combinations should be used to describe the dependence. For this purpose, we introduce the concept of a vine. Since it was invented in the field of graphical modelling (Bedford & Cooke 2002), it is convenient to use diagrams referred to as graphs.

2.2.2 Vine

In order to define it, we have to introduce some graph-theoretical terminologies. We start from the definition of a tree in graph theory.

DEFINITION 1. **(tree)**
Consider a set of d nodes. When a graph T is connected and has no cycles, T is a tree.

An example of a tree is presented in Fig. 1.
Based on the concept of tree graph, we define a vine.

DEFINITION 2. **(vine)**
A vine $\mathcal{V}$ on d elements $\{1, 2, \ldots, d\}$ is a set of trees T_i ($i = 1, \ldots, d - 1$), which satisfies the following conditions:

  (i) T_1 is a connected tree that have $\{1, 2, \ldots, d\}$ as a set of nodes and E_1 as a set of edges,
  (ii) For $i = 2, \ldots, d - 1$, T_i is a tree that have E_{i-1} as a set of nodes and E_i as a set of edges.

DEFINITION 3. **(regular vine)**
If two nodes in tree T_{i+1} are joined by an edge, the corresponding edges in tree i share a node. This is referred to as the proximity condition.

Following discussions will be restricted to regular vines without any loss of generality, since the class of regular vines is still so large that it can treat most of the practical cases. The structure of a regular vine is schematically described in Fig. 2. The term 'vine' is named after the fact that its botryoidal structure looks similar to a cluster of grapes in its appearance (see Fig. 2; e.g. Chapter 1 of Kurowicka & Joe 2011). We note that the tree structure is not strictly necessary for applying the pair-copula methodology, but it helps with identifying the different pair-copula decompositions (Aas et al. 2009).
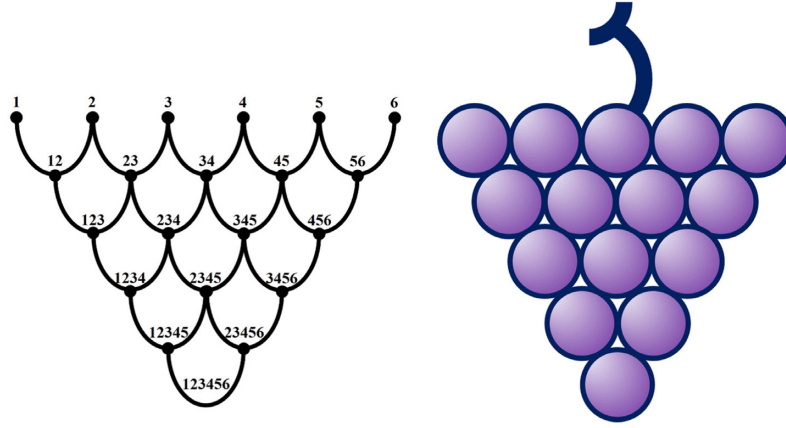
Figure 2. A diagrammatic presentation of a concept of vines. As an example, we show the case with six variables. This is called a vine because it looks like a grape (right-hand panel).

2.2.3 Frequently used vines

In practice, two subclasses of vines are frequently used. They are so-called D-vine and C-vine, introduced as follows.

DEFINITION 4. **D (drawable)-vine**

Any joint PDF $f(x_1, \ldots, x_d)$ can be written down by D-vine as follows.

$$f(x_1, \cdots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j)|(i+1),\ldots,(i+j-1)} \left[F(x_i|x_i, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{x_{i+j-1}}) \right] \prod_{k=1}^{d} f_k(x_k) , \tag{15}$$

where index j identifies the trees, and i runs over each tree (Bedford & Cooke 2001).

The definition of D-vine means that for any tree T_i, the number of edges connected to each nodes never exceeds 2. To have a concrete idea, we present examples for $d = 3, 4,$ and 5.

$$f(x_1, x_2, x_3) = c_{13|2} \left[F(x_1|x_2)F(x_3|x_2) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{23} \left[F_2(x_2), F_3(x_3) \right]$$
$$f_1(x_1)f_2(x_2)f_3(x_3) , \tag{16}$$

$$f(x_1, x_2, x_3, x_4) = c_{14|23} \left[F(x_1|x_2, x_3)F(x_4|x_2, x_3) \right]$$
$$c_{13|2} \left[F(x_1|x_2), F_3(x_3|x_2) \right] c_{24|3} \left[F(x_2|x_3), F(x_4|x_3) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{23} \left[F_2(x_2), F_3(x_3) \right] c_{34} \left[F_3(x_3), F_4(x_4) \right]$$
$$f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4) , \tag{17}$$

$$f(x_1, x_2, x_3, x_4, x_5) = c_{15|234} \left[F(x_1|x_2, x_3, x_4)F(x_5|x_2, x_3, x_4) \right]$$
$$c_{14|23} \left[F(x_1|x_2, x_3)F(x_4|x_2, x_3) \right] c_{25|34} \left[F(x_2|x_3, x_4)F(x_5|x_3, x_4) \right]$$
$$c_{13|2} \left[F(x_1|x_2), F_3(x_3|x_2) \right] c_{24|3} \left[F(x_2|x_3), F(x_4|x_3) \right] c_{35|4} \left[F(x_3|x_4), F(x_5|x_4) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{23} \left[F_2(x_2), F_3(x_3) \right] c_{34} \left[F_3(x_3), F_4(x_4) \right] c_{45} \left[F_4(x_4), F_5(x_5) \right]$$
$$f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4)f_5(x_5) . \tag{18}$$

A diagrammatic representation of a D-vine with five variables is shown in Fig. 3. This describes the dependence structure of the D-vine well. In this case it has four layers of tree structure, labelled as T_i ($i = 1, \ldots, 4$). Each edge is associated with a pair copula.

DEFINITION 5. **C (canonical)-vine**

Any joint PDF $f(x_1, \ldots, x_d)$ can be written down by C-vine as follows.

$$f(x_1, \cdots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,(j+i)|1,\ldots,(j-1)} \left[F(x_j|x_1, \ldots, x_{j-1}), F(j_i|x_1, \ldots, x_{j-1}) \right] \prod_{k=1}^{d} f_k(x_k) . \tag{19}$$

Each tree T_j has a unique node connected to $d - j$ edges.

When we know a particular variable is a key that governs the interaction in the data set, the C-vine has a great advantage. We can decide this 'pivot' variable at the root of the C-vine. We show examples for $d = 3, 4,$ and 5.
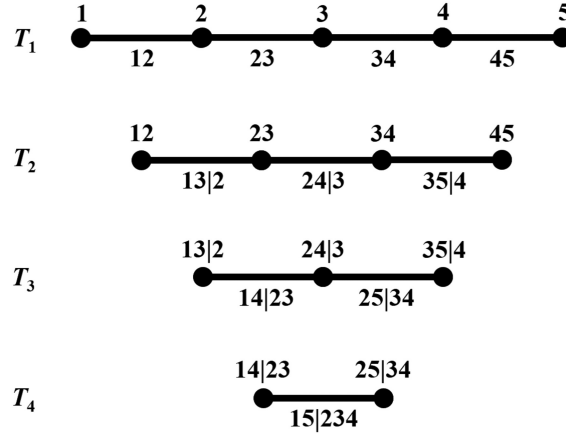
Figure 3. A diagrammatic representation of a D-vine with five variables. It contains 4 trees and 10 edges. Each edge is associated with a pair copula.
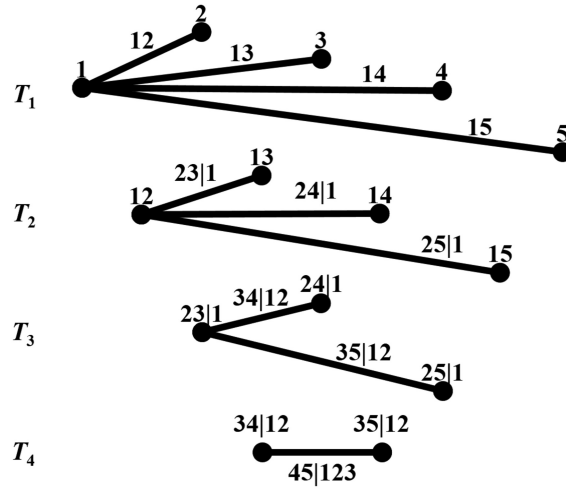
Figure 4. Same as Fig. 3 but for a C-vine with five variables.

$$f(x_1, x_2, x_3) = c_{23|1} \left[F(x_2|x_3) F(x_3|x_1) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{13} \left[F_1(x_1), F_3(x_3) \right]$$
$$f_1(x_1) f_2(x_2) f_3(x_3) , \tag{20}$$

$$f(x_1, x_2, x_3, x_4) = c_{34|12} \left[F(x_3|x_1, x_2) F(x_4|x_1, x_2) \right]$$
$$c_{23|1} \left[F(x_2|x_1), F(x_3|x_1) \right] c_{24|1} \left[F(x_2|x_1), F(x_4|x_1) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{13} \left[F_1(x_1), F_3(x_3) \right] c_{14} \left[F_1(x_1), F_4(x_4) \right]$$
$$f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) , \tag{21}$$

$$f(x_1, x_2, x_3, x_4, x_5) = c_{45|123} \left[F(x_4|x_1, x_2, x_3) F(x_5|x_1, x_2, x_3) \right]$$
$$c_{34|12} \left[F(x_3|x_1, x_2) F(x_4|x_1, x_2) \right] c_{35|12} \left[F(x_3|x_1, x_2) F(x_5|x_1, x_2) \right]$$
$$c_{23|1} \left[F(x_2|x_1), F(x_3|x_1) \right] c_{24|1} \left[F(x_2|x_1), F(x_4|x_1) \right] c_{25|1} \left[F(x_2|x_1), F(x_5|x_1) \right]$$
$$c_{12} \left[F_1(x_1), F_2(x_2) \right] c_{13} \left[F_1(x_1), F_3(x_3) \right] c_{14} \left[F_1(x_1), F_4(x_4) \right] c_{15} \left[F_1(x_1), F_5(x_5) \right]$$
$$f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) f_5(x_5) . \tag{22}$$

A diagrammatic representation of a C-vine with five variables is shown in Fig. 4. For the case of $d = 3$, the general expression for vine structure is expressed as equation (16) or (20). It is valid for both D- and C-vines. There are six ways with a permutation between x_1, x_2, and x_3, but only three of them yield different pair-copula decompositions. Further, each of the three correspond both to D- and C-vine. Namely, both D- and C-vines cover the whole possible structures of pair copula decompositions for $d = 3$. For $d = 4$, there are 24 regular vine decompositions,

of which 12 are D-vine and 12 are C-vine ones. There is no overlap between any of the D- and C-vines. Further, there are no other regular vine decompositions. This guarantees the wide range of applicability of D- and C-vines for the pair-copula decomposition. For $d = 5$, there are 60 different D-vines and 60 different C-vines. Between any of these 60 D-vines and 60 C-vines, there is no overlap. However, unlike $d \leq 4$, there are 120 more regular vines that are not D- nor C-vines. Hence in total there are 240 different possible pair-copula decompositions. In general, for any d, there are $d!/2$ D-vines and the same number of C-vines. Detailed explanations on these results are presented in Aas et al. (2009).

2.3 Likelihood for vines

Practically, we can safely restrict the pair-copula decompositions to D- and C-vines. When we have n data sample $\{\vec{x}_m\} = \{(x_{m,1}, x_{m,2}, \ldots, x_{m,d})\}$ $(m = 1, \ldots, n)$, the log likelihood of a D-vine for the parameter estimation is

$$
\begin{aligned}
\ln \mathcal{L}(\vec{\theta}; \vec{x}_m, m = 1, \ldots, n) &= \sum_{m=1}^{n} \ln f(\vec{\theta}; \vec{x}_m) \\
&= \sum_{m=1}^{n} \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \ln c[\vec{\theta}(\text{copula})]_{i,(i+j)|(i+1),\ldots,(i+j-1)} \left[F(x_i | x_i, \ldots, x_{i+j-1}), F(x_{i+j} | x_{i+1}, \ldots, x_{x_{i+j-1}}) \right] \\
&\quad + \sum_{k=1}^{d} \ln f_k[\vec{\theta}(\text{marginal}); x_k] ,
\end{aligned}
\tag{23}
$$

where we denote the parameter vectors for copulae and marginals in a symbolic way for saving space. Similarly, the log likelihood of a C-vine is

$$
\begin{aligned}
\ln \mathcal{L}(\vec{\theta}; \vec{x}_m, m = 1, \ldots, n) &= \sum_{m=1}^{n} \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \ln c[\vec{\theta}(\text{copula})]_{j,(j+i)|1,\ldots,(j-1)} \left[F(x_i | x_i, \ldots, x_{i+j-1}), F(x_{i+j} | x_{i+1}, \ldots, x_{x_{i+j-1}}) \right] \\
&\quad + \sum_{k=1}^{d} \ln f_k[\vec{\theta}(\text{marginal}); x_k] .
\end{aligned}
\tag{24}
$$

We should maximize the log likelihood equation (23) or (24) to estimate the parameter set $\vec{\theta} = [\vec{\theta}(\text{copula}), \vec{\theta}(\text{marginal})]$. In principle, both $\vec{\theta}(\text{copula})$ and $\vec{\theta}(\text{marginal})$ can be estimated simultaneously. In the research fields such as economics, however, the likelihood estimation of the marginals is often found to be difficult or even implausible. Hence, they do not use the exact form of the likelihood but instead maximize the so-called pseudo-likelihood (Aas et al. 2009). In contrast, in astrophysics, we have a rich field of research on the estimation of the marginals, e.g. the LF or MF of galaxies at a certain observed wavelength (e.g. Takeuchi et al. 2000; Johnston 2011). Then, we can simply use the estimated marginals and plug in them for the log-likelihood. Namely, we can omit the estimation step for the marginals when we try to estimate copula parameters in a different sense from other research fields.

On this step, we can obtain the error of each parameter and the goodness-of-fit indicator(s) of the model, since we have the likelihood ellipsoid and maximum likelihood value.

2.4 Model selection with vines

The likelihood estimation we discussed above is only one of the steps of the full estimation problem. Schematically, the model to be specified has a following structure as

Model = structure (trees) + copula families + copula parameters.
$\tag{25}$

Namely, for the estimation procedure we should consider

  (i) selection of a specific decomposition,
  (ii) choice of pair-copula types,
  (iii) estimation of the copula parameters.

This is schematically described in Fig. 5. As for the copula types, since we have an infinitely large degree of freedom for the choice of copulae, the model selection is fundamentally important. Often we have to choose an optimal model from large choice of candidate models with different number of parameters. In such a case, usual goodness-of-fit method does not work, since obviously more parameters give a better fit. For such a case, a model selection procedure should be used instead. The most popular tool for the model selection is the information criterion (e.g. Takeuchi 2000, and references therein). One of the first and most widely used information criterion is the Akaike Information Criterion (AIC; Akaike 1974),

$$
\text{AIC}(q) = -2[\ln \mathcal{L}(\hat{\vec{\theta}}) - q],
\tag{26}
$$

where $\hat{\vec{\theta}}$ stands for the parameter that maximizes the likelihood, and q is the number of parameters. The model that gives the smallest AIC is selected. This is a natural extension of the classical maximum likelihood estimation, corrected for the bias introduced by the parameter
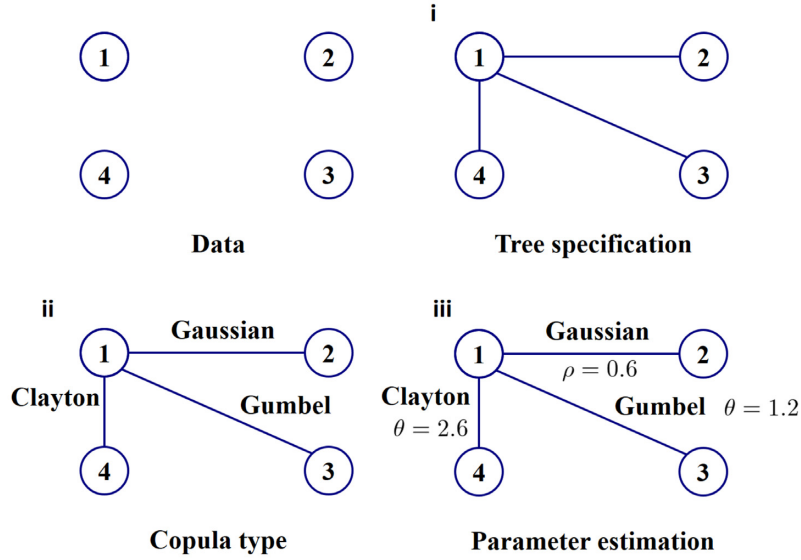
Figure 5. A schematic example for the description of the estimation with vines. Gaussian, Clayton, and Gumbel stand for three popular copula types often used in practice. Each of them has some specifying parameters, and they are estimated by usual statistical procedure in Step iii.

estimation step (Akaike [1974]). Some other information criteria are also used. Among them, the Bayesian Information Criterion (BIC)

$$\text{BIC}(q) = -2 \left[\ln \mathcal{L}(\hat{\theta}) - \frac{q}{2} \ln n \right] \tag{27}$$

(n: sample size) is also often used (Schwarz [1978]). This model selection is performed in the step of copula selection. Each copula is determined by the evaluation of such information criterion among all possible copula types, as well as the classical goodness-of-fit indicators like χ^2-statistic.

As we saw in Section 2.2, the tree determination and subsequent specification of copulae would be computationally heavy because of the factor $d!/2$. Thanks to the present-day development of software, we can treat a problem with a dimension up to $d \sim 500$. Some software packages are available for this problem (e.g. Brechmann & Schepsmeier [2013]). The construction and estimation of a multivariate PDF based on the vine copula is completed by this step.

3 APPLICATION TO CONSTRUCT A MULTIVARIATE MASS FUNCTION OF GALAXIES

The relation between stars, atomic gas, and molecular gas mass is one of the most important issues in galaxy evolution, and studied very extensively. For this aim, a multivariate MF is obviously a fundamental tool (see e.g. an elaborate work of Rodríguez-Puebla et al. [2020]). Here we present one astrophysically interesting example of multivariate MF, a three-variate mass function of stellar, atomic gas, and molecular gas mass. The formal procedure is exactly the same for the case of luminosity functions.

3.1 Multivariate MF

First, we prepare some notations for the multivariate MF. A MF of galaxies is defined as a number density of galaxies whose mass lies between a logarithmic interval[4] $[\log M, \log M + \text{d}\log M]$:

$$\phi^{(1)}(M) \equiv \frac{\text{d}n}{\text{d}\log M} \ . \tag{28}$$

For mathematical simplicity, we define the as *being normalized*, i.e.

$$\int \phi^{(1)}(M)\text{d}\log M = 1 \ . \tag{29}$$

Hence, this corresponds to a PDF. We also define the cumulative MF as

$$\Phi^{(1)}(M) \equiv \int_{\log M_{\min}}^{\log M} \phi^{(1)}(M')\text{d}\log M' \ , \tag{30}$$

where $M_{\min}$ is the minimum mass of galaxies considered. This corresponds to the DF.

---

[4]We denote $\log x \equiv \log_{10} x$ and $\ln x \equiv \log_e x$.

Table 1. Parameters for the gas mass function.

MF	α	ϕ_* (Mpc^{-3} dex^{-1})	M_* (M$_\odot$)	Reference
Atomic gas	-1.25 ± 0.02	$(4.5 \pm 0.2) \times 10^{-3}$	8.7×10^9	Jones et al. (2018)
Molecular gas	-1.30 ± 0.16	$(5.9 \pm 2.8) \times 10^{-3}$	9.4×10^9	Keres et al. (2003)

Table 2. Parameters for the stellar mass function.

α_1	ϕ_{1*} (Mpc^{-3} dex^{-1})	M_{1*} (M$_\odot$)	α_2	ϕ_{2*} (Mpc^{-3} dex^{-1})	M_{2*} (M$_\odot$)	Reference
1.082	6.0×10^{-2}	4.1×10^{10}	1.120	2.5×10^{-3}	9.9×10^{10}	D'Souza et al. (2015)

If we denote univariate MFs as $\phi_k^{(1)}(M_k)\,(k = 1, \ldots, d)$, the joint multivariate PDF $\phi^{(d)}(M_1, \ldots, L_d)$ is described by a differential copula $c(u_1, \ldots, u_k)$ as

$$\phi^{(d)}(M_1, \ldots, M_d) \equiv c\left[\Phi_1^{(1)}(M_1), \ldots, \Phi_d^{(1)}(M_d)\right] \phi_1^{(1)}(M_1) \cdots \phi_d^{(1)}(M_d) \,. \tag{31}$$

For this analysis, we made use of the R package VINECOPULA[5] (Aas et al. 2009). It provides statistical inference of C- and D-vine copulae. This package enables us to construct copula density and tree structures from multivariate data. The optimal combination of copulae and their parameters are chosen through AIC, BIC, and maximum likelihood estimation with `RVineStructureSelect` function.

3.2 The stellar–atomic gas–molecular gas multivariate MF

As marginals of the multivariate MF, we should determine the univariate MFs for stellar mass, atomic gas mass, and molecular gas mass, respectively. The univariate MF for atomic gas mass, and molecular gas mass are known to be well described by the Schechter function (Schechter 1976).

$$\phi^{(1)}(M) = (\ln 10)\, \phi_* \left(\frac{M}{M_*}\right)^{1-\alpha} \exp\left[-\left(\frac{M}{M_*}\right)\right] \,, \tag{32}$$

For the atomic gas mass, we took the parameters from Jones et al. (2018) but in a normalized form with equation (29), i.e. we did not use the normalization factor $\phi_{\mathrm{H\,I}*}$. Similarly, we took the Schechter function parameters for the molecular gas from Keres, Yun & Young (2003). These parameters are summarized in Table 1.

Recent studies revealed that the stellar MF is, however, better described by a double Schechter function (e.g. D'Souza, Vegetti & Kauffmann 2015)

$$\phi^{(1)}(M) = \frac{\phi_{1*}}{M_{1*}} \left(\frac{M}{M_{1*}}\right)^{-\alpha_1} \exp\left[-\left(\frac{M}{M_{1*}}\right)\right] + \frac{\phi_{2*}}{M_{2*}} \left(\frac{M}{M_{2*}}\right)^{-\alpha_2} \exp\left[-\left(\frac{M}{M_{2*}}\right)\right] \,. \tag{33}$$

We should note that D'Souza et al. (2015) defined the double Schechter function (equation 33) for a linear mass interval dM, *not* dlog$M$. The parameters are shown in Table 2.

3.3 Data

In this work, we used a subsample of the combined data set compiled by Calette et al. (2018). Their original sample consists of Golden, Silver, and Bronze Categories both for H I and H$_2$. Full details of the original sample are found in Appendix of Calette et al. (2018). We briefly describe the data set used here.

3.3.1 *The compiled galaxy sample with H I information*

We used the following data sets for H I information.
Golden Category

(i) *GALEX* Arecibo SDSS Survey (GASS; Catinella et al. 2013): an optically selected subsample of 760 galaxies more massive than 10^{10} M$_\odot$ taken from a parent SDSS DR6 sample volume limited in the redshift range $0.025 < z < 0.05$ and cross-matched with the ALFALFA and *GALEX* surveys.

(ii) Field galaxies from the Herschel Reference Survey (HRS; Boselli et al. 2010; Boselli, Cortese & Boquien 2014; Boselli et al. 2014b, c): a *K*-band volume-limited ($15 \leq D\,[\mathrm{Mpc}] \leq 25$) sample of 323 galaxies complete to $K_s = -12$ and -8.7 mag for late-type galaxies and early-type galaxies, respectively.

(iii) Field early-type galaxies from the ATLAS^{3D} H I sample (Serra et al. 2012): a sample of 166 local early-type galaxies observed in detail with integral field unities (IFUs; Cappellari et al. 2011). The distance range of the sample is between 10 and 47 Mpc; the sample includes 39 galaxies from the Virgo Cluster, but for the Golden category, the early type galaxies in the Virgo cluster core were excluded by Calette et al. (2018).

Bronze Category

(i) Analysis of the interstellar Medium of Isolated GAlaxies (AMIGA; Lisenfeld et al. 2011): a redshift-limited sample ($1500 \leq v_{\rm rec}[\mathrm{km\,s^{-1}}] \leq 5000$) consisting of 273 isolated galaxies with reported multiband imaging and CO data.

3.3.2 The compiled galaxy sample with CO (H$_2$) information

We used the following data sets for H I information.
Golden Category

(i) Field galaxies from the HRS: the same sample described above (excluding Virgo Cluster core), with 155 galaxies with available CO information (101 detections and 54 non-detections).

(ii) CO Legacy Database for GASS (COLD GASS; Saintonge et al. 2011): a programme aimed at observing CO(1–0) line fluxes with the IRAM 30 m telescope for galaxies from the GASS survey described above. From the CO fluxes, the total CO luminosities, (and hence the H$_2$ masses) were calculated for 349 galaxies.

(iii) Field early-type galaxies from the ATLAS^{3D} H$_2$ sample (Young et al. 2011): the same sample described above (excluding the Virgo Cluster core) but with observations in CO using the IRAM 30 m Radio Telescope. The sample amounts for 243 early type galaxies with CO observations.

Bronze Category

(i) AMIGA (Lisenfeld et al. 2011): the same sample described above. The authors carried out their own observations of CO(*J*: 1–0) with the IRAM 30 m or the 14 m FCRAO telescopes for 189 galaxies; 87 more were compiled from the literature.

3.4 Result

We present the constructed M_*–$M_{\rm H_I}$–$M_{\rm H_2}$ MF with a three-dimensional vine copula. The goodness-of-fit analysis of the obtained three-dimensional MF to the data set gave the *p*-value $p = 0.41$ that suggests that the fit is appropriate. The M_*–$M_{\rm H_I}$–$M_{\rm H_2}$ MF is described by a C-vine copula. The structure is presented in Fig. 6. The structure of the estimated C-vine is labelled as (3, 1), (3, 2), and (2, 1|3) in Fig. 6. Here, 1 corresponds to M_*, 2 to $M_{\rm H_I}$, and 3 to $M_{\rm H_2}$, respectively.

The relation between M_* and $M_{\rm H_2}$ was well described by the BB8 copula

$$C(u_1, u_2; \theta, \delta) = \frac{1}{\delta} \left\{ 1 - \left[1 - \frac{1}{1 - (1-\delta)^\theta} \left(1 - (1 - \delta u_1)^\theta \right) \left(1 - (1 - \delta u_2)^\theta \right) \right]^{\frac{1}{\theta}} \right\}, \quad (\theta \geq 1, \ \delta \in (0, 1]) \tag{34}$$

with parameters $\theta = 6.00 \pm 1.76$ and $\delta = 0.49 \pm 0.10$. Kendall's τ of this pair is 0.35, reflecting the broad distribution with a loose correlation. The scatter plot of the data on the $M_*\mbox{--}M_{\rm H}{}_2$ plane and the corresponding bivariate PDF are presented in Fig. 7.

In contrast, the relation between $M_{\rm H_I}$ and $M_{\rm H_2}$ was found to be reproduced by the Frank copula,

$$C(u_1, u_2; \delta) = -\frac{1}{\delta} \log \left[1 - \frac{\left(e^{-\delta u_1} - 1 \right) \left(e^{-\delta u_2} - 1 \right)}{e^\delta - 1} \right], \quad (\delta \in \mathbb{R} \setminus \{0\}) \tag{35}$$

with a parameter $\delta = 8.80 \pm 0.43$. Kendall's τ of this pair is 0.63. The scatter plot of the data on the $M_{\rm H_I}$–$M_{\rm H_2}$ plane, and the corresponding bivariate PDF are presented in Fig. 8. On this plane, we see a moderately strong dependence between these two variables, reflected to $\tau = 0.63$.

For the conditional copula between M_* and $M_{\rm H_I}$ for given $M_{\rm H_2}$, the Student-*t* copula

$$C(u_1, u_2; \rho, \nu) = \int_{-\infty}^{u_1} \int_{-\infty}^{u_2} \frac{1}{2\pi \sqrt{1 - \rho^2}} \left[1 + \frac{s^2 + t^2 - 2\rho s t}{\nu(1 - \rho^2)} \right]^{-\frac{\nu+2}{2}} \, \mathrm{d}s \, \mathrm{d}t, \quad (\rho \in (-1, 1), \ \nu > 2) \tag{36}$$

was selected as the appropriate copula with $\rho = 0.29 \pm 0.03$ and $\nu = 30.0$, respectively.

Though it is not easy to visualize the 3D structure of the PDF, its heavily asymmetric structure is well described in Fig. 9. Of course it requires a further analysis for a physical interpretation, the copula method can provide us with a fundamentally important tool to understand
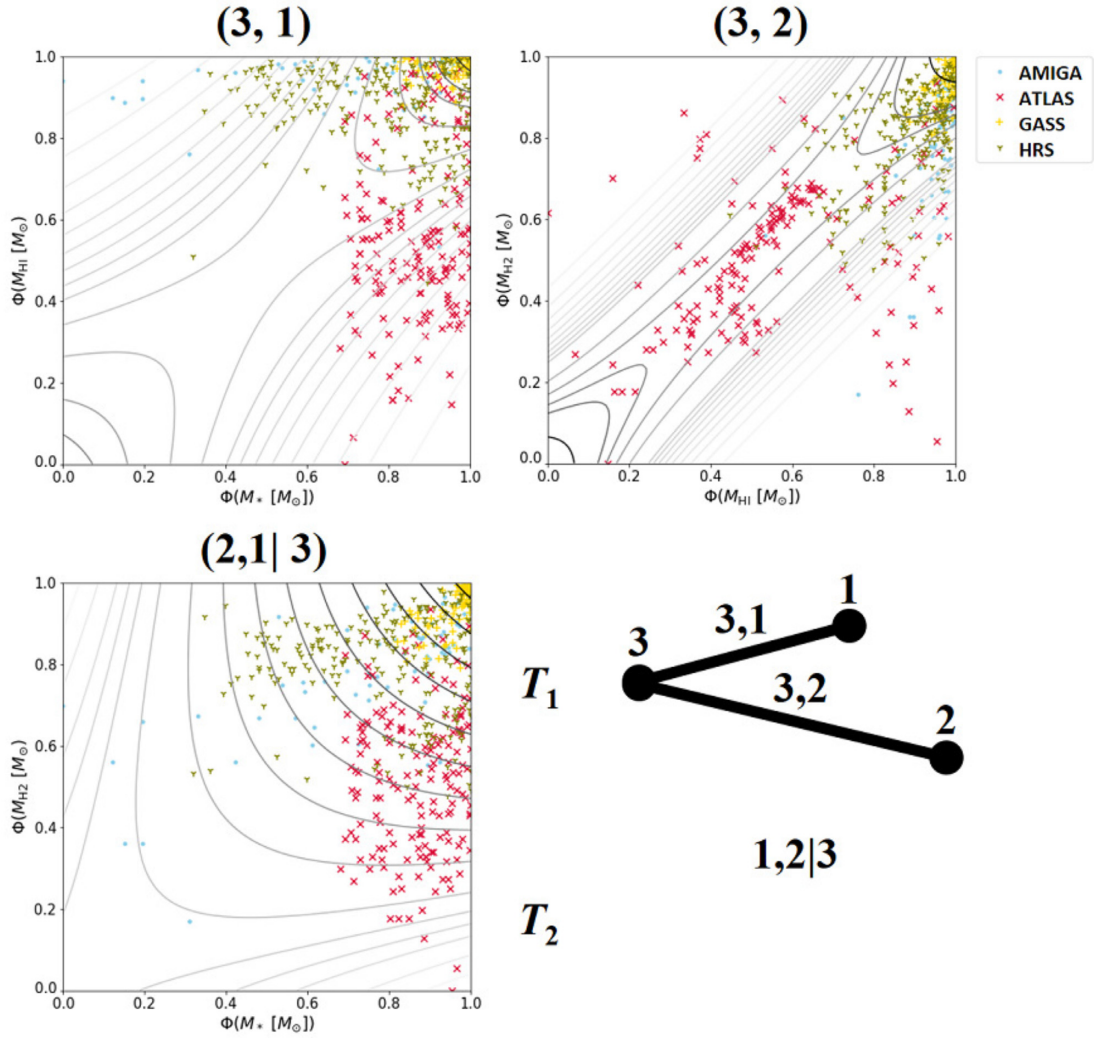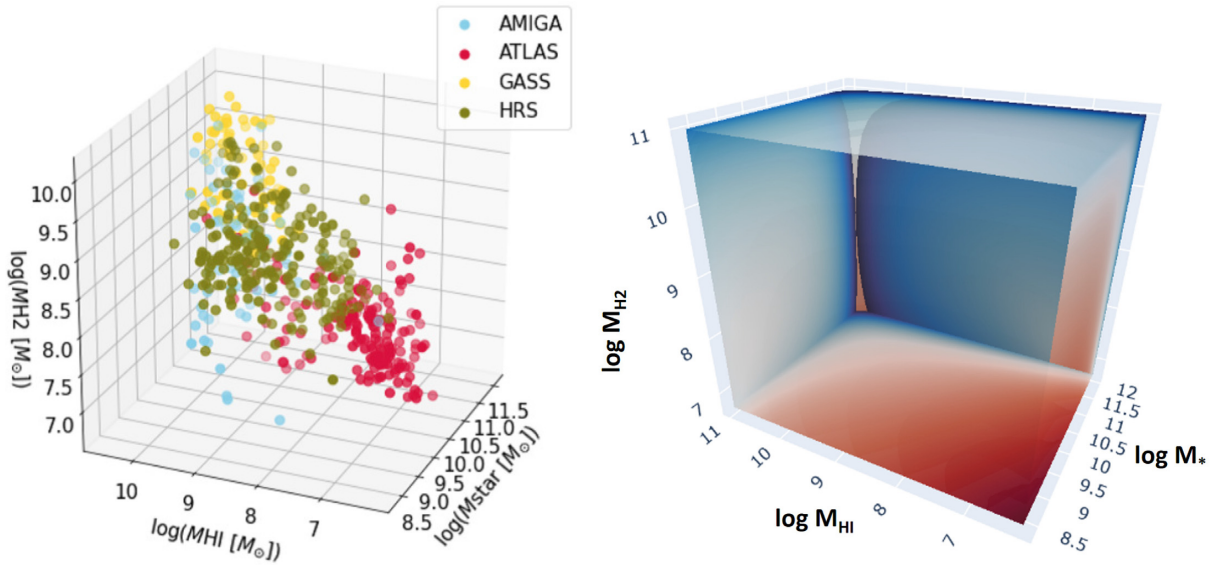
Figure 6. Vine copulae estimated from the M_*–M_{H_I}–M_{H_2} data. Top-left: the estimated copula for M_{H_2}–M_* relation, top-right: the estimated copula for M_{H_2}–M_{H_I} relation, bottom-left: the estimated conditional copula for M_*–M_{H_I} relation with M_{H_2} given, and bottom-right: tree structure of vine copulae. The structure of the estimated C-vine is labelled as (3, 1), (3, 2), and (2, 1|3), where node 1 corresponds to M_*, node 2 to M_{H_I}, and node 3 to M_{H_2}, respectively.
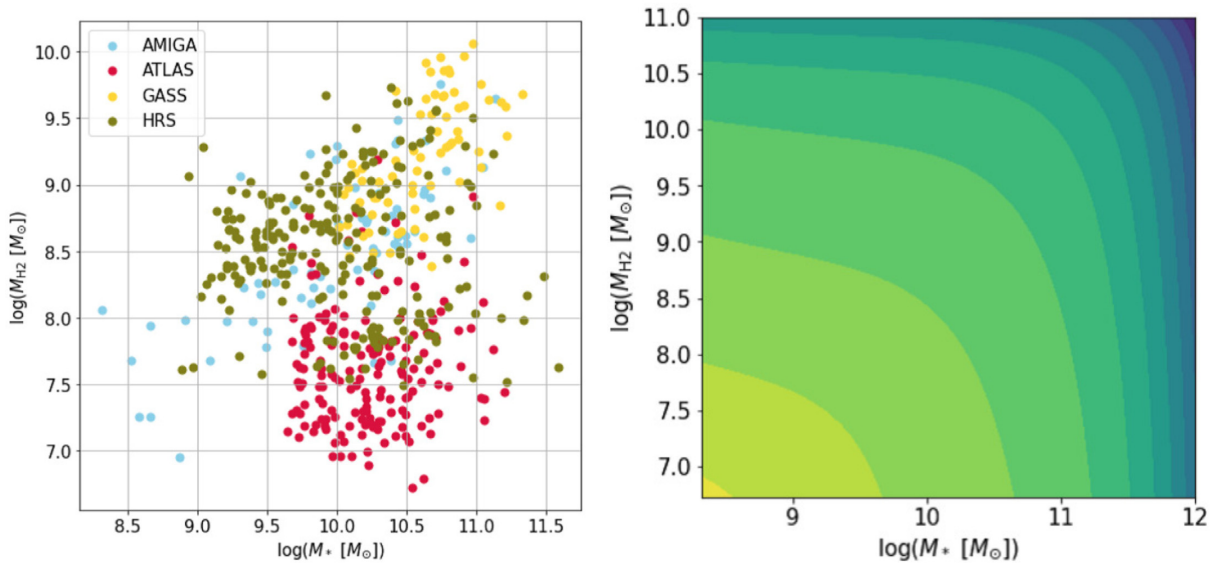
Figure 7. Relation between M_* and M_{H_2}. Left: the data distribution on the M_*–M_{H_2} plane. Right: estimated bivariate PDF. The selected copula is the BB8 copula with a parameter of $\theta = 6.00 \pm 1.76$ and $\delta = 0.49 \pm 0.10$. Data are taken from Calette et al. (2018).
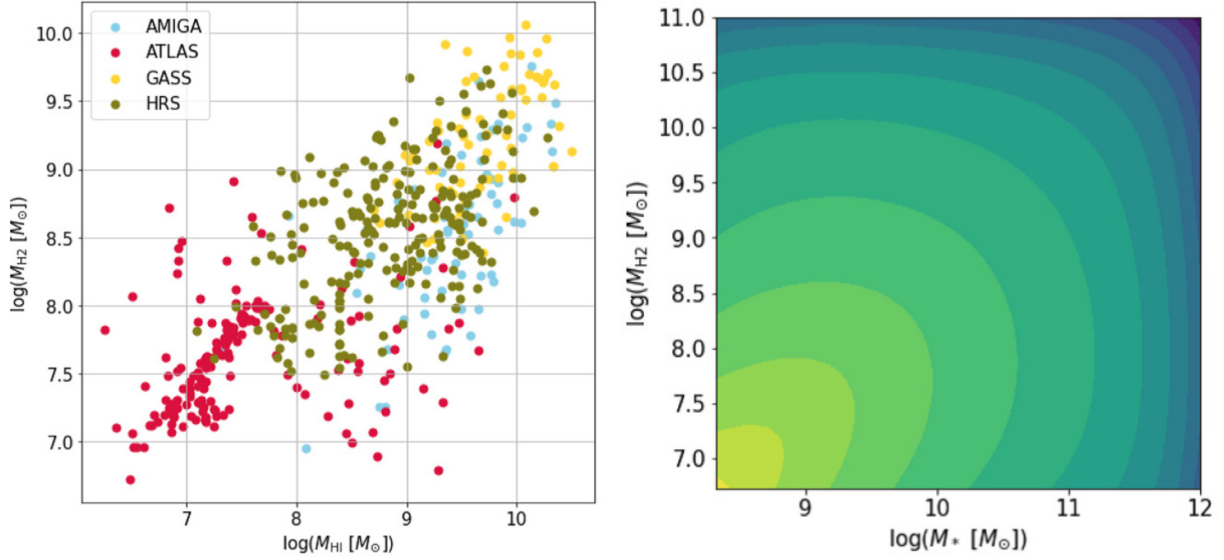
Figure 8. Relation between M_{H_I} and M_{H_2}. Left: the data distribution on the $M_{H_I}-M_{H_2}$ plane. Right: the corresponding bivariate PDF. The selected copula is the Frank copula with a parameter of $\delta = 8.80 \pm 0.43$. Data are taken from Calette et al. (2018).

Figure 9. The $M_*-M_{H_I}-M_{H_2}$ three-dimensional MF obtained from the vine copula likelihood estimation. Left: the data distribution in the mass space. Right: the estimated three-dimensional MF (PDF).

the physical processes behind the multivariate LF/MF. The result here is just a demonstration of how the vine copula performs well for the multidimensional LF/MF estimation. More physical discussion is planned to be presented in our next work.

If we go to much higher dimensions, human-intuitive approach may not work anymore even for a simple visualization. Very plausibly, we will confront the need for such a tremendously large data analysis. A machine-aided method will be a promising way to address such issues. We will discuss such a strategy in our future works.

4 SUMMARY AND CONCLUSIONS

We have proposed a systematic method to build a bivariate luminosity or MF of galaxies by using a copula (Takeuchi 2010: T10). It allows us to construct a distribution function when only its marginal distributions are available and the dependence structure should be estimated from data.

Though proposed a promising way to construct multivariate PDFs by a copula, the main limitation of the method is that it is not easy to extend a joint function to higher dimensions ($d > 2$), except some special cases like Gaussian. Even if we find such a multivariate analytic

function in some special case, it would probably be very inflexible and impractical. In this work, we introduced a systematic method to extend the copula to unlimitedly higher dimensions by a pair-copula decomposition method, referred to as the vine copula.

The vine copula method is extremely flexible because all the dependence structures are decomposed and reduced into pair dependence relations. We first formulated the factorization of a multidimensional DF/PDF. This factorization is not unique, and we introduced a vine structure as a systematic method to sort out the complicated structure. The vine copula can be described intuitively by a diagrammatic method, as shown in Figs 2, 3, and 4. We also presented that the likelihood parameter estimation of each copula and the model selection procedure can be performed simultaneously.

Then, as an interesting and important example, we applied the vine copula PDF method to estimate the 3D PDF of M_*, M_{H1}, and M_{H_2}. The vine copula can describe the PDF very well, and it can be used for the physical interpretation of the PDF as well as the evaluation of the complicated selection effect. We conclude that the vine copula method provides us with a promising way for the data analysis of unprecedentedly large surveys in the future. A machine-aided method will be a promising way to tackle such problems.

DATA AVAILABILITY STATEMENTS

The data underlying this article are available in the public domain mentioned in the main text. The data sets were derived from sources of this domain.

REFERENCES

Aas K., 2016, Econometrics, 4, 43.
Aas K., Czado C., Frigessi A., Bakken H., 2009, Insur. Math. Econ., 44, 182
Acar E. F., Czado C., Lysy M., 2019, Econom. Stat., 12, 181
Akaike H., 1974, IEEE Trans. Autom. Contrib., 19, 716
Alidoost F., Su Z., Stein A., 2019, Weather Clim. Extremes, 26, 100227
Allen D. E., McAleer M., Singh A. K., 2017, Sustainability, 9, 1762
Almeida C., Czado C., Manner H., 2016, Appl. Stoch. Models Bus. Ind., 32, 62
Andreani P., Boselli A., Ciesla L., Vio R., Cortese L., Buat V., Miyamoto Y., 2018, A&A, 617, A33
Andreani P. et al., 2014, A&A, 566, A70
Bedford T., Cooke R. M., 2001, Ann. Math. Artif. Intell., 32, 245
Bedford T., Cooke R. M., 2002, Ann. Stat., 30, 1031
Benabed K., Cardoso J.-F., Prunet S., Hivon E., 2009, MNRAS, 400, 219
Bhatawdekar R., Conselice C. J., Margalef-Bentabol B., Duncan K., 2019, MNRAS, 486, 3805
Binggeli B., Sandage A., Tammann G. A., 1988, ARA&A, 26, 509
Blanton M. R. et al., 2001, AJ, 121, 2358
Boselli A., Cortese L., Boquien M., 2014, A&A, 564, A65
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, A&A, 564, A66
Boselli A. et al., 2010, PASP, 122, 261
Boselli A. et al., 2014, A&A, 564, A67
Brechmann E. C., Schepsmeier U., 2013, J. Stat. Softw., 52, 1
Calette A. R., Avila-Reese V., Rodríguez-Puebla A., Hernández-Toledo H., Papastergis E., 2018, Rev. Mex. Astron. Astrofis., 54, 443
Callau Poduje A. C., Haberlandt U., 2018, Water, 10, 862
Caplar N., Lilly S. J., Trakhtenbrot B., 2018, ApJ, 867, 148
Cappellari M. et al., 2011, MNRAS, 413, 813
Catinella B. et al., 2013, MNRAS, 436, 34
Chapman S. C., Helou G., Lewis G. F., Dale D. A., 2003, ApJ, 588, 186
Chołoniewski J., 1985, MNRAS, 214, 197
de Lapparent V., Galaz G., Bardelli S., Arnouts S., 2003, A&A, 404, 831
Dutta S., Khandai N., Dey B., 2020, MNRAS, 494, 2664
D'Souza R., Vegetti S., Kauffmann G., 2015, MNRAS, 454, 4027
Gräler B., 2014, Spatial Stat., 10, 102
Gräler B., Pebesma E., 2011, Procedia Environ. Sci., 7, 206
Gunawardhana M. L. P. et al., 2015, MNRAS, 447, 875
Jiang I.-G., Yeh L.-C., Chang Y.-C., Hung W.-L., 2009, AJ, 137, 329
Jiang I.-G., Yeh L.-C., Hung W.-L., 2015, MNRAS, 449, L65

Jo H.-H., 2019, Phys. Rev. E, 100, 012306

Jo H.-H., Lee B.-H., Hiraoka T., Jung W.-S., 2019, Phys. Rev. E, 100, 022307

Johnson N. L., Kotz S., 1977, Commun. Stat. – Theory Methods, 6, 485

Johnston R., 2011, A&AR, 19, 41

Johnston R., Teodoro L., Hendry M., 2012, MNRAS, 421, 270

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2

Jäger Morales Nápoles O., 2017, ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A: Civil Eng., 3, 04017014

Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659

Khuntia S. R., Rueda J. L., van der Meijden M. A. M. M., 2019, Wind Energy, 1

Kloubert M.-L., 2020, Energies, 13, 1727

Koen C., 2009, MNRAS, 393, 1370

Koen C., Bere A., 2017, MNRAS, 471, 2771

Koprowski M. P. et al., 2017, MNRAS, 471, 4155

Kurowicka D., Joe H., 2011, Dependence Modeling: a Vine Copula Handbook. World Scientific, Singapore

Lake S. E., Wright E. L., Tsai C.-W., Lam A., 2017, AJ, 153, 189

Lin C.-A., Kilbinger M., 2015, A&A, 583, A70

Lin G. D., Dou X., Kuriki S., Huang J. S., 2014, J. Stat. Distrib. Appl., 1, 14

Lisenfeld U. et al., 2011, A&A, 534, A102

López-Sanjuan C. et al., 2017, A&A, 599, A62

Mashian N., Oesch P. A., Loeb A., 2016, MNRAS, 455, 2101

Mejdoub H., Ben Arab M., 2018, Res. Int. Bus. Finance, 45(C), 208

Mobasher B., Sharples R. M., Ellis R. S., 1993, MNRAS, 263, 560

Moffett A. J. et al., 2016, MNRAS, 462, 4336

Nagler T., Bumann C., Czado C., 2019, J. Multivariate Anal., 172, 180

Nelsen R. B., 2006, An Introduction to Copulae, 2nd ed. Springer, New York

Peters G. W., Dong A. X. D., Kohn R., 2014, Insur. Math. Econ., 59, 258

Rodríguez-Puebla A., Calette A. R., Avila-Reese V., Rodriguez-Gomez V., Huertas-Company M., 2020, PASA, 37, e024

Saintonge A. et al., 2011, MNRAS, 415, 32

Sato M., Ichiki K., Takeuchi T. T., 2010, Phys. Rev. Lett., 105, 251301

Sato M., Ichiki K., Takeuchi T. T., 2011, Phys. Rev. D, 83, 023501

Saunders W. et al., 2000, MNRAS, 317, 55

Schafer C. M., 2007, ApJ, 661, 703

Schechter P. L., 1976, ApJ, 203, 297

Scherrer R. J., Berlind A. A., Mao Q., McBride C. K., 2010, ApJ, 708, L9

Schwarz G., 1978, Ann. Stat., 6, 461

Serra P. et al., 2012, MNRAS, 422, 1835

Shi P., Yang L., 2018, J. Am. Stat. Assoc., 113, 122

Simon P., Schneider P., 2017, A&A, 604, A109

Sklar A., 1959, Publ. Inst. Stat. Univ. Paris, 8, 229

Sriboonchitta S., Liu J., Kreinovich V., Nguyen H. T., 2014, in Huynh V. N., Kreinovich V., Sriboonchitta S., eds, Modeling Dependence in Econometrics. Advances in Intelligent Systems and Computing, Vol. 251, Springer, Cham

Takeuchi T. T., 2000, Ap&SS, 271, 213

Takeuchi T. T., 2010, MNRAS, 406, 1830 (T10)

Takeuchi T. T., Sakurai A., Yuan F.-T., Buat V., Burgarella D., 2013, EP&S, 65, 281

Takeuchi T. T., Yoshikawa K., Ishii T. T., 2000, ApJS, 129, 1

Takeuchi T. T., Yoshikawa K., Ishii T. T., 2003, ApJ, 587, L89

Torabi S., Dourandish A., Daneshvar Kakhki M., Kianirad A., Mohammadi H., 2020, in Rashidghalam M., ed., The Economics of Agriculture and Natural Resources. Perspectives on Development in the Middle East and North Africa (MENA) Region. Springer, Singapore

Trivedi P. R., Zimmer D. M., 2005, Found. Trends Econom., 1, 1

Vallini L., Gruppioni C., Pozzi F., Vignali C., Zamorani G., 2016, MNRAS, 456, L40

Vio R., Nagler T. W., Andreani P., 2020, A&A, in press, preprint (arXiv:2006.06268)

Willmer C. N. A. et al., 2006, ApJ, 647, 853

Wright A. H. et al., 2017, MNRAS, 470, 283

Xu D., Wei Q., Elsayed E. A., Chen Y., Kang R., 2017, Qual. Reliab. Eng. Int., 33, 803

Xu M., Hua L., Xu S., 2017, Technometrics, 59, 508

Young L. M. et al., 2011, MNRAS, 414, 940

Yuan Z., Wang J., Worrall D. M., Zhang B.-B., Mao J., 2018, ApJS, 239, 33

This paper has been typeset from a TEX/LATEX file prepared by the author.