Environmental and Ecological Statistics (2022) 29:315–338
https://doi.org/10.1007/s10651-021-00525-0

Modelling multivariate data using product copulas and minimum distance estimators: an exemplary application to ecological traits

Eckhard Liebscher[1] · Franziska Taubert[2] · David Waltschew[1] · Jessica Hetzer[2]

Received: 28 May 2021 / Accepted: 20 November 2021 / Published online: 4 January 2022

Abstract

Modelling and applying multivariate distributions is an important topic in ecology. In particular in plant ecology, the multidimensional nature of plant traits comes with challenges such as wide ranges in observations as well as correlations between several characteristics. In other disciplines (e.g., finances and economics), copulas have been proven as a valuable tool for modelling multivariate distributions. However, applications in ecology are still rarely used. Here, we present a copula-based methodology of fitting multivariate distributions to ecological data. We used product copula models to fit multidimensional plant traits, on example of observations from the global trait database TRY. The fitting procedure is split into two parts: fitting the marginal distributions and fitting the copula. We found that product copulas are well suited to model ecological data as they have the advantage of being asymmetric (similar to the observed data). Challenges in the fitting were mainly addressed to limited amount of data. In view of growing global databases, we conclude that copula modelling provides a great potential for ecological modelling.

Keywords Multivariate distributions · Copula · Plant traits · Trait distributions · Vegetation modelling

Handling Editor: Pierre R. L. Dutilleul.

✉ Eckhard Liebscher
eckhard.liebscher@hs-merseburg.de

[1] Department of Engineering and Natural Sciences, University of Applied Sciences Merseburg, Eberhard-Leibnitz-Straße 2, 06217 Merseburg, Germany

[2] Department of Ecological Modelling, Helmholtz Centre for Environmental Research GmbH - UFZ, Permoserstraße 15, 04318 Leipzig, Germany

1 Introduction

In the last two decades the concept of copulas has become well-established for describing multivariate relationships between several attributes of a system. Copula models have been successfully applied in many fields, for instance in the fields of finance (Embrechts 2009), econometrics (Fan and Patton 2014), system reliability (Zhang and Wilson 2016) and astronomy (Vio et al. 2020) giving deep insights in the multivariate structure of the data. Copula-based approaches allow to split the structure of multivariate distributions into two parts: first, the one-dimensional marginal distributions of the attributes separately, and secondly, the copula which provides a complete description of the dependencies of all attributes among each other. For example in finance, copula approaches enables to estimate the risk of port-folios of assets more precisely, by providing insights on the dependencies between assets.

Recent studies have emphasized the potential of copulas for modelling multivariate distributions in ecology (Ghosh et al. 2020, and 2021; Anderson et al. 2019). Copula approaches have been utilized to model the relationship between several environmental quantities. Thereby, studies broaden knowledge on copulas presenting methodical frameworks to use copulas (Anderson et al. 2019), and discussing new findings and challenges applying them in ecology (Ghosh et al. 2020). She and Xia (2018) modelled the relationship between drought duration and severity by standard Archimedean copulas to investigate drought processes on the Loess Plateau of China, see also Dayal et al. (2020). In Anderson et al. (2019), fish population abundances are considered. Emura and Michimae (2017) analyzed data of salamander metamorphoses where the copula describes the dependence of the two variables in the censoring setup. The paper by Chang and Joe (2019) deals with vine copulas and their application to abalone data, see also Michimae et al. (2020). Since copulas have only been applied to a limited number of ecological field data yet, insights are still lacking on the generality of those statements, concerning different modelling approaches as well as regarding different types of observations.

Earlier studies mostly considered Archimedean or/and elliptical copulas (Ghosh et al. 2020; Anderson et al. 2019; She and Xia 2018 among others). One main characteristic of these copulas is that they are exchangeable, which means that the copula remains the same if the components are permuted. As a consequence of exchangeability, their bivariate marginal distributions are symmetric with regard to the main diagonal, and all bivariate correlations are identical. However, scatterplots and empirical correlations of practical data often do not show these symmetries. Recent studies show that copulas in ecology are typically asymmetric (see for instance Ghosh et al. (2020), Dayal et al. (2020)). In the paper by Ghosh et al. (2020), the rationale behind the asymmetries in the empirical copula is investigated. Here asymmetry of copulas means that they do not feature symmetric properties explained in Section 5.1. In this paper, we employ product copula models according to Liebscher (2008) to overcome the disadvantages of Archimedean or elliptical copula models. We show the great potential of product copulas for modelling the distribution of ecological data.

The aim of this paper is twofold: first we introduce a copula-based methodology of fitting multivariate distributions. We consider data vectors of an arbitrary dimension contrary to most papers on this subject. Secondly, we analyze data from a widely used ecological database of vegetation attributes to give an example of how to utilize the proposed methodology. In contrast to other papers, the model parameters are fit by means of a minimum distance method by use of Cramér-von-Mises divergence. This method exhibits some advantages. We do not need copula densities having a sophisticated and numerically inconvenient structure in many cases, especially in higher dimensions and for product copulas. Moreover, the minimum distance method provides an approximation coefficient assessing the fit quality. Earlier studies have focused on the maximum likelihood method to fit the copulas (see Hofert et al. (2021) and Ghosh et al. (2020), for example) which appears to be an alternative method. The maximum likelihood method gives theoretically best results whenever the underlying distribution belongs to the model family. The latter is rather seldom the case in practice. The minimum Cramér-von-Mises distance estimation method puts emphasis on fitting the copula rather than parameters, see Weiß (2011) for a comparison of several estimation methods.

The proposed approach is used to describe dependencies of a plant community. In vegetation ecology, a community of interacting plants can be described by the distribution of abundant plant traits. Plant traits (measured at the individual plant level) specify morphological, anatomical, biochemical, physiological or phenological properties of a plant (Violle et al. 2007) and characterize, for example, growth dynamics and functions of the vegetation (Kattge et al. 2011, 2020; Violle et al. 2007). The development of large databases like the global plant trait network GlopNet (Wright et al. 2004), the plant trait data-base for invasive species BiolFlor (Kü hn et al., 2004), the LEDA data-base of life-history traits (Kleyer et al. 2008), and the global plant trait database TRY (Kattge et al. 2011, 2020) enabled new insights on plant properties at the local scale, across biomes and to the global scale. Earlier studies found that the (marginal) distributions of plant traits tend to be lognormal (Kattge et al. 2011), but can markedly change due to climate (Butler et al. 2017), management regime (Herz et al. 2017), and fine-scale soil conditions (Bruelheide et al. 2018).

To investigate links between plant traits, one typical approach was to correlate empirical trait data pairwise (e.g., by using linear or power-law regression) and investigate the coefficient of determination (Paul et al. 1999; Reich et al. 1992, 1997). Theories about plant strategies (e.g., (Grime 1979)) were tested and revealed trade-offs between plant traits related to environmental conditions. For example, plants with large seeds tend to have a higher chance to establish in a community. In turn, the number of their seeds is usually lower resulting in a reduced probability that soil conditions promote their establishment (Heisse et al. 2007). More recent studies investigated plant traits not only pairwise but also in a higher dimension (Wright et al. 2004) and showed trade-offs in multivariate distributions.

However, going to higher dimensions is still challenging. Firstly, datasets can have limitations in the number of pairwise corresponding trait samples for each species. A method to reduce the range of attributes to those that have a high influence is, for example, the Principal Component Analysis (PCA) (Lever et al. 2017). Studies revealed plant traits associated with resource management (leaf traits that determine capture, usage and release of resources like light, carbon and nitrogen) as well as size related traits (e.g., leaf size, canopy height) to mainly determine plant ecosystem functioning (Díaz et al. 2004).

Second, skewed (marginal-) distributions of traits can span large ranges. In contrast to other multivariate statistics, which exclude outer ranges of observed values as outliers, copulas are able to cover any distribution (Anderson et al. 2019), not only separately but also combined in a variable number of dimensions.

Here we used a Germany-wide dataset of measured plant traits with a focus on 20 herbaceous species (Herz et al., 2017) and investigate marginal distributions of plant traits independently and combined in copulas to answer the following questions:

- Which statistical marginal distributions can be used to describe plant attributes?
- How do plant traits differ in their marginal distributions?
- Which plant traits correlate with each other under the given marginal distributions?

The analyzed dataset is part of the global trait database TRY (Kattge et al. 2011, 2020) which covers plant ecosystems like forests, croplands and grasslands worldwide. We introduce a new methodology of modeling multivariate distributions of data vectors from natural science applications. This methodology is rather general and flexible for data with diverse origins, and is demonstrated here on the example of the TRY database. By this, we developed a novel copula not modeled so far.

The paper is structured as follows: In Section 2 we present a description of the database used for numerical studies. Section 3 is devoted to the modelling of marginal distributions where the focus is on the Weibull distribution and on the Gamma distribution. Modelling the copula is subject of Section 4. Details to the results are collected in the appendices.

2 Ecological trait data: An exemplary data set from TRY

TRY plant trait database is one of the main plant trait databases (Kattge et al. 2011, 2020). The globally collected data includes plant attributes on an individual level across different ecosystems such as forests, crops and grasslands, and across climatic gradients.

In this study, we used a dataset on individual plant traits of 20 species (10 grass species and 10 forbs) in managed grasslands (Herz et al. (2017), two censuses measured in 2014 and 2015). Plant traits were measured across a climatic gradient in Germany covered by three study sites of the German Biodiversity Exploratories ((Fischer et al. 2010), Schorfheide-Chorin, Hainich, Schwäbische Alb). We focused the analysis on five plant traits explained in Table 1.

RSR was calculated as the fraction of the measured ‘dry mass aboveground biomass’ [g] and ‘dry mass roots total’ [g] (Herz et al. 2017). SLA was determined as leaf area per leaf dry mass. LCNC was calculated by the ratio of leaf carbon concentration to leaf nitrogen concentration and RCNC analogously by the ratio of root carbon concentration and root nitrogen concentration. HW was determined by the ratio of plant standing height [mm] and the maximum of plant diameter in north-south direction [mm] and east-west direction [mm]. All traits were derived for each census and each study site separately. The analyzed data are presented in a shortened form in Table 4.

Table 1 Table of attributes

Trait	Description	Unit
RSR	Ratio of aboveground plant biomass to belowground biomass	–
SLA	specific leaf area	m^2/kg
RCNC	Carbon–nitrogen ratio of roots	–
LCNC	Carbon–nitrogen ratio of leaves	–
HW	Height–width ratio of plant	–

3 Multivariate distributions

Let $Y = (Y^{(1)}, \dots, Y^{(d)})^T$ be a d-dimensional random vector having the joint distribution function H:

$$H(y_1, \dots, y_d) = \mathbb{P}\left\{Y^{(1)} \le y_1, \dots, Y^{(d)} \le y_d\right\} \quad \text{for } y_i \in \mathbb{R}.$$

In our context $Y^{(1)}, \dots, Y^{(d)}$ are the measured values of the traits. They are comprised in the data vector Y. All probabilities related to the random vector Y can be computed in terms of H. We denote the marginal distribution function of $Y^{(j)}$ by F_j ($j = 1, \dots, d$):

$$F_j(t) = \mathbb{P}\left\{Y^{(j)} \le t\right\} \quad \text{for } t \in \mathbb{R}.$$

We assume that the marginal distributions are continuous. Then f_j denotes the marginal density of $Y^{(j)}$:

$$f_j(t) = F_j'(t) \quad \text{for } t \in \mathbb{R}.$$

Sklar's theorem (Sklar, 1959) implies that

$$H(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)) \quad \text{for } y_i \in \mathbb{R}. \tag{1}$$

This formula shows that the joint distribution function comprises two parts: the marginal distribution functions $F_1, \dots, F_d$ and the so-called copula C. The copula C describes the dependence of the attributes irrespectively of the marginal distributions. Therefore, the methodology is divided into two parts related to the modeling of the marginal distributions and the copula, respectively. The challenge of the approach is to find suitable parametric models for $F_1, \dots, F_d$ and C. Using these functions, probabilities of rectangles $[y_1, z_1] \times \dots \times [y_d, z_d]$ can be evaluated by

$$\begin{aligned}&\mathbb{P}\left\{y_1 \le Y^{(1)} \le z_1, \dots, y_d \le Y^{(d)} \le z_d\right\}\\ &\quad = \sum_{\delta_1=0}^{1} \dots \sum_{\delta_d=0}^{1} (-1)^{d+\delta_1+\dots+\delta_d} C(F_1(w_{\delta_1,1}), \dots, F_d(w_{\delta_d,d})),\end{aligned}$$

where $w_{0,j} = y_j$, $w_{1,j} = z_j$. This formula shows how to compute probabilities using $F_1, \ldots, F_d$ and C.

In the next section we consider the problem of modelling and fitting the marginal distributions by use of the data. Modeling and fitting the copulas is then the subject of Section 5.

4 Analysis of the marginal distributions

4.1 Methodology of fitting the marginals

In a first step the data is divided into several subsets. Each subset corresponds to one class, here to one species. m_c denotes the median (alternatively the mean) of class c. In the following we consider the measured values of trait j. Let $Z_1, \ldots, Z_n$ be the data sample of the considered trait; i.e. Z_i is the i-th measured value of the trait under consideration. $c(1) \ldots c(n)$ denote the classes of the sample items. For each class c, we compute the empirical median $\hat{m}_c$.

In the second step we normalize the data by:

$$\tilde{Z}_i = \frac{Z_i}{\hat{m}_{c(i)}} \tag{2}$$

for $i = 1, \ldots, n$ in order to make the data comparable. Normalized data have the property that the empirical median equals 1.

In the last step, the Weibull distribution or the gamma distribution (see Appendix A) is fit to the transformed data $\tilde{Z}_1, \ldots, \tilde{Z}_n$ using the maximum-likelihood method. According to the maximum-likelihood method, the estimated parameters of θ_j are evaluated by:

$$\hat{\theta} = \arg\max_{\theta \in \Theta_1} \sum_{i=1}^{n} \ln f_j(\tilde{Z}_i \mid \theta) \tag{3}$$

where $f_j(. \mid \theta)$ denotes density model of trait j and f_j is the Weibull or the gamma distribution. In case of Weibull or gamma distribution, no explicit formulas for the estimators are available. The minimization problem (3) has to be solved numerically deploying an efficient optimization algorithm (see e.g. Meeker and Escobar (1998), Chapters 8 and 11). To find the best model, one can compare the values of the Akaike and the Bayesian information criterion (AIC, BIC). The smallest value gives the best model in the considered case.

Finally, $f_j(. \mid \hat{\theta})$ is the best-fit model density for the normalized marginal data. For the original data, the best-fit density for the original marginal data of class c is $t \mapsto \hat{m}_c^{-1} f_j(t\, \hat{m}_c^{-1} \mid \hat{\theta})$ and its median is about $\hat{m}_c$.

The computations were performed using R software. For distribution fitting and goodness-of-fit, the packages MASS (function fitdistr) and goftest are available in R.

4.2 Results for marginal distributions of the observed data

Looking at the density plots of the variables for the various species, we see that shapes of these curves are rather similar, see Figure 5 in Appendix A2 depicting the densities of the variable RSR. Further in case of some species, there is a sizable difference between the mean and the median being an indicator for the presence of outliers (cf. Table 5 in Appendix A2, exemplified for variable RSR and several grass species). We found similarities between species in density plots of all tested traits. Hence, the whole transformed data of any trait can be regarded all as coming from the same statistical population of the trait. Its distribution can be considered as a baseline one.

The results of the baseline distribution fitting are provided in Tables 6 and 7, see Appendix A2. All the estimated distribution models are checked using the Anderson-Darling goodness-of-fit test (AD-Test in Tables 6,7). All checks were successful since all p-values are higher than 0.05. Additionally, the AIC and BIC show mostly a small difference such that both models, gamma and Weibull distribution, are reasonable for modeling marginal distributions in most cases. Throughout the results reveal that the estimated models give good approximations of the marginal distributions.

4.3 Discussion

The mean is widely used as location parameter of the distribution. If outliers are present in the dataset, then the median should be preferred over the mean as description of the location since the median represents a more robust statistic. Because of the variety of the species medians m_j , normalization according to (2) makes the trait values comparable.

In our investigations concerning the marginal distributions, we found that two types of one-dimensional distributions are relevant in the context of the TRY database: the Weibull distribution and the Gamma distribution. These two distribution families play an important role in modelling positive random variables in the framework of biology and ecology. This is shown in the papers by Taubert et al. (2013), by Hagey et al. (2016) and by de Freitas Costa et al. (2021), for example. The Anderson-Darling goodness-of-fit test is proved to be a powerful method for testing how good the distribution fit is, see Stephens (1986). In this book alternative tests are discussed such as the Kolmogorov test and the Cramér-von Mises test.

5 Methodology of modeling and fitting the copula

5.1 Basics of copulas

Formula (1) provides a link between the joint distribution function H of the random vector $Y = (Y^{(1)}, \ldots, Y^{(d)})^T$ and the copula. Suppose that the marginal distributions are continuous (i.e. F_m has a density f_m). Then the copula C is uniquely determined.

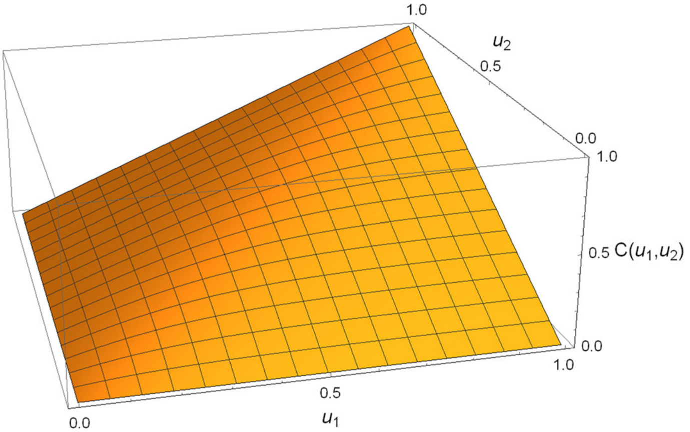

Fig. 1 three-dimensional plot of a bivariate Frank copula $C(u_1, u_2)$ with parameter 6, see Table 8 of copula models

A function $C : [0, 1]^d \to [0, 1]$ is a *copula* if it has the following properties (cf. Nelsen 2006, p. 45):

$$
\begin{aligned}
&1)\ C(u) = 0 \text{ if } u_j = 0 \text{ for any } j,\\
&2)\ C(u) = u_j \text{ if } u_l = 1 \text{ for all } l \neq j \text{ and any } j,\\
&3)\quad \sum_{\delta_1=0}^{1} \dots \sum_{\delta_d=0}^{1} (-1)^{d+\delta_1+\ldots+\delta_d} C(u + \delta \cdot h) \geq 0
\end{aligned}
$$

for all $u, h \in [0, 1]^d : u + h \in [0, 1]^d$, where $\delta = (\delta_1, \dots, \delta_d)^T, u = (u_1, \dots, u_d)^T$.

Condition 3) ensures that C is a multivariate distribution function. This implies that C is increasing in each component. Moreover, according to condition 2), C has marginals which are uniformly distributed on [0, 1] (see Fig. 1). Concerning the theory of copulas we refer to the monograph by Nelsen (2006).

We introduce

$$U_j = F_j(Y^{(j)}) \quad \text{for } j = 1 \dots d.$$

Then U_j has a uniform distribution on [0, 1]. The random vector $U = (U_1, \dots, U_d)^T$ has the distribution function C. Thus the random vector U involves the information about the dependence of the attributes. In bivariate copula plots this idea is utilized.

There are several copula-based measures describing the dependence of two components. Spearman's ρ_S and Kendall's τ are the most popular coefficients in this framework. The formulas for the population version of the coefficients are given by

$$\rho_S = 12 \int_0^1 \int_0^1 C(u, v)\, \mathrm{d}u\mathrm{d}v - 3,$$

$$\tau = 4 \int_0^1 \int_0^1 C(u, v)\, \mathrm{d}C(u, v) - 1,$$

where C is the copula of two attributes under consideration. Both association coefficients can be estimated by use of the data.

Two multivariate copulas are of special interest, especially in Proposition 5.1:

$$\begin{aligned} \textit{independent copula}: \quad & \Pi(u) = u_1 u_2 \ldots u_d, \\ \textit{comonotonicity copula}: \ & M(u) = \min\{u_1, u_2, \ldots, u_d\} \end{aligned}$$

The copula M is also known as Fréchet-Hoeffding upper bound copula. If U has distribution M, then $U_1 = U_2 = \ldots = U_d$ holds with probability 1. Theorems 2.10.14 and 5.1.9 in Nelsen (2006) lead to the following proposition:

Proposition 5.1 *1) If $Y^{(1)}, \ldots, Y^{(d)}$ are independent then*

a) $U_1, \ldots, U_d$ are independent and $C = \Pi$,

b) all bivariate Spearman's ρ_S and Kendall's τ of two components of Y are equal to 0.

c) If $C = \Pi$ then $Y^{(1)}, \ldots, Y^{(d)}$ are independent.

2) Each of the variables $Y^{(1)}, \ldots, Y^{(d)}$ is a strictly increasing function of any of the others with probability 1 if and only if $C = M$. Furthermore, in this case, all Spearman's ρ_S and Kendall's τ of two components of Y equal 1.

From the point of view of the visualization of the distribution and the dependence, a diagram of the copula density gives more information, see figure below. The copula density is evaluated by

$$c(u) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C(u) \text{ for } u \in [0, 1]^d.$$

Several concepts of symmetry of copulas can be distinguished (see Nelsen (1993)). We mention only two of them here. A copula is referred to as *exchangeable*, if

$$C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$$

for any permutation $(\pi(1), \ldots, \pi(d))$ of $(1, \ldots, d)$ and all $u \in [0, 1]^d$. Interchanging two components u_j leads to the same copula. Especially in the case $d = 3$, exchangeability means

$$\begin{aligned} & C(u, v, w) = C(v, u, w), \quad C(u, v, w) = C(u, w, v), \\ & C(u, v, w) = C(w, v, u) \text{ for } u, v, w \in [0, 1]. \end{aligned}$$

Corresponding equalities hold true for copula densities. Exchangeability is one kind of symmetry of a copula. As a consequence, all bivariate correlations of exchangeable copulas are identical. All the commonly used copula models of Table 8 in Appendix B are exchangeable.

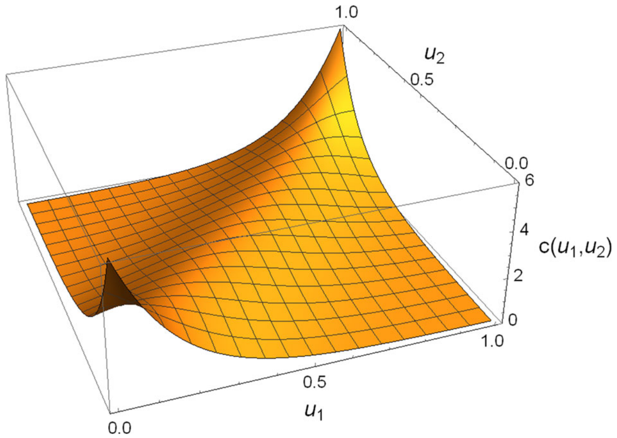

Fig. 2 three-dimensional plot of the density $c(u_1, u_2)$ of a bivariate Frank copula with parameter 6

A copula is called *radially symmetric (or tail symmetric)*, if $(U_1, \dots, U_d)^T$ and $(1 - U_1, \dots, 1 - U_d)^T$ have the same distribution. In the case $d = 2$, bivariate copulas are radially symmetric if C coincides with the survival copula:

$$C(u, v) = u + v - 1 + C(1 - u, 1 - v) \text{ for } u, v \in [0, 1].$$

The bivariate Frank copula is the only one of Table 8 which is radially symmetric. Elliptical copulas are radially symmetric, too.

5.2 Product copula models

The classical copulas of Table 8 in Appendix B are used as basic models for the construction of product models. The theory of this model class of asymmetric copulas was established in Liebscher (2008). Let C_1 and C_2 be two copulas. Then the formula for the product copula is given by

$$C(u) = C_1(u_1^{\alpha_1}, \dots, u_d^{\alpha_d}) \cdot C_2(u_1^{1-\alpha_1}, \dots, u_d^{1-\alpha_d}) \tag{4}$$

where $\alpha_1, \dots, \alpha_d \in [0, 1]$ are parameters inducing the asymmetry of C. In summary, the copula in (4) has $d + 2$ parameters: the two parameters of the copulas C_1, C_2, and the exponents $\alpha_1, \dots, \alpha_d$. Identity (4) gives the basic model for fitting it to ecological data from TRY database. The copulas C_1 and C_2 will be taken then from Table 8.

Notice that most of the copulas of Table 8 have pairwise positive correlation. If negative correlations occur, then one should transform some variables to obtain pairwise positive correlations. To achieve this, one simple idea would be to consider the negative of some variables instead of these variables. The methodology of fitting is the subject of the next sections.

5.3 Fitting the copula

Let $Y_1, \dots, Y_n$ be the sample of d-dimensional random vectors with distribution function H and copula C. Vector $Y_i = (Y_{i1}, \dots, Y_{id})^T$ includes the normalized values of the traits of the i-th sample item (i-th measurement). Y_i is the normalized i-th row of the data matrix (see Table 4). Note that the data preprocessing described in Section 4.1 has to be performed before the copula fitting. We denote the joint empirical distribution function by $\hat{H}_n$ (estimator for H):

$$\hat{H}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Y_{i1} \le y_1, \dots, Y_{id} \le y_d\}$$

for $y = (y_1, \dots, y_d)^T \in \mathbb{R}^d$, where $\mathbf{1}\{A\}$ is the indicator of the event A; i.e. $\mathbf{1}\{A\} = 1$ if A occurs and $= 0$ if A does not occur. Let $\tilde{F}_n(y) = (F_{1n}(y_1), \dots, F_{dn}(y_d))^T$ be the vector of the marginal empirical distribution functions

$$F_{jn}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left\{Y_{ij} \le t\right\}$$

for $t \in \mathbb{R}$. The values $F_{jn}(Y_{1j}), \dots, F_{jn}(Y_{nj})$ represent a permutation of the relative ranks $\{i/n : 1 = 1 \dots n\}$, and they mimic so the uniform distribution. Now a copula plot of the j-th and the μ-th component can be drawn to show the dependence of both variables. The plot contains the points $(F_{jn}(Y_{1j}), F_{\mu n}(Y_{1\mu})), \dots, (F_{jn}(Y_{nj}), F_{\mu n}(Y_{n\mu}))$ (see Figure 6, Appendix D).

The empirical copula $\hat{C}_n$ estimates the copula and its formula is given by

$$\hat{C}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{F_{1n}(Y_{i1}) \le u_1, \dots, F_{dn}(Y_{id}) \le u_d\}$$

for $u \in [0, 1]^d$. In the case $d = 2$, $\hat{C}_n(u_1, u_2)$ is just the relative frequency of points in the copula plot lying in $[0, u_1] \times [0, u_2]$.

We consider now the family $\{C_\gamma : \gamma \in \Theta\}$ of copulas where γ is the q-dimensional parameter vector and $\Theta \subset \mathbb{R}^q$ the parameter space. To fit this family to the given data, we search for a copula $C_{\hat{\gamma}}$ of the family that fits best to the data. In this paper the fit of the copulas is based on the *Cramér-von-Mises divergence*, see Liebscher (2009). A divergence is something like a distance, but it does not fulfil some of the mathematical requirements of a distance. The estimated *Cramér-von-Mises divergence* is given by:

$$\widehat{\mathcal{D}}_n(C_\gamma) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{H}_n(Y_i) - C_\gamma(\tilde{F}_n(Y_i))\right)^2 \tag{5}$$

for $\gamma \in \Theta$. This divergence measures the distance between the nonparametrically estimated distribution function and the parametric model of it. It is an interesting feature that we do not need $\hat{C}_n$ in (5). Fitting the copula model means minimizing $\widehat{\mathcal{D}}_n(C_\gamma)$ with respect to the parameter γ. The minimizer $\hat{\gamma}$ of $\gamma \mapsto \widehat{\mathcal{D}}_n(C_\gamma)$ is called a *minimum-distance estimator* for γ (cf. Tsukahara 2005; Liebscher 2009) if the following identity holds true:

$$\widehat{\mathcal{D}}_n(C_{\hat{\gamma}}) = \min_{\gamma \in \Theta} \widehat{\mathcal{D}}_n(C_\gamma). \tag{6}$$

Assessing the goodness-of-approximation is discussed in Liebscher (2015). For comparisons, we compute the *approximation coefficient*:

$$\hat{\rho} = 1 - \frac{\widehat{\mathcal{D}}_n(C_{\hat{\gamma}})}{\widehat{\mathcal{D}}_n(\Pi)}. \tag{7}$$

$\widehat{\mathcal{D}}_n(\Pi)$ is the Cramér-von-Mises divergence (5) for the independent copula Π introduced in Section 5.1. Since $\hat{\rho} \leq 1$ holds, the coefficient $\hat{\rho}$ is a normed measure and can be regarded as a counterpart to the coefficient of determination in regression. If $\Pi \in \{C_\gamma\}$ then $\hat{\rho} \in [0, 1]$. Ideally, $\hat{\rho}$ is close to 1 indicating a perfect fit.

Using R software the computations for fitting the copulas were performed. We deployed the function optim (R package stats) for optimization.

5.4 Discussion

There are several alternative construction concepts for asymmetric copulas. Among them, nested Archimedean Copulas (McNeil 2008) are asymmetric in general but they have the disadvantage being exchangeable in some components, see also Grimaldi and Serinaldi (2006).

We apply the Cramer-von Mises divergence to fit the copulas. Alternatively, one can use maximum-likelihood estimation for fitting. This may lead to rather different parameters in comparison to minimum-distance method especially in the often occurring case that C does not belong to the family $\{C_\gamma : \gamma \in \Theta\}$. Maximum-likelihood estimation has the disadvantage that one needs the density of the model copula. According to formula (5), the Cramér-von-Mises divergence needs only the empirical distribution functions $\hat{H}_n, F_{1n}, \ldots, F_{dn}$ for computations. Concerning estimation, we refer to Hofert et al. (2012) and Liebscher (2015).

6 Fitting copulas to the observed ecological data

6.1 Results

First we give some comments on the choice of the data. To comprise species with similar bivariate correlations, we concentrate on the grass species Al.pr, An.od, Ar.el, Av.pu, Cy.cr, Da.gl, Fe.pr, Lo.pe and the herb species Ga.mo, Ru.ac, Ce.ja, Ga.ve, Pl.la (see Appendix C for abbreviations of species). From a glance at the correlations, it is evident that variable RCNC can be regarded as independent of the remaining variables

Table 2 Fitting results (copula of variables RSR, SLA and minus HW): the approximation coefficient $\hat{\rho}$ (defined in (7)) for various basic copula models abbreviated according to Table 8

Category	n	Copula	$\hat{\rho}$	Category	n	Copula	$\hat{\rho}$
Grass	546	F	0.924	Herbs	356	AMH	0.864
		GH	0.921			C	0.860
		AMH	0.906			C	0.848

Table 3 Fitting results (copula of variables RSR, SLA and minus HW): the approximation coefficient $\hat{\rho}$ (defined in (7)) for various product copula models with copulas C_1 and C_2 abbreviated according to Table 8

category	n	C_1	C_2	$\hat{\rho}$	Category	n	C_1	C_2	$\hat{\rho}$
Grass	546	J	F	0.974	Herbs	356	C	C	0.907
		F	GH	0.972			C	F	0.905
		GH	GH	0.970			C	Π	0.904

(confidence intervals cover zero), see Table 9 in Appendix B2. The variable LCNC is omitted because of a lack of sufficient amount of data. Therefore, in this section, we consider the 3-dimensional distribution of the normalized data vector $(Y^{(1)}, Y^{(2)}, Y^{(3)})$ including the variables RSR, SLA, and minus HW (normalization according to Section 4.2).

Next we fit copulas C from Table 8 and product copulas according to (4) to the data vector of the variables RSR, SLA and minus HW. The copulas C_1 and C_2 in (4) are taken from Table 8. $Y^{(3)}$ is taken to be the negative of HW in order to obtain positive bivariate correlations (see Table 12 for correlations) which are present for most of the copula models in Table 8. The parameter vector of product copulas consists of the parameters γ_1, γ_2 of C_1 and C_2, and the exponents $\alpha_1, \alpha_2, \alpha_3$. The best fitting results are given in Tables 2 and 3, in Tables 10 and 11 in more detail.

Approximation coefficients greater than 0.9 indicate a very good fitting accuracy, and the best fits of Tables 2 and 3 exhibit this property. Next we focus on the grass data. The estimated copula density of the components SLA and minus HW depicted in Figure 3 reveals a significant asymmetry with respect to the main diagonal. This effect is especially shown in Figure 4 where the difference $\Delta(u_1, u_2) = c(u_1, s(u_1)) - c(s(u_1), u_1)$ $(u_1 \in [0, 1])$ of copula density values along cross sections $u_2 = s(u_1)$ is depicted. In case of a symmetrical bivariate density, we have $C(u, v) = C(v, u)$ for $u, v \in [0, 1]$, and therefore $\Delta \equiv 0$.

Next we show the asymmetry of the distribution of RSR, SLA and minus HW using probabilities and tail indices λ_L, and λ_U (for the definition see Ghosh et al. (2020), p. 420). For this purpose, we calculate the following two quantities

$$\gamma_1 = \mathbb{P}\{0.75 \leq U_1 \leq 1, 0 \leq U_2 \leq 0.25\} = 0.25 - C(0.75, 0.25),$$
$$\gamma_2 = \mathbb{P}\{0 \leq U_1 \leq 0.25, 0.75 \leq U_2 \leq 1\} = 0.25 - C(0.25, 0.75).$$

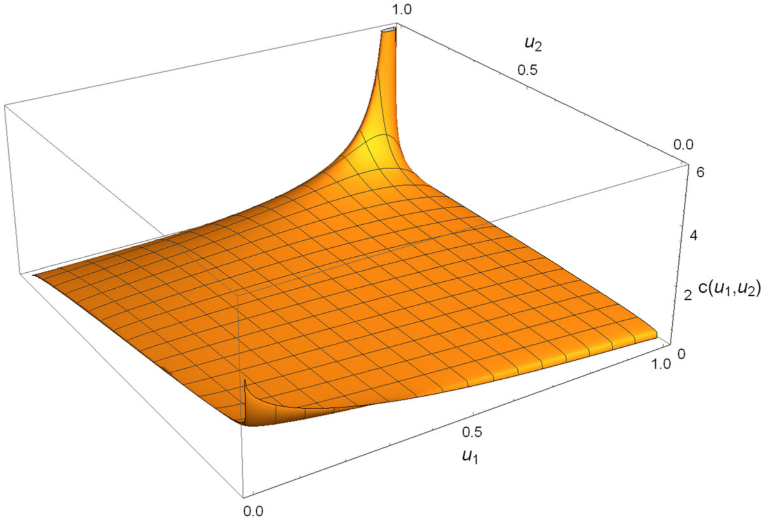

Fig. 3 estimated bivariate copula density of RSR and SLA, product model J-F from Table 3

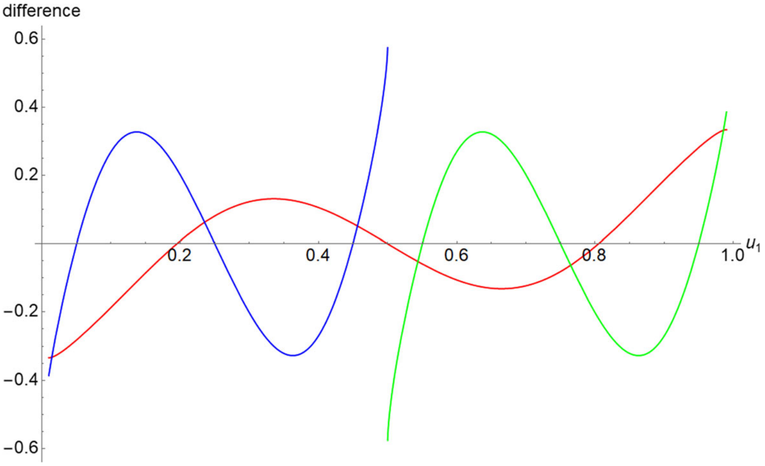

Fig. 4 sections parallel to the secondary diagonal (section along $u_2 = 1 - u_1$ in red, section along $u_2 = 1.5 - u_1$ in green, section along $u_2 = 0.5 - u_1$ in blue) of the estimated bivariate copula density of RSR and SLA, model J-F from Table 3

In case of a symmetric copula, γ_1 and γ_2 are identical. The corresponding results are given in Table 12 together with the empirical bivariate Spearman correlations and these correlations in the estimated copula model. The results provided in Table 12 exhibit that the multivariate distribution is non-exchangeable and with regard to the tails (upper and lower). The correlations are significantly different which is not the case for the commonly used basic models (see Table 8, Appendix B1). The noticeable difference between γ_1 and γ_2 for the pair SLA and -HW gives a further evidence for asymmetry. Furthermore, we see from Table 12 that the combination of small SLA and small HW is less probable in comparison to large SLA and large HW. Table 12 reveals that the difference between empirical correlations and estimated model-driven ones are reasonably small showing the good quality of estimation results. On the other hand, bivariate Spearman's ρ_S coefficients differ from each other significantly which also exhibits the asymmetry of the multivariate distribution.

6.2 Discussion

Looking at the results of the previous section, the dependence structure of the copula fit reflects the asymmetric dependence structure in the data. Similar results can be obtained by incorporating variable LCNC. Since symmetric copulas exhibit identical Spearman coefficients for all pairs of variables, it is comprehensible that the use of asymmetric copulas gives better fitting results in our context. A further analysis showed that the dependence structure varies over the species. To identify these differences more data are needed. We omitted to perform goodness-of-fit tests for several reasons. In higher dimensions they are hard to apply because of a sophisticated structure of variances of test statistics (often bootstrapping has to be applied). Moreover, goodness-of-fit tests do not yield a ranking of the models under consideration based on appropriate statistics.

Typically, copulas are fitted using big data (for example in finance sciences using daily stock exchange data). Ecological plant measurements are limited because of high efforts for data gathering. In our study, we used a state-of-the art data set, having (60-80 samples per species). If sample size is too small, then there will be a degree of uncertainty in the results.

7 Conclusions

The paper provides a methodology for analyzing environmental data. The description of the multivariate distribution is split up into the one-dimensional marginal distributions and the copula. In the case of the example data set from the TRY database, the data analysis exhibits good to very good accuracy of fitting results. The results have demonstrated that copulas are a valuable tool to analyze ecological data. Product copulas according to Liebscher (2008) prove to be flexible model classes for accurate fitting even in the case of higher dimensions $d > 2$. Product copulas have not been considered up to now in the context of ecological data to our knowledge. They are also flexible to describe asymmetries in several directions occurring in most applications.

We see challenges in the accuracy, due to the limited number of samples and mainly caused by the exclusion of samples in which some traits were missing. General approaches on how to deal with missing data would be a first step towards more reliable estimates. In light of the increasing effort in observing traits reflected by growing trait databases (e.g., TRY) we see a high potential to combine ecological data with copulas to gain deeper insights in measured characteristics.

Acknowledgements The authors are very grateful to the referees for their valuable suggestions and helpful comments.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix A1: Models for marginal distributions

In Appendix A1 we introduce parametric models for the marginal distribution functions F_j and their density f_j, $j = 1, \dots, d$. In the paper the focus is on the following two specific classes of distributions:

1) *Weibull distribution* with shape parameter β and scale parameter $\tau : \beta, \tau > 0$: The density is given by

$$\varphi_\theta(t) = \begin{cases} \tau^{-\beta} \beta t^{\beta-1} e^{-(t/\tau)^\beta} & \text{for } t \geq 0 \\ 0 \quad \text{for } t < 0. \end{cases}$$

The parameter τ is the scale parameter having the property that with a probability of 63.2%, the random variable is smaller than τ. The expectation is equal to $\tau\Gamma(1+1/\beta)$ where Γ is the Gamma function.

2) *Gamma distribution* with shape parameter β and scale parameter $\tau : \beta, \tau > 0$: The density reads as

$$\varphi_\theta(t) = \begin{cases} \frac{1}{\tau^\beta \Gamma(\beta)} t^{\beta-1} \exp(-t/\tau) & \text{for } t \geq 0 \\ 0 \quad \text{for } t < 0. \end{cases}$$

The expectation is equal to $\beta\tau$.

For both models, $\theta = (\beta, \tau)^T \in \Theta_1$ is the parameter vector where $\Theta_1 = (0, \infty) \times (0, \infty)$. The corresponding distribution function is evaluated by

$$\Phi_\theta(t) = \int_0^t \varphi_\theta(u) du.$$

Appendix A2: Data structure and results for marginal distributions

Table 4 Data structure. species according to Appendix C, year=year of mesurement, explo=location of plants, variables RSR...HW according to Section 2, NA=missing data, SCH=Schorfheide-Chorin, ALB=Schwäbische Alb

Case no.	Species	Year	Explo	Class	RSR	SLA	RCNC	LCNC	HW
1	Ac.mi	2014	ALB	Herb	2.0965517	47.813043	24.84785181	NA	1.1666667
2	Al.pr	2014	ALB	Grass	1.6901172	46.396631	37.31636093	NA	0.2692308
3	An.od	2014	ALB	Grass	10.6013513	60.880000	46.32642950	NA	0.3076923
4	Ar.el	2014	ALB	Grass	0.9147059	41.738247	56.89432999	NA	0.1153846
5	Av.pu	2014	ALB	Grass	4.1107595	40.357500	0.13203634	NA	0.5625000
6	Ce.ja	2014	ALB	Herb	1.2247557	40.233897	33.21107228	NA	0.5454545
7	Cy.cr	2014	ALB	Grass	1.6989568	28.933615	25.60340398	NA	0.5333333
8	Da.gl	2014	ALB	Grass	0.4216145	39.252406	56.42985378	NA	0.1794872
9	Fe.pr	2014	ALB	Grass	5.0397422	32.543976	33.32396372	NA	1.0000000
10	Lo.pe	2014	ALB	Grass	2.9369369	42.614286	34.85883363	NA	0.6666667
11	Pl.la	2014	ALB	Herb	1.8306773	36.148805	0.09060719	NA	0.4000000
12	Po.pr	2014	ALB	Grass	2.7578125	30.450000	0.37286034	NA	0.1304348
13	Po.tr	2014	ALB	Grass	49.4000000	19.450000	28.07972777	NA	0.3076923
14	Ra.ac	2014	ALB	Herb	4.8484849	43.162121	16.90244166	NA	2.0000000
…									…
1858	Ru.ac	2015	SCH	Herb	1.328375	25.259630	32.93124	14.48004	1.1016949
1859	Ve.ch	2015	SCH	Herb	1.326945	36.069670	25.88047	22.08081	0.5652174

Table 5 Mean and median of RSR trait for several grass species (see Appendix C for full Latin names of abbreviations for species)

Species	Number of items	Mean	Median $\hat{m}_j$
Al.pr	64	6.0758	3.8494
An.od	73	1.9031	1.3719
Ar.el	67	6.8079	2.0636
Av.pu	73	2.7255	1.7663
Cy.cr	69	1.6819	1.0523
Da.gl	94	2.4613	1.2039
Fe.pr	64	4.8886	1.4413
Lo.pe	73	1.6383	1.0231
Po.pr	73	2.8481	2.0231
Po.tr	67	5.0539	1.7865

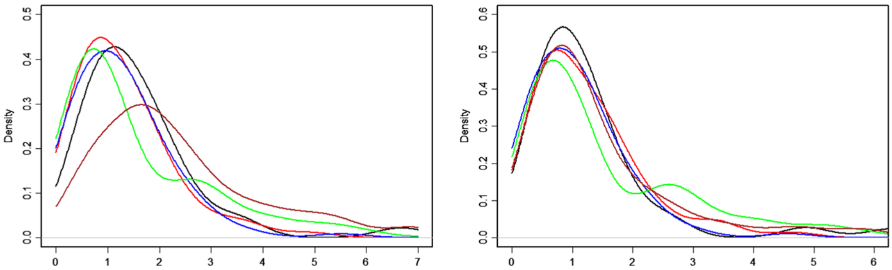

Fig. 5 non-parametric kernel density estimator for variable RSR, left: unnormalized, right: normalized; colors of the lines indicate the different species: black=An.od, red=Cy.cr, blue=Da.gl, green=Lo.pe, brown=Po.pr; for abbreviations, see Appendix C

Table 6 Detailed fitting results for marginal distributions of grass species: estimated parameters (standard errors in parentheses), p-values of the Anderson-Darling test of goodness-of-fit, values of AIC and BIC criterions for model selection

Trait	Sample length *n*	Model	Shape parameter	Scale parameter	p-value of AD test	AIC	BIC
RSR	717	Gamma	1.7854 (0.0893)	0.6955 (0.0829)	0.244	1541.1	1550.1
		Weibull	1.3466 (0.0388)	1.3621 (0.0411)	0.258	1564.8	1573.9
SLA	691	Gamma	6.055 (0.319)	0.1776 (0.0097)	0.879	734.2	743.2
		Weibull	2.3602 (0.0610)	1.2114 (0.0208)	0.873	838.8	847.9
RCNC*	730	Gamma	12.639 (0.658)	0.08398 (0.00446)	0.907	262.8	272.0
		Weibull	3.5457(0.0953)	1.1754 (0.0132)	0.410	367.4	376.6
LCNC	226	Gamma	19.032 (1.717)	0.05410 (0.00494)	0.940	−15.22	−8.379
		Weibull	4.2449 (0.1986)	1.1262 (0.0187)	0.467	26.38	33.22
HW	688	Gamma	2.6457 (0.1357)	0.4510 (0.0255)	0.432	1318.2	1327.3
		Weibull	1.6342 (0.0457)	1.3416 (0.0334)	0.434	1364.1	1373.1

REMARK: Outliers greater than 5 are removed from analysis.
* 24 very small data values < 0.1 are excluded from the analysis

Table 7 Detailed fitting results for marginal distributions of herb species: estimated parameters (standard errors in parentheses), p-values of the Anderson-Darling test of goodness-of-fit, values of AIC and BIC criterions for model selection

Trait	Sample length n	Model	Shape parameter	Scale parameter	p-value of AD test	AIC	BIC
RSR	668	Gamma	1.9231 (0.1014)	0.6280 (0.0378)	0.919	1343.7	1352.6
		Weibull	1.3957 (0.0417)	1.3329 (0.0407)	0.0372	1369.4	1378.2
SLA	632	Gamma	4.8369 (0.2636)	0.2226 (0.0128)	0.626	800.7	809.6
		Weibull	2.2560 (0.0643)	1.216 (0.0227)	0.524	847.1	856
RCNC*	690	Gamma	9.992 (0.533)	0.1052 (0.0057)	0.656	386.7	395.8
		Weibull	3.2414 (0.0886)	1.1690 (0.0147)	0.193	453.4	462.4
LCNC	228	Gamma	16.994 (1.539)	0.06023 (0.00553)	0.900	6.657	13.52
		Weibull	3.9733 (0.1804)	1.1222 (0.0198)	0.528	48.54	55.40
HW	648	Gamma	2.5584 (0.1350)	0.4627 (0.0270)	0.114	1245.4	1254.3
		Weibull	1.6210 (0.0472)	1.3297 (0.0344)	0.328	1282.4	1291.3

REMARK: Outliers greater than 5 are removed from analysis.
* 15 very small data values < 0.1 are excluded from the analysis

Appendix B1: Copula Models

Table 8 Copulas (the abbreviations are given below the family name) and their parameters

Family	Formula for $C(u)$	Parameter	$C = \Pi$ for	$C = M$ for
Clayton C	$\left(\sum_{i=1}^{d} u_i^{-\gamma} - d + 1\right)^{-1/\gamma}$	$\gamma > 0$	$\gamma \to 0$	$\gamma \to \infty$
Frank F	$-\frac{1}{\gamma}\ln\left(1 + \left(e^{-\gamma} - 1\right)\prod_{i=1}^{d}\frac{e^{-\gamma u_i}-1}{e^{-\gamma}-1}\right)$	$\gamma \neq 0$	$\gamma \to 0$	$\gamma \to \infty$
Joe J	$1 - \left(1 - \prod_{i=1}^{d}\left(1 - (1 - u_i)^{\gamma}\right)\right)^{1/\gamma}$	$\gamma \geq 1$	$\gamma = 1$	$\gamma \to \infty$
Ali-Mikhail- Haq AMH	$(1 - \gamma)\left(\prod_{i=1}^{d}\frac{1-\gamma(1-u_i)}{u_i} - \gamma\right)^{-1}$	$-1 \leq \gamma < 1$	$\gamma = 0$	-
Gumbel- Hougaard GH	$\exp\left(-\left(\sum_{i=1}^{d}(-\ln u_i)^{\gamma}\right)^{1/\gamma}\right)$	$\gamma \geq 1$	$\gamma = 1$	$\gamma \to \infty$

M is the comonotonicity copula. We refer to Hofert et al. (2012) for a long list.

Appendix B2: Results of copula fitting

Table 9 Empirical Spearman correlations with RCNC and 95%-confidence intervals for grass data

	Correlations
RSR	0.02637 [-0.05782, 0.11019]
SLA	-0.000006 [-0.084073, 0.08406]
HW	0.00223 [-0.08185, 0.08629]

Table 10 Fitting results for various basic copula models (copula of variables RSR, SLA and minus HW): estimated parameters and the approximation coefficient $\hat{\rho}$

Category	n	Copula	$\hat{\rho}$	Estimated parameter
grass	546	F	0.924	1.5377
		GH	0.921	1.2114
		AMH	0.906	0.60313
herbs	356	AMH	0.864	0.49490
		C	0.860	0.28428
		C	0.848	1.1459

Table 11 Fitting results for various product copula models (copula of variables RSR, SLA and minus HW): estimated parameters and the approximation coefficient $\hat{\rho}$ (defined in (7)); in the last line $\hat{\gamma}_2$ is not present; copulas C_1 and C_2 are chosen for model (4) and abbreviated according to Table 8

category	n	C_1	C_2	$\hat{\rho}$	estimated parameters $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$
grass	546	J	F	0.974	1.6627, 2.5963, 0.03237, 0.37216, 0.92333
		F	GH	0.972	3.4219, 1.3201, 0.86894, 0.44546, 0.00439
		GH	GH	0.970	1.3919, 1.4199, 0.06256, 0.40173, 1
Herbs	356	C	C	0.907	0.11768, 7.1482, 0.88526, 0.66026, 0.54809
		C	F	0.905	0.13309, 4.1267, 0.82187, 0.46801, 0.34606
		C	Π	0.904	3.3265, 0.21626, 0.44266, 0.63990

Table 12 subset of grass species: indicator variables for asymmetries, empirical Spearman correlations $\hat{\rho}_S$ (including 95% confidence intervals) and the Spearman correlations using the best model fit; lcb=lower confidence bound, ucb=upper confidence bound

Pair of variables	best fit model					empirical correlation		
	λ_L	λ_U	γ_1	γ_2	ρ_S	$\hat{\rho}_S$	lcb	ucb
RSR, SLA	0	0.02852	0.03011	0.03277	0.31931	0.32578	0.24867	0.39879
RSR, -HW	0	0.03026	0.05439	0.05291	0.08222	0.05581	−0.02823	0.13907
SLA, -HW	0	0.25449	0.04155	0.03198	0.27146	0.22419	0.14297	0.30241

Appendix C: Names of species in the database

Grass

Al.pr...Alopecurus pratensis
An.od...Anthoxanthum odoratum
Ar.el...Arrhenatherum elatius
Av.pu...Helictotrichon pubescens
Cy.cr...Cynosurus cristatus
Da.gl...Dactylis glomerata
Fe.pr...Festuca pratensis
Lo.pe...Lolium perenne
Po.pr...Poa pratensis
Po.tr...Poa trivialis

Herbs

Ac.mi...Achillea millefolium
Be.pe...Bellis perennis
Ce.ja...Centaurea jacea
Ga.mo...Galium mollugo
Ga.ve...Galium verum
Pl.la...Plantago lanceolata
Ra.ac...Ranunculus acris
Ra.bu...Ranunculus bulbosus
Ru.ac...Rumex acetosa
Ve.ch...Veronica chamaedrys

Appendix D

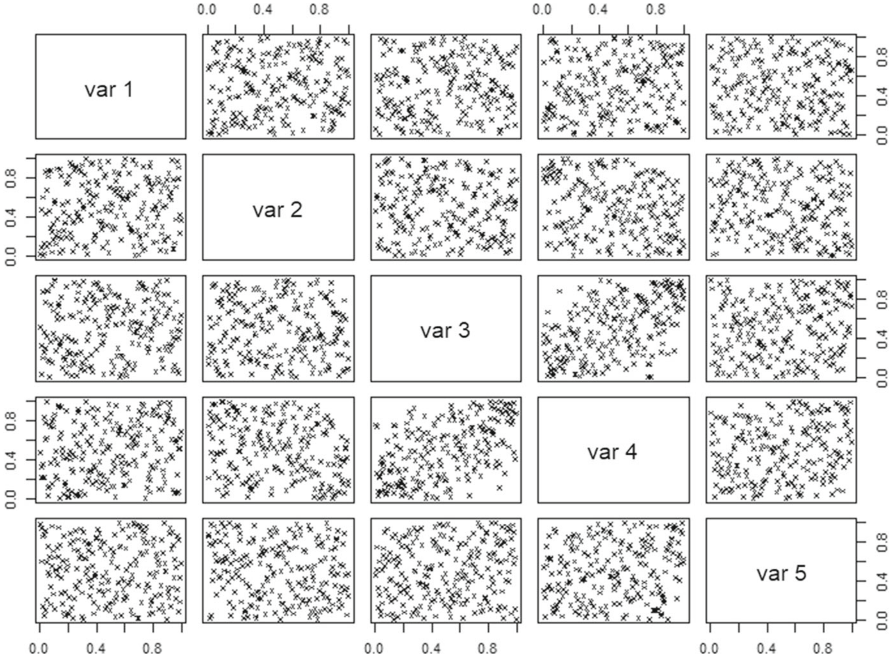

Fig. 6 pairwise copula plots of herb species, the variables are RSR, SLA, RCNC, LCNC, HW

References

Anderson MJ, de Valpine P, Punnett A, Miller AE (2019) A pathway for multivariate analysis of ecological communities using copulas. Ecol Evol 9(6):3276–3294. https://doi.org/10.1002/ece3.4948

Bruelheide H, Dengler J, Purschke O, Lenoir J, Jiménez-Alfaro B, Hennekens SM, Botta-Dukát Z, Chytrý M, Field R, Jansen F, Kattge J, Pillar VD, Schrodt F, Alvarez-Dávila E, Khan MASA, Attorre F, Aubin I, De Sanctis M, Díaz S et al (2018) Global trait-environment relationships of plant communities. Nat Ecol Evol 2:1906–1917. https://doi.org/10.1038/s41559-018-0699-8

Chang B, Joe H (2019) Prediction based on conditional distributions of vine copulas. Comput Stat Data Anal 139:45–63. https://doi.org/10.1016/j.csda.2019.04.015

Dayal, KS, Deo RC, Apan, AA (2020) Development of copula-statistical drought prediction model using the Standardized Precipitation-Evapotranspiration Index. In: Handbook of Probabilistic Models, pp 141-178. https://doi.org/10.1016/B978-0-12-816514-0.00006-0

de Freitas Costa E, Schneider S, Carlotto GB et al (2021) Zero-inflated-censored Weibull and gamma regression models to estimate wild boar population dispersal distance. Jpn J Stat Data Sci. https://doi.org/10.1007/s42081-021-00124-0

Butler EE, Datta A, Flores-Moreno H, Chen M, Wythers KR, Fazayeli F, Banerjee A, Atkin OK, Kattge J, Amiaud B, Blonder B, Boenisch G, Bond-Lamberty B, Brown KA, Byun C, Campetella G, Cerabolini BEL, Cornelissen JHC, Craine JM et al (2017) Mapping local and global variability in plant trait distributions. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1708984114

Díaz S, Hodgson JG, Thompson K, Cabido M, Cornelissen JHC, Jalili A, Montserrat-Martí G, Grime JP, Zarrinkamar F, Asri Y, Band SR, Basconcelo S, Castro-Díez P, Funes G, Hamzehee B, Khoshnevi M, Pérez-Harguindeguy N, Pérez-Rontomé MC, Shirvany FA et al (2004) The plant traits that drive ecosystems: evidence from three continents. J Veg Sci 15(3):295–304. https://doi.org/10.1111/j.1654-1103.2004.tb02266.x

Embrechts P (2009) Copulas: a personal view. J Risk Insur 76:639–650. https://doi.org/10.1111/j.1539-6975.2009.01310.x

Emura T, Michimae H (2017) A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae. Environ Ecol Stat 24:151–173. https://doi.org/10.1007/s10651-017-0364-4

Fan Y, Patton AJ (2014) Copulas in econometrics. Annu Rev Econ 6:179–200. https://doi.org/10.1146/annurev-economics-080213-041221

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze ED, Weisser WW (2010) Implementing large-scale and long-term functional biodiversity research: the Biodiversity Exploratories. Basic Appl Ecol 11(6):473–485. https://doi.org/10.1016/j.baae.2010.07.009

Ghosh S, Sheppard LW, Holder MT, Loecke TD, Reid PhC (2020) Copulas and their potential for ecology. Adv Ecol Res. https://doi.org/10.1016/bs.aecr.2020.01.003

Ghosh S, Cottingham KL, Reuman DC (2021) Species relationships in the extremes and their influence on community stability. Philos Trans R Soc Lond Ser B 376(1835):20200343. https://doi.org/10.1098/rstb.2020.0343

Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. Adv Water Resour 29:1155–1167. https://doi.org/10.1016/j.advwatres.2005.09.005

Grime JP (1979) Plant strategies and vegetation processes. Wiley, New York

Hagey TJ, Puthoff JB, Crandell KE, Autumn K, Harmon LJ (2016) Modeling observed animal performance using the Weibull distribution. J Exp Biol 219(11):1603–1607. https://doi.org/10.1242/jeb.129940

Heisse K, Roscher C, Schumacher J, Schulze ED (2007) Establishment of grassland species in monocultures: different strategies lead to success. Oecologia 152(3):435–447. https://doi.org/10.1007/s00442-007-0666-6

Herz K, Dietz S, Haider S, Jandt U, Scheel D, Bruelheide H (2017) Drivers of intraspecific trait variation of grass and forb species in German meadows and pastures. J Veg Sci 28(4):705–716. https://doi.org/10.1111/jvs.12534

Hofert M, Mächler M, Mc Neil AJ (2012) Likelihood inference for Archimedean copulas in high dimensions under known margins. J Multivar Anal 110:133–150. https://doi.org/10.1016/j.jmva.2012.02.019

Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ, Cornelissen JHC, Violle C, Harrison SP, Van Bodegom PM, Reichstein M, Enquist BJ, Soudzilovskaia NA, Ackerly DD, Anand M et al (2011) TRY—a global database of plant traits. Glob Chang Biol 17(9):2905–2935. https://doi.org/10.1111/j.1365-2486.2011.02451.x

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, Acosta ATR, Adamidis GC, Adamson K, Aiba M, Albert CH, Alcántara JM, Alcázar CC, Aleixo I, Ali H et al (2020) TRY plant trait database—enhanced coverage and open access. Glob Chang Biol 26(1):119–188. https://doi.org/10.1111/gcb.14904

Kleyer M, Bekker RM, Knevel IC, Bakker JP, Thompson K, Sonnenschein M, Poschlod P, Van Groenendael JM, Klimeš L, Klimešová J, Klotz S, Rusch GM, Hermy M, Adriaens D, Boedeltje G, Bossuyt B, Dannemann A, Endels P, Götzenberger L et al (2008) The LEDA Traitbase: a database of life-history traits of the Northwest European flora. J Ecol 96(6):1266–1274. https://doi.org/10.1111/j.1365-2745.2008.01430.x

Kühn I, Durka W, Klotz S (2004) BiolFlor—a new plant-trait database as a tool for plant invasion ecology. Diversity Distrib 10(5–6):363–365. https://doi.org/10.1111/j.1366-9516.2004.00106.x

Lever J, Krzywinski M, Altman N (2017) Principal component analysis. Nat Methods 14(7):641–642. https://doi.org/10.1038/nmeth.4346

Liebscher E (2008) Construction of asymmetric multivariate copulas. J Multivar Anal 99:2234–2250. https://doi.org/10.1016/j.jmva.2008.02.025

Liebscher E (2009) Semiparametric estimation of the parameters of multivariate copulas. Kybernetika 6:972–991

Liebscher E (2015) Goodness-of-approximation of copulas by a parametric family. In: Stochastic models, statistics and their applications. Springer Proceedings in Mathematics Statistics 122: 101–109. https://doi.org/10.1007/978-3-319-13881-7_12

McNeil A (2008) Sampling nested Archimedean copulas. J Stat Comput Simul 78(6):567–581. https://doi.org/10.1080/00949650701255834

Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York

Michimae H, Matsunami M, Emura T (2020) Robust ridge regression for estimating the effects of correlated gene expressions on phenotypic traits. Environ Ecol Stat 27:41–72. https://doi.org/10.1007/s10651-019-00434-3

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer Series in Statistics 139. Springer, New York

Paul S, Lansing E, Service USF, Biology, O (1999) Generality of leaf trait relationships: A test across six biomes. Ecology 80(6):1955–1969. https://doi.org/10.1890/0012-9658(1999)080[1955:GOLTRA]2.0.CO;2

Reich PB, Walters MB, Ellsworth DS (1992) Leaf life-span in relation to leaf, plant, and stand characteristics among diverse ecosystems. Ecol Monogr 62(3):365–392. https://doi.org/10.2307/2937116

Reich PB, Walters MB, Ellsworth DS (1997) From tropics to tundra: global convergence in plant functioning. Proc Natl Acad Sci USA 94(25):13730–13734. https://doi.org/10.1073/pnas.94.25.13730

She D, Xia J (2018) Copulas-based drought characteristics analysis and risk assessment across the Loess Plateau of China. Water Resour Manag 32:547–564. https://doi.org/10.1007/s11269-017-1826-z

Taubert F, Hartig F, Dobner H-J, Huth A (2013) On the challenge of fitting tree size distributions in ecology. PLoS ONE 8(2):e58036. https://doi.org/10.1371/journal.pone.0058036

Tsukahara H (2005) Semiparametric estimation in copula models. Can J Stat 33:357–375. https://doi.org/10.1002/cjs.5540330304

Stephens MA (1986) Tests based on EDF statistics. In: D'Agostino RB, Stephens MA (eds) Goodness of fit techniques. Marcel Dekker, New York

Violle C, Navas M-L, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E (2007) Let the concept of trait be functional! Oikos 116(5):882–892. https://doi.org/10.1111/j.2007.0030-1299.15559.x

Try Plant Traint Database https://www.try-db.org/

Vio R, Nagler TW, Andreani P (2020) Modeling high-dimensional dependence in astronomical data. Astronomy and Astrophysics 642(A156):1–10. https://doi.org/10.1051/0004-6361/202038585

Weiß G (2011) Copula parameter estimation by maximum-likelihood and minimum-distance estimators: a simulation study. Comput Stat 26:31–54. https://doi.org/10.1007/s00180-010-0203-7

Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JHC, Diemer M, Flexas J, Garnier E, Groom PK, Gulias J, Hikosaka K, Lamont BB, Lee T, Lee W, Lusk C et al (2004) The worldwide leaf economics spectrum. Nature 428:821–827. https://doi.org/10.1038/nature02403

Zhang X, Wilson A (2016) System reliability and component importance under dependence: a copula approach. Technometrics 59:215–224. https://doi.org/10.1080/00401706.2016.1142907

Eckhard Liebscher is currently working as Professor of Stochastics and Data Analysis at the University of Applied Sciences Merseburg. His research interests are focused on Multivariate Statistics (copulas, dependence measures), Reliability and E-Learning. He teaches basic courses on Mathematics, and courses on Data Analysis, Reliability and Machine Learning (Bachelor and Master level).

Franziska Taubert is the leader of the working group Grassland Modelling at the Helmholtz Centre for Environmental Research GmbH - UFZ. Her main area of expertise covers vegetation ecology, mathematical ecology and ecological modelling with specific focus on grassland ecology, modelling and biodiversity.

David Waltschew is currently working as software engineer at the University of Applied Sciences Merseburg

Jessica Hetzer is a PhD student at the Helmholtz Centre for Environmental Research GmbH - UFZ. Her main area of expertise is Applied Mathematics. Her current research focuses on vegetation models in light of scale, plant traits and climate.