Computational Statistics and Data Analysis 169 (2022) 107417

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Approximate Bayesian conditional copulas

Clara Grazian [a,b,*], Luciana Dalla Valle [c], Brunero Liseo [d]

[a] University of Sydney, School of Mathematics and Statistics, Carslaw Building F07, Camperdown, 2006, NSW, Australia
[b] School of Mathematics and Statistics, University of Sydney, and ACEMS, Australia
[c] School of Engineering, Computing and Mathematics, University of Plymouth, United Kingdom
[d] MEMOTEF Department, Sapienza Università di Roma, Italy

ARTICLE INFO

Article history:
Received 4 July 2021
Received in revised form 11 November 2021
Accepted 30 December 2021
Available online 14 January 2022

Keywords:
Approximate Bayesian computation
Bayesian inference
Dependence modelling
Gaussian processes
Empirical likelihood
Splines

ABSTRACT

Copula models are flexible tools to represent complex structures of dependence for multivariate random variables. According to Sklar's theorem, any multidimensional absolutely continuous distribution function can be uniquely represented as a copula, i.e. a joint cumulative distribution function on the unit hypercube with uniform marginals, which captures the dependence structure among the vector components. In real data applications, the interest of the analyses often lies on specific functionals of the dependence, which quantify aspects of it in a few numerical values. A broad literature exists on such functionals, however extensions to include covariates are still limited. This is mainly due to the lack of unbiased estimators of the conditional copula, especially when one does not have enough information to select the copula model. Several Bayesian methods to approximate the posterior distribution of functionals of the dependence varying according covariates are presented and compared; the main advantage of the investigated methods is that they use nonparametric models, avoiding the selection of the copula, which is usually a delicate aspect of copula modelling. These methods are compared in simulation studies and in two realistic applications, from civil engineering and astrophysics.

1. Introduction

Copula models have received an increasing interest since the work of Sklar (1959). Sklar's theorem is a probability result which states that every multivariate cumulative distribution function (CDF, hereafter) can be represented as a copula, i.e. a joint cumulative distribution function on the unit hypercube $[0, 1]^d$ with uniform marginals, capturing the dependence structure among the components of the random vector. This result is very important in statistical modelling, especially when it is reasonable and useful to separately model the marginal distributions and the potentially complex multivariate dependence structure, or when the degree of information about the marginals and their dependencies is different, since, in general, more information can be gathered on marginal aspects of the problem at hand.

Sklar's theorem proves that every multivariate distribution function $F_{\mathbf{Y}}(\cdot)$ of a random variable $\mathbf{Y} = (Y_1, \ldots, Y_d)$ can be represented by a copula $C(\cdot) : [0, 1]^d \to [0, 1]$ depending on d univariate marginal distributions

$$F_{\mathbf{Y}}(y_1, \ldots, y_d) = C\left(F_1(y_1|\boldsymbol{\theta}_1), \ldots, F_d(y_d|\boldsymbol{\theta}_d) \mid \boldsymbol{\psi}\right), \tag{1.1}$$

* Corresponding author.
E-mail address: c.grazian@unsw.edu.au (C. Grazian).

https://doi.org/10.1016/j.csda.2021.107417

where $F_j(\cdot)$ is the marginal CDF of Y_j, depending on the parameter vector $\boldsymbol{\theta}_j$, for $j = 1, \ldots, d$, and $\boldsymbol{\psi}$ is the copula parameter vector. This representation is unique if $F_{\mathbf{Y}}$ is continuous. In the case of continuous random variables, Equation (1.1) admits the following density, which shows how the copula absorbs all the dependence of the model

$$f_{\mathbf{Y}}(y_1, \ldots, y_d) = c(F_1(y_1|\boldsymbol{\theta}_1), \ldots, F_d(y_d|\boldsymbol{\theta}_d) \mid \boldsymbol{\psi}) \cdot f_1(y_1|\boldsymbol{\theta}_1) \cdot \ldots \cdot f_d(y_d|\boldsymbol{\theta}_d),$$

where $c(\cdot)$ is the density of the copula $C(\cdot)$. In this work, we will focus on the bivariate case $f_{\mathbf{Y}}(y_1, y_2)$, however all the methods can be easily extended to the d-dimensional case. We refer the reader to Nelsen (2007) for a detailed description of copula theory and methods. Therefore, the density of the d-dimensional absolutely continuous distribution can be represented as the product of the marginal densities times the copula density applied to the marginal distribution functions.

Patton (2006) extends the definition of copula in the presence of covariates, to describe situations where the marginals and their dependence structure are influenced by the values of other variables, that is:

$$F_{\mathbf{Y}|X}(y_1, \ldots, y_d|\mathbf{X} = \mathbf{x}) = C_{\mathbf{X}}\left(F_{1|\mathbf{X}}(y_1|X = \mathbf{X}), \ldots, F_{d|\mathbf{X}}(y_d|\mathbf{X} = \mathbf{x}) \mid \mathbf{X} = \mathbf{x}\right),$$

where $\mathbf{X} \in \mathbb{R}^p$ represents a vector of covariates, $C_X(\cdot)$ is the conditional copula, which may potentially vary with $\mathbf{X}$, and $F_{j|\mathbf{X}}$ is the conditional CDF of Y_j, for $j = 1, \ldots, d$. Here and in the following, the dependence on the parameters $(\boldsymbol{\psi}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d)$ is left implicit in the notation of the CDFs. The introduction of covariates may be useful in many applications where the dependence structure varies over the space of the observations (Acar et al., 2011). Moreover, conditional copulas are the building blocks of vine copulas (Czado, 2010), where situations of dependence among the variables on which copulas are conditioned on are common in real applications; in these cases, a "Simplifying Assumption" (Czado, 2019) is often introduced in order to make statistical analysis easier. According to the Simplifying Assumption, the conditional copula is assumed constant, as in Gijbels et al. (2015). Several contributions to the literature aim at exploring and testing violations of this assumption, such as in Haff et al. (2010), Acar et al. (2012), Acar et al. (2013), Killiches et al. (2016), Killiches et al. (2017) and Kurz and Spanhel (2017). However, Levi and Craiu (2018) shows that violations of the Simplifying Assumption may be due to the omission of important covariates, rather than to a real dependence on the included covariates. This result suggests that, in practical situations, it is safer to assume the potential dependence of the copula on the values of the available covariates.

A standard approach to model the influence of covariates on copulas is based on a parametric model which assumes a functional relationship between copula parameters and covariates, such that $C_{\mathbf{X}}(\cdot) = C_{\psi(\mathbf{X})}(\cdot)$, where $\psi \in \Psi$ is the copula parameter, assumed to be a function of the covariates $\mathbf{X}$, supposing for simplicity that the copula depends on one parameter. In this setting, the parameter is associated to the covariates through a link function $\zeta : \Psi \to \mathbb{R}$, such that $\psi(\mathbf{X}) = \zeta^{-1}(\eta(\mathbf{X}))$, where $\eta(\cdot)$ is a real-valued calibration function. The calibration function may assume different forms. A parametric form is adopted, for example, by Genest et al. (1995), while a nonparametric form is suggested by Acar et al. (2011), which employs a local polynomial-based approach, and Craiu and Sabeti (2012), Vatter and Chavez-Demoulin (2015), Klein and Kneib (2016) and Stander et al. (2019), which propose additive conditional copula regression specifications with predictors defined using splines.

Different approaches are considered in the literature for the estimation of conditional copulas. Abegaz et al. (2012) and Gijbels et al. (2012) have proposed semiparametric and nonparametric methodologies within the frequentist framework to model the influence on copulas of covariates taking values in complex spaces; in both papers the authors consider the statistical properties of conditional copula estimators, establishing consistency and asymptotic normality results.

In the Bayesian framework, Dalla Valle et al. (2018) has proposed to nonparametrically estimate the conditional copula density in the bivariate case, introducing a generalization of ideas presented in Wu et al. (2015). The authors assume that the unknown conditional copula density can be represented as an infinite mixture of Gaussian copulas, where the correlation parameter is defined as a (linear or non-linear) function of a covariate:

$$c_X(u_1, u_2 \mid X = x) = \sum_{\iota=1}^{\infty} \pi_\iota c_{\rho(x)}(u_1, u_2 \mid X = x),$$

where $c_{\rho(x)}(\cdot)$ denotes the Gaussian copula densities with correlation coefficient $\rho(x)$ depending on the covariate X, $u_1 = F_{1|X}(y_1 \mid X = x)$, $u_2 = F_{2|X}(y_2 \mid X = x)$, $\sum_{\iota=1}^{\infty} \pi_\iota = 1$ and $0 < \pi_\iota < 1$.

Levi and Craiu (2018) proposes to jointly estimate the marginal distributions and the copula using Gaussian process (GP) models, where the calibration function follows *a priori* a single-index model based on GPs, to handle high-dimensional covariates. In more details, the authors assume that the copula is characterized by its own calibration function $\boldsymbol{\eta}(\mathbf{x}_i) = \boldsymbol{\eta}(\mathbf{x}_i'\boldsymbol{\beta})$, for $i = 1, \ldots, n$, where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of coefficients that must be normalized for identifiability reasons ($||\boldsymbol{\beta}|| = 1$). The calibration function follows a Gaussian process prior, centred around zero and with covariance structure depending on a kernel which is a function of the distance among the covariates, e.g. the squared exponential kernel. Although the GP approach is very attractive for its flexibility, the idea of modelling the parameters of a known copula as a function of covariates implies the need to choose the copula family. Several model selection methods, which may be applied to both the choice of the copula family and the choice of the form of the calibration function, are available. One approach compares the average prediction power of different models using the cross-validated pseudo-marginal likelihood (CVML) proposed by Geisser and Eddy (1979); another approach is based on the Watanabe-Akaike information criterion, proposed by Watanabe (2013). Both measures can be generalized to consider covariates.

As a general comment, we point out that the choice of a statistical model for the distribution of a multivariate random vector is generally complex. In particular, when differences among copula models are strong in the tails of the distributions, it is possible that there is a limited number of observations in applied settings in order to clearly identify the correct model. We will show this limitation in the simulation study of Section 3. For this reason, it is often the case that the researcher prefers to reconsider the inference goals on some low dimensional functional of the copula, i.e., for example Kendall's τ, Spearman's ρ or some tail dependence indices; in such case, the complete dependence structure could be considered as a nuisance parameter.

In this paper we explore several ways to make inference on functionals of the dependence, in particular the Spearman's ρ and the Kendall's τ, in presence of covariates. The goal of the paper is to avoid the selection of the copula. We discuss and propose three methods of increasing relaxation of the distributional assumptions on the functional of interest: the former is based on GPs, where, as opposed to Levi and Craiu (2018), the choice of a specific copula family is avoided; the second is a direct generalization of the approach of Grazian and Liseo (2017) and makes use of the empirical likelihood of the functional of interest, either implementing an inconsistent estimator of the conditional copula or a linearized model; the latter makes use of Bayesian splines to approximate the behaviour of the functional of the copula. For each method, we discuss advantages and disadvantages. GPs show the best performance in the simulation study, they are applicable in presence of repetitions of observations for each level of the covariate, therefore the method is affected by the curse of dimensionality and may be not applicable in real applications in the presence of an even small number of covariates. The method based on the Bayesian use of the empirical likelihood shows a worse performance and it relies on the existence of moment conditions, however it can be applied to data without the need of replications for each level of the covariate. Finally, the Bayesian splines show a good performance. However they need a certain level of tuning and, again, replications for each level of the covariate; moreover, they show a tendency to over-smooth the function of interest, as it will be shown on real datasets.

The `BICC` R-package R Core Team (2013) has been prepared to implement all the methodologies presented in this work. The package `BICC` is publicly available at the page https://github.com/cgrazian/BICC.

The remainder of this paper is organised as follows: Section 2 presents several methods to perform inference on functionals of the dependence based on nonparametric representations of the copula. Each of these methods is compared in Section 3 and is contrasted to the conditional method proposed by Levi and Craiu (2018) in the case of two covariates. Two datasets showing non-linear dependence on one or two covariates are analysed in Section 4, showing that areas of applications include a broad range spacing from civil engineering and energy management to astrophysics. Finally, Section 5 concludes the paper.

2. Bayesian analysis for functionals of conditional copulas

2.1. Conditional dependence measures

Let us consider the bivariate case and assume that for each level of a covariate $(x_1, x_2, \ldots, x_k)$ we observe n_ℓ replications Y_{1i} and Y_{2i}, with $\ell = 1, \ldots, k$, and compute the probability integral transforms $u_{ji} = F_{j|X_i}(y_j|X_i = x_i)$ for $j = 1, 2$ so to obtain:

$$\left[\left(u_{1,1}, u_{2,1}\right), \ldots, \left(u_{1,n_\ell}, u_{2,n_\ell}\right)\right], \qquad \ell = 1, \ldots, k,$$

such that n_ℓ is the sample size of (u_1, u_2) at location x_ℓ and $\sum_{\ell=1}^{k} n_\ell = n$. The joint distribution is defined through a conditional copula:

$$F_{Y_1,Y_2|X}(y_1, y_2|X = x) = C_X(u_1, u_2|X = x).$$

We also assume that the strength of dependence between vectors U_1 and U_2 can be modelled as a smooth function of X and that we are not able to assume any specific parametric form for C_X. Also, we consider the situation where one is mainly interested in making inference on a synthetic dependence measure of C_X, say

$$\varphi(C_X; x) = \mathbb{E}[v(U_1, U_2)\big|X = x],$$

for some function v. The quantities of interest are in general, functionals of the dependence, such as Kendall's τ, Spearman's ρ or tail dependence indices. Kendall's τ is a measure of similarity of the orderings of the data. Given two independent bivariate random variables $(Y_{1,1}, Y_{2,1})$ and $(Y_{1,2}, Y_{2,2})$, Kendall's τ is defined as

$$\tau = \mathbb{E}\left(\text{sgn}\left[(Y_{1,1} - Y_{1,2})(Y_{2,1} - Y_{2,2})\right]\right).$$

This dependence measure can also be defined in terms of copulas as

$$\tau(Y_1, Y_2) = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1. \tag{2.1}$$

The Spearman's ρ is an alternative nonparametric measure of rank correlation, assessing how well the dependence among variables can be described by a monotonic function; it is defined as

$$\rho(Y_1, Y_2) = \text{Corr}\left(F_{Y_1}(Y_1), F_{Y_2}(Y_2)\right);$$

the corresponding copula expression is

$$\rho(Y_1, Y_2) = 12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3. \tag{2.2}$$

Other measures of interest are tail dependence indices or the conditional expected value of U_1 for a given level of X and U_2, see Levi and Craiu (2018). In addition, dependence measures such as the Randomized Dependence Coefficient (RDC) was proposed by Lopez-Paz et al. (2013) for non-monotonic associations. The rest of the paper is focused on bivariate Spearman's ρ and Kendall's τ, however other measures of dependence can be analysed, as well as multivariate extensions to dimension d; it is important to notice, however, that multivariate functionals of the dependence are often not uniquely defined Schmid and Schmidt (2007). Moreover, the methodologies proposed in this work require the definition of estimators of the quantities of interest. While the literature provides estimators for the measures of monotonic dependence, such as the Spearman's ρ and the Kendall's τ, robust nonparametric estimation of other measures has not been considered thoroughly. Recently, Goegebeur et al. (2020) proposes a robust and nonparametric estimation of the coefficient of tail dependence in presence of covariates, based on density power divergence. However, while Goegebeur et al. (2020) proves the existence, convergence and asymptotic normality of the estimator, it does not provide an easy-to-use formula, as in the case provided in this work. For these reasons, Bayesian nonparametric analysis of conditional tail dependence is left for further research.

Since the above dependence indices can be directly defined through their copula, the conditional versions of (2.1) and (2.2) can be easily derived in terms of conditional copulas:

$$\tau(x) = 4 \iint_{[0,1]^2} C_X(u_1, u_2) dC_X(u_1, u_2) - 1, \qquad \text{and}$$
$$\rho(x) = 12 \iint_{[0,1]^2} C_X(u_1, u_2) du_1 du_2 - 3. \tag{2.3}$$

Consequently, the most common estimators of measures of conditional dependence are expressed in terms of estimators of the conditional copula. In particular, estimators of $\tau(x)$ and $\rho(x)$ are obtained in terms of

$$C_{X;h}(u_1, u_2) = \sum_{i=1}^{n} \pi_{n;i}(x, h_n) \mathbb{I}\left[U_1 \le u_1, U_2 \le u_2\right], \tag{2.4}$$

where $\mathbb{I}$ is an indicator function and $\{\pi_{n;i}(x, h_n)\}$ is a sequence of weights that smooth over the covariate space (for example, the Nadaraya-Watson or the local-linear weights) and $h_n > 0$ is a bandwidth which is assumed to vanish as the sample size increases. See Gijbels et al. (2011) for a detailed definition. Moreover, weights depend on a kernel smoothing over the covariate space, such as the triweight kernel and the Gaussian kernel.

Based on these estimators of the copula, Gijbels et al. (2011) propose the nonparametric version of the conditional Kendall's τ:

$$\hat{\tau}(x) = -1 + \frac{4}{1 - \sum_{i=1}^{n} \pi_{n;i}^2(x, h_n)} \sum_{i=1}^{n} \sum_{t=1}^{n} \pi_{n;i}(x, h_n) \pi_{n;t}(x, h_n) \mathbb{I}(Y_{i,1} < Y_{t,1}, Y_{i,2} < Y_{t,2}),$$

and of the conditional Spearman's ρ

$$\hat{\rho}(x) = 12 \sum_{i=1}^{n} \pi_{n;i}(x, h_n)(1 - \hat{U}_{i,1})(1 - \hat{U}_{i,2}) - 3,$$

where $\hat{U}_{i,1} = \hat{F}_{1|X}(y_{i,1}|X_i = x_i)$ and $\hat{U}_{i,2} = \hat{F}_{2|X}(y_{i,2}|X_i = x_i)$.

Unfortunately, estimator (2.4) is biased and the size of the bias depends on the role of the covariates on the marginal distributions. In order to reduce the influence of the covariates, Gijbels et al. (2011) have proposed an alternative estimator, where the marginal distributions are similarly approximated using Equation (2.4); in this case, a different sequence of weights must be adopted. Nonetheless, this latter estimator is biased as well: whereas, in particular settings, it can be drastically less biased than the former estimator, there is no guarantee that the bias will always be smaller than the bias of the former. In addition, it is necessary to choose three different bandwidth parameters in order to implement it. We refer to Veraverbeke et al. (2011) for a discussion of the asymptotic properties of these estimators. In conclusion, the problem of evaluating the bias of the nonparametric functionals of the dependence is still open.

Grazian and Liseo (2016) shows a preliminary extension of the method proposed in Grazian and Liseo (2017) to the conditional case, based on simulations. However, the lack of theoretical warranties makes the performance of the procedure rather "problem-specific" and difficult to evaluate in general.

In this work, the goal of the analysis is to approximate the posterior distribution of $\varphi(u_1, u_2; x) = \varphi(x)$ where $\varphi(\cdot)$ is a measure of conditional dependence, when a nonparametric consistent estimator of $\varphi(x)$ is not available. We propose to work on a transformation of $\varphi(x)$, encouraging normality, e.g. the Fisher's transformation which maps indices defined in $[-1, 1]$, like Spearman's ρ or Kendall's τ, into the real line:

$$Z(x) = \frac{1}{2}\log\frac{1+\varphi(x)}{1-\varphi(x)}. \tag{2.5}$$

The sample version of $Z(x)$ is then defined by computing the (possibly biased) estimates of φ:

$$W(x) = \frac{1}{2}\log\frac{1+\hat{\varphi}(x)}{1-\hat{\varphi}(x)}.$$

2.2. *Gaussian processes approach*

In this section, we present a method based on a Gaussian representation of a transformation of the functionals of the dependence. While the methodology is already available in the literature, to the best of our knowledge, this is the first application to a copula setting. First, we assume that $Z(x)$ follows - *a priori* - a GP

$$Z(x) \sim \mathcal{GP}\left(\mathbf{g}(x)^T\boldsymbol{\beta}, \sigma^2\mathcal{K}(x, x'; \xi)\right). \tag{2.6}$$

Here, the location parameter of the Gaussian process is

$$\mathbb{E}[Z(x)] = \mathbf{g}(x)^T\boldsymbol{\beta},$$

where $\mathbf{g}(x) = \left(g_1(x), \ldots, g_q(x)\right)^T$ is a vector of known functions, $x \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^q$. Common choices for the basis function $\mathbf{g}(x)$ are $\mathbf{0}$, $(1, \mathbf{x})$ or $(1, \mathbf{x}, \mathbf{x}^2)$, and so on.
Also, $\mathcal{K}(x, x'; \xi)$ is a generic correlation kernel depending on a parameter ξ, and σ^2 is a positive scale parameter, so that

$$\mathbb{C}ov(Z(x), Z(x')) = \sigma^2\mathcal{K}(x, x'; \xi).$$

Without loss of generality, we will consider the squared exponential kernel

$$\mathbb{C}ov\left(Z(x), Z(x')\right) = \sigma^2\mathcal{K}(x, x'; \xi) = \sigma^2\exp\left(-\frac{1}{2}\sum_{\iota=1}^{m}\frac{d(x_\iota, x'_\iota)}{\xi_\iota}\right) = \sigma^2\exp\left(-\frac{1}{2}\sum_{\iota=1}^{m}\frac{(x_\iota - x'_\iota)^2}{\xi_\iota}\right).$$

For ease of notation, we will consider the case where $m = 1$, that is

$$\mathcal{K}(x, x'; \xi) = \exp\left(-\frac{1}{2}\frac{(x-x')^2}{\xi}\right);$$

however, generalizations to $m > 1$ are straightforward.

We estimate the functional $\varphi(x_\ell)$ using the unconditional consistent estimator of φ, say $\hat{\varphi}$ and assume that each vector of observations, $(y_{i,1}, y_{i,2}, x_\ell)$, $i = 1, \ldots, n_\ell$, generates a noisy version of $Z(x_\ell)$, say $W(x_\ell)$, which depends on a statistic evaluated at location x_ℓ; in practice, $W(x)$ is a noisy observation of the signal $Z(x)$. This means that for each level of the covariate, n_ℓ replications are needed. It is possible to explicitly model the noise through some parametric assumption. For instance, the case of compensating errors can be modelled through a Gaussian distribution

$$W(x_\ell) = Z(x_\ell) + \varepsilon_\ell \qquad \ell = 1, \ldots, k,$$

where $Z(x_\ell)$ is defined as in Equation (2.6) and $\varepsilon_\ell \sim \mathcal{N}(0, \kappa_\ell^2)$, with $\kappa_\ell^2 = \kappa^2/n_\ell$ and $\varepsilon_\ell \perp \varepsilon_{\ell'}, \ell \neq \ell'$.

Consequently, the observations follow a normal distribution

$$W(x_\ell) \sim \mathcal{N}\left(\mathbf{g}(x_\ell)^T\boldsymbol{\beta}, \sigma^2 + \kappa_\ell^2\right) \qquad \ell = 1, \ldots, \kappa. \tag{2.7}$$

From Equation (2.7), it follows that the likelihood associated to the observations related to locations $x_1, \ldots, x_k$, with a number of observations $n_1, \ldots, n_k$ each, is

$$L(\boldsymbol{\beta}, \sigma^2, \xi, \kappa^2) = \mathcal{N}\left(\mathbf{g}(\mathbf{x})^{\mathbf{T}}\boldsymbol{\beta}, \boldsymbol{\sigma}^2\boldsymbol{\Sigma}_\xi + \kappa^2\tilde{\mathbf{I}}\right),$$

where $\tilde{\mathbf{I}} = \text{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_k}\right)$ is a diagonal matrix with element $\left(\frac{1}{n_1}, \ldots, \frac{1}{n_k}\right)$ on the diagonal, and

$$\boldsymbol{\Sigma}_\xi = \begin{pmatrix} \mathcal{K}(x_1, x_1; \xi) & \cdots & \mathcal{K}(x_1, x_k; \xi) \\ \vdots & \ddots & \vdots \\ \mathcal{K}(x_k, x_1; \xi) & \cdots & \mathcal{K}(x_k, x_k; \xi) \end{pmatrix}.$$

Standard Bayesian approaches to Gaussian processes can then be developed to estimate the parameters of this model $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\kappa}, \xi)$ (Goldstein and Wooff, 2007, Cressie, 1992). First, the variance matrix can be reformulated following Paulo (2005):

$$\sigma^2 \boldsymbol{\Sigma}_\xi + \kappa^2 \tilde{\mathbf{I}} = \sigma^2 \left(\boldsymbol{\Sigma}_\xi + \frac{\kappa^2}{\sigma^2} \tilde{\mathbf{I}} \right) = \sigma^2 \left(\boldsymbol{\Sigma}_\xi + \lambda \tilde{\mathbf{I}} \right) = \sigma^2 \mathbf{M},$$

where $\lambda = \kappa^2/\sigma^2$ and $\mathbf{M}$ is a positive definite matrix depending on parameters ξ and λ.

Then, the likelihood can be rewritten as

$$L(\boldsymbol{\beta}, \sigma^2, \lambda, \xi) = \frac{1}{\sigma^k \mid \mathbf{M} \mid^{1/2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{W} - \tilde{\mathbf{X}}\boldsymbol{\beta})^T \mathbf{M}^{-1} (\mathbf{W} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \right), \tag{2.8}$$

where $\tilde{\mathbf{X}} = \mathbf{g}(\mathbf{x})^{\mathbf{T}}$ is a $k \times q$ matrix of known constants.

Integrating Equation (2.8) with respect to $\boldsymbol{\beta}$ (using a noninformative prior distribution $\pi(\boldsymbol{\beta}) \propto 1$), the integrated likelihood function for the variance parameters is obtained as

$$\begin{aligned} L^I(\sigma^2, \lambda, \xi) &= \int_B L(\boldsymbol{\beta}, \sigma^2, \lambda, \xi) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} \\ &\propto \sigma^{-(k-q)} \mid \mathbf{M} \mid^{-1/2} \mid \tilde{\mathbf{X}}^T \mathbf{M}^{-1} \tilde{\mathbf{X}} \mid^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \tilde{\mathbf{S}}_\xi^2 \right), \end{aligned}$$

where $\tilde{\mathbf{S}}_\xi^2 = \mathbf{W}^T \tilde{\mathbf{Q}} \mathbf{W}$, $\tilde{\mathbf{Q}} = \mathbf{M}^{-1} \tilde{\mathbf{P}}$, $\tilde{\mathbf{P}} = \mathbf{I} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \mathbf{M}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{M}^{-1}$, and $\mathbf{I}$ is the identity matrix.

It is also possible to integrate with respect to the scale parameter σ^2, assuming, *a priori*, that it follows an inverse gamma distribution with shape parameter α and scale parameter r

$$\pi(\sigma^2) \propto \frac{1}{\sigma^{\alpha+1}} \exp\left(-\frac{r}{2\sigma^2} \right).$$

Then, the integrated likelihood for ξ and λ is (see Paulo (2005)):

$$\begin{aligned} L^I(\xi, \lambda) &= \int_0^\infty L^I(\sigma^2, \lambda, \xi) \pi(\sigma^2) d\sigma^2 \\ &\propto \mid \mathbf{M} \mid^{-1/2} \mid \mathbf{X}^T \mathbf{M}^{-1} \mathbf{X} \mid^{-1/2} \frac{1}{\left(\tilde{\mathbf{S}}_\xi^2 + r \right)^{\frac{n-q}{2} + \alpha}}. \end{aligned} \tag{2.9}$$

Expression (2.9) can be also interpreted as the joint density of the observations $(w_1, \ldots, w_k)$ conditionally on the hyperparameters ξ, λ.

In practice, the GP modelling assumptions may not hold (e.g. the bias of the estimator can be asymmetric, the tails may be non-Gaussian, the variance may vary over the parameter space). In order to consider the case of a varying covariance structure depending on the parameter space, we may define

$$\mathbf{W} \sim \mathcal{N}\left(\mathbf{g}(\mathbf{x})\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma}_\xi + \kappa^2 \tilde{\mathbf{I}} \exp(\varsigma(\mathbf{x})) \right),$$

where $\varsigma(\mathbf{x})$ is again modelled as a GP: $\varsigma(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathcal{K}^*(x, x'))$. The variance of the Fisher's transform of the copula functional is allowed to vary smoothly as a function of the covariate; the exponential function is used since the variance needs to be positive, then the logarithm is modelled as a GP. Again, $\mathcal{K}^*$ can be chosen to be the squared exponential kernel. The covariance matrix may be reparametrized by fixing either σ^2 or κ^2 equal to one in order to make the model identifiable. Model selection techniques can be used to select the functional form for $\mathbf{g}(\cdot)$ and $\varsigma(\cdot)$; see, for example, Vehtari and Ojanen (2012).

The posterior sample of (ξ, λ) is then used to approximate the posterior predictive distribution of W at new locations x^*, using the following expression:

$$f\left(w^*|x^*,\mathbf{W}\right)=\int\limits_{\xi,\lambda} f\left(w^*,\xi,\lambda|x^*,\mathbf{W}\right)d\xi d\lambda$$
$$=\int\limits_{\xi,\lambda} f\left(w^*|x^*,\xi,\lambda\right)f(\xi,\lambda|\mathbf{W})d\xi d\lambda.$$

This approximation can be computed for each value x^* defined over a grid. A natural summary value for the predictive distribution of W^* is the expected value $\mathbb{E}(W|x^*)\approx\bar{W}^*(x^*)=\sum_{t=1}^{N}w_t^* f(w_t^*|x^*)$, where N is the number of Monte Carlo simulations, which is also a point estimate of $W^*=Z^*+\varepsilon^*\approx Z^*$.

Finally, the estimation of the functional of interest φ can be obtained as

$$\hat{\varphi}(x^*)=\frac{\exp\{2\bar{W}^*(x^*)\}-1}{\exp\{2\bar{W}^*(x^*)\}+1}. \tag{2.10}$$

Alternatively, a full sample of $\varphi(x^*)$ can be derived from the posterior predictive distribution of $W^*(x^*)$.

The need for replications for each level of the covariate may be unrealistic in practice, and it means that this method is affected by the course of dimensionality: in order to reach a constant level of accuracy, more and more replications will be needed for each combination of covariate levels. In particular, in case of continuous covariates, it is necessary to identify group levels, and the grouping may be arbitrary. A similar comment applies to the Bayesian spline method described in Section 2.4.

2.3. Empirical likelihood approach

We now present a method based on a Bayesian use of the empirical likelihood. This approach represents a natural extension of the work of Grazian and Liseo (2017) to the conditional case and it is a novel contribution.

The assumption of normality of the Fisher's transform of the copula functionals may be admittedly too strict in many situations, in particular because the bias of the conditional copula estimator is not analytically known.

In recent years, there has been an interest in finding ways to derive the posterior distribution of the parameters of a model by substituting the likelihood function with an approximation. In this setting, Price et al. (2018) propose to use the approximation provided by a synthetic likelihood, for which the distribution of some (not necessarily sufficient) summary statistics of the model is assumed to be Gaussian. Elsewhere (see, for example, Gutmann and Corander, 2016) simulator-based models are used, which compare the observed datasets with datasets generated from the model; the likelihood function is then approximated by assuming a specific model – for example, a Gaussian distribution – for the discrepancy between observed and simulated data (possibly evaluated in terms of summary statistics). Another proposal is available in Mengersen et al. (2013), where the empirical likelihood is employed as a nonparametric approximation of the likelihood function of the parameter of interest. See Grazian and Fan (2020) for a recent review of these approaches.

Recently, Grazian and Liseo (2017) have introduced the use of the empirical likelihood in the specific setting of copula models, by proposing a semiparametric procedure where the posterior distribution of (low-dimensional) functionals of the dependence is derived and the structure of dependence of the underlying joint multivariate distribution is taken as a nuisance parameter defined on a infinite-dimensional space. This approach has two main advantages: i) in many settings, the interest lies in particular indices of the dependence (Spearman's ρ, Kendall's τ or tail dependence indices) which may be in complex relationship with the parameters of the copula and therefore it can be difficult to derive a likelihood function for them and ii) the selection of a specific copula family can be difficult in applied contexts and a semiparametric approach would avoid the need of choosing among alternative copula models. Grazian and Liseo (2017) has derived an approximation of the posterior distribution $\pi(\varphi;\mathbf{y})$ of the functional of interest φ using

$$\pi(\varphi|\mathbf{y})\propto\pi(\varphi)\hat{L}(\varphi;\mathbf{y}),$$

where $\pi(\varphi)$ is a prior distribution and $\hat{L}(\cdot)$ is a nonparametric approximation of the likelihood function. The choice of the prior distribution for ϕ is relatively easy in this setting, since functionals of the dependence are usually defined in a compact space, e.g. the Spearman's ρ and the Kendall's τ are defined in $[-1,1]$ and tail dependence indices are defined in $[0,1]$; however a formal objective prior for these quantities has not yet been derived in the literature. Therefore we adopt, throughout the paper, uniform prior distributions, defined on these compact spaces, as a surrogate of an ignorance prior.

Grazian and Liseo (2017) uses the exponentially tilted empirical likelihood proposed by Schennach (2005). This version of the empirical likelihood allows for a Bayesian interpretation involving an implicit nonparametric process prior on the infinite-dimensional nuisance parameter (the copula structure). Other version of the empirical likelihood (Owen, 2001) can be used: see, for example, Mengersen et al. (2013). This approach produces a good approximation of the posterior distribution (and of the likelihood function) if the generalized moment condition - which is implicit in the maximization problem associated with the definition of the empirical likelihood - is satisfied; this condition can be interpreted as a sort of unbiasedness requirement. In order to achieve this goal, a consistent estimator of the quantity of interest φ is needed; while this is easy to do in an unconditional setting, common estimators of the conditional copula – upon which the estimators for the functionals are built – have been shown to display some degree of inconsistency (Gijbels et al., 2011).

Suppose the model $F_{\mathbf{Y},\psi}(\cdot)$ as expressed in (1.1) for a multivariate random variable $\mathbf{Y}$ is indexed by a parameter ψ which can be defined as $\psi = (\varphi, \nu)$ and the interest of the analysis is in φ while ν is considered as a nuisance parameter. Then the goal of the analysis is to derive the posterior distribution of φ

$$\pi(\varphi|\mathbf{y}) = \int_N \frac{f(\mathbf{y}|\nu,\varphi)\pi(\nu|\varphi)\pi(\varphi)}{m(\mathbf{y})} d\nu,$$

where $\varphi \in \Phi$ and $\nu \in N$, $\pi(\nu|\varphi)\pi(\varphi)$ is the joint prior distribution of (φ, ν) and $m(\mathbf{y})$ is the distribution of the data $\mathbf{y}$ marginalized over the parameter space.

In the specific setting of this work, we assume that the copula $C(\cdot)$ is parametrized by (φ, C^*) where φ is some functional of the dependence in the joint distribution and C^* belongs to an infinite dimensional metric space (H, d_H). The empirical likelihood, in particular its Bayesian exponentially tilted version proposed by Schennach (2005), L_{BEL}, may be seen as the derivation of the integrated likelihood for φ

$$L_{BEL}(\varphi; \mathbf{y}) \propto \int_N L(\varphi, \nu; \mathbf{y})\pi(\nu|\varphi)d\nu,$$

where the prior distribution $\pi(\nu|\varphi)$ is a stochastic process constructed in such a way to give preference to distributions with a large entropy; following Schennach (2005), the empirical likelihood is defined in terms of a vector of weights $\{\omega_i\}_{i=1}^n$ obtained as the solution (for each fixed value of φ), of the maximization problem

$$\max_{(\omega_1,\ldots,\omega_n)} \sum_{i=1}^n \left(-\omega_i \log \omega_i\right),$$

under the constraints $0 \le \omega_i \le 1$ for $i = 1, \ldots, n$, $\sum_{i=1}^n \omega_i = 1$, and a moment constraint

$$\sum_{i=1}^n q(\mathbf{y}_i, \varphi)\omega_i = 0,$$

for some function q. A common choice for $q(\cdot)$ is a moment condition such as $\mathbb{E}[\phi - \hat{\phi}] = 0$; therefore $\hat{\phi}$ should be at least a consistent estimator of ϕ for the moment condition to be respected. The approximate Bayesian procedure for deriving an approximate posterior distribution for φ is then defined by i) selecting a prior distribution $\pi(\varphi)$; ii) selecting a nonparametric estimator $\hat{\varphi}$, which must be, at least, asymptotically unbiased; iii) computing the empirical likelihood $L_{BEL}(\varphi)$; iv) deriving the posterior distribution $\pi(\varphi|\mathbf{y})$ through a simulation process. Unfortunately, it is not always easy to perform step ii), i.e. to find an (at least, asymptotically) unbiased estimator of the quantity of interest, in order to satisfy the moment condition. In the setting of conditional functional of the dependence, as we have stated in Section 1, nonparametric estimators of the copula are biased, then a fully nonparametric approach cannot be directly implemented by using the empirical likelihood approximation, along the lines of Grazian and Liseo (2017).

The Fisher's transform of the observations can be defined as a function of the covariates, through a Taylor's expansion in terms of a polynomial of degree p. Assume that $W(\cdot)$ is differentiable p times in a neighbourhood of an interior point x_0; then

$$W(x_h) \approx W(x_0) + W(x_0)'(x_h - x_0) + \ldots + \frac{W(x_0)^{(p)}}{p!}(x_h - x_0)^p \equiv \mathbf{x}_{h,x_0}^{*T}\boldsymbol{\beta}, \tag{2.11}$$

where $\mathbf{x}_{h,x_0}^{*T} = (1, (x_h - x_0), \ldots, (x_h - x_0)^p)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$.

Similarly, a spline approximation, along the lines of Craiu and Sabeti (2012), can be adopted, using a cubic spline as a model for the calibration function of the copula parameter

$$W(x_h) \approx \sum_{\iota=1}^3 \alpha_\iota x_h^\iota + \sum_{s=1}^S \delta_s (x_h - \gamma_s)_+^3, \tag{2.12}$$

where $a_+ = \max(0, a)$ and $\{\gamma_s\}_{s=1}^S$ is a vector of knots. In the approach of Craiu and Sabeti (2012), the copula family is chosen through a model selection procedure. Then, consistent estimators of the coefficients of the Taylor's expansion or of the spline function, which can be used to define the moment condition in the definition of the empirical likelihood, can be derived.

For the parameters $\boldsymbol{\beta}$ of the linearised model, it is common to define weakly informative priors, such as, for instance, $\mathcal{N}(0, \sigma^2)$, where σ^2 is some large value, and combine them with the weights defined by the empirical likelihood approach described in Grazian and Liseo (2017) in order to obtain a sample of size G of $(\beta_0, \ldots, \beta_l | w_1, \ldots, w_k)$ such that

$$\begin{bmatrix} \beta_0^{(1)}, \ldots, \beta_l^{(1)}, \omega^{(1)} \\ \vdots \\ \beta_0^{(G)}, \ldots, \beta_l^{(G)}, \omega^{(G)} \end{bmatrix},$$

where ω_g, with $g = 1, \ldots, G$, represents the weights of $(\beta_0^{(g)}, \ldots, \beta_l^{(g)})$ induced by the empirical likelihood. The combinations of $(\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(G)})$ and the weights $(\omega_1, \ldots, \omega_G)$ produce an approximation of the joint posterior distribution for $\boldsymbol{\beta}$. On the other hand, if a cubic spline as in Equation (2.12) is used, weakly informative priors for the parameters $\{\alpha_\iota\}_{\iota=1}^{3}$ and $\{\delta_s\} =_{s=1}^{S}$ can be considered, such as, for example, independent $\mathcal{N}(0, \sigma^2)$ for some large value σ^2.

The approximate posterior distribution of $\boldsymbol{\beta}$ can be used to approximate the posterior predictive distribution

$$\begin{aligned} f\left(w^*|x^*, \mathbf{W}\right) &= \int_{\mathbf{B}} f\left(w^*, \boldsymbol{\beta}|x^*, \mathbf{W}\right) d\boldsymbol{\beta} \\ &= \int_{\mathbf{B}} f\left(w^*|x^*, \boldsymbol{\beta}\right) f(\boldsymbol{\beta}|\mathbf{W}) d\boldsymbol{\beta} \\ &\approx \sum_{g=1}^{G} f\left(w^*|x^*, \boldsymbol{\beta}^{(g)}\right) \bar{\omega}^{(g)}, \end{aligned}$$

where $\bar{\omega}^{(g)} = \frac{\omega^{(g)}}{\sum_{g=1}^{G} \omega^{(g)}}$ is the normalized weight for the g-th iteration. As in Section 2.2, the approximation can be computed on a grid of possible values of x^* and the estimator of the functional of interest can be derived as in Equation (2.10).

2.4. Bayesian splines approach

Finally, we present the approach based on Bayesian splines. Similarly to the approach proposed in Section 2.2, the methodology is already available in the literature, however, up to our knowledge, it has never been applied to the specific setting of conditional copula.

An alternative to the empirical likelihood approach of Section 2.3, still avoiding parametric assumptions, is to model the functional of the dependence as a regression function using regression splines. In this setting, it is possible to introduce assumptions on the type of function, i.e. its shape and degree of smoothness.

Again, it is useful to work with a transformation of the functional of the dependence, as defined in Equation (2.5):

$$Z(x) = f(x) + \varepsilon,$$

where f is assumed to be smooth; in addition, constraints to force the function to be monotone or convex can be included.

For a given vector of m knots, $a = d_0 < \ldots < d_{m-1} = b$, spline basis functions $\mathbf{s}(x) = (s_1(x), \ldots, s_m(x))$ are defined on $[a, b]$. Possible choices of splines are quadratic I-splines (Ramsay et al., 1988), cubic I-splines, and C-splines (Meyer, 2008).

The function $f(\cdot)$ of the predictor x is modelled as

$$f(x) = \sum_{j=1}^{m} \beta_j \boldsymbol{\delta}_j.$$

The $\boldsymbol{\delta}_j$ for $j = 1, \ldots, m$ are basis vectors corresponding to the shape-restricted basis function for f. The coefficients β_j are supposed to follow a priori a normal distribution with zero mean and large variance M. In addition, a vague gamma prior distribution can be assumed for the precision of the error term. For a full description of the method used in the implementation, we refer the reader to Meyer et al. (2011).

Prior distributions for the coefficients β_j, and to impose shape restrictions, the prior for the coefficients can be restricted to be positive. Noninformative camma prior distributions for the coefficients are chosen.

Shape-restricted regression splines have been introduced by Meyer (2008). By using I-splines and C-splines, the shape-restrictions are imposed by constraining the spline basis functions to be non-negative. This method has been shown to be robust in terms of choices of the knots, provided that the number of knots is large enough to capture the behaviour of the data. In this work, we have associate the number of knots to the sample size, since Stone and Huang (2002) proved that the optimal number of knots for unrestricted regression splines is $n^{1/(2r+1)}$ where r is the order of the regression splines (here $r = 3$).

The Bayesian estimation of the regression coefficients and the knots locations can be implemented via Gibbs sampler.

In presence of multiple covariates, an additive regression model can be used, so that

$$Z(x_1, \ldots, x_p) = f_1(x_1) + \ldots + f_p(x_p) + \varepsilon,$$

and each function f_k can be estimated with by imposing shape restrictions (which may differ from function to function).

Table 1
Frequencies of correct selection of the copula model using leaving-one-out cross-validation, where ρ is a linear function of the covariate. The rows represent the copula the data have been simulated from, and the columns the possible models.

	Clayton	*Frank*	*Gumbel*	*Joe*	*Gaussian*	*Plackett*	*Student t*
Clayton	42	0	0	0	0	1	7
Frank	0	10	0	0	0	38	3
Gumbel	0	0	39	0	0	0	11
Gaussian	0	0	0	0	8	0	42

Table 2
Frequencies of correct selection of the copula model using leaving-one-out cross-validation, where ρ is a sine function of the covariate. The rows represent the copula the data have been simulated from, and the columns the possible models.

	Clayton	*Frank*	*Gumbel*	*Joe*	*Gaussian*	*Plackett*	*Student t*
Clayton	39	0	0	0	0	0	11
Frank	5	1	7	21	1	0	15
Gumbel	0	0	6	0	0	0	44
Gaussian	1	0	6	27	0	0	16

3. Simulation study

Here we present a detailed simulation study, including set ups with one and two covariates.

3.1. One covariate

Values of the covariate were simulated from a uniform distribution and the Spearman's ρ was associated to the covariate through a linear or a sine relationship

- $\rho = 0.8x - 2$ with $x \sim \mathcal{U}\text{nif}(2, 5)$;
- $\rho = \sin(x)$ with $x \sim \mathcal{U}\text{nif}(-5, 5)$.

For each model, we had 20 levels of the covariate. We simulated two-dimensional pseudo-observations from four different copula models: Clayton, Frank, Gumbel and Gaussian. For each level of the covariate, we simulated 100 data points with specific parameter values.

Each simulation set-up was repeated 50 times. We first performed a model selection approach based on the `xvCopula` function of the `R` `copula` package (Kojadinovic et al., 2010), performing the leave-one-out cross-validation for a set of hypothesized parametric copula models, using maximum pseudo-likelihood estimation. Specifically, we considered, as potential generating models, the following copulas: Clayton, Frank, Gumbel, Joe, Gaussian, Plackett, and Student t. Table 1 and 2 show the frequencies of correct selection of the copula model, when ρ is defined as a linear function and as a sine function respectively. It is apparent that the selection of the copula model is not expected to be an easy task. On the other hand, methods based on an assumption about the copula model describing the data, strongly rely on this assumption for inference on the functionals of dependence, such as Spearman's ρ, Kendall's τ, or tail dependence indices. Moreover, when selecting the copula based on the observed data, one often assumes that the copula does not change with the level of the covariate; conversely, when this possibility is taken into account, the number of observations for each level of the covariate are limited, and the selection problem becomes more difficult. As shown in Grazian and Liseo (2017), a parametric method may outperform nonparametric methods when the correct model can be selected; therefore, if enough information is available to select the correct copula model a parametric method should be used, however, when information is limited Grazian and Liseo (2017) also shows that estimating functionals of the dependence under miss-specification may lead to strong biases and, in this situation, a nonparametric method may be preferred in terms of robustness.

Tables 3, 4, 5 and 6 show the results of the comparison among the four approaches described in Section 2: GPs, the empirical likelihood method based on an inconsistent estimator of the copula, the empirical likelihood method based on a linearised model for the functional of the dependence, and Bayesian splines. An approximation of the integrated mean squared error (IMSE) is used to compare the approaches, where the integration is taken with respect to all the possible samples and all the possible covariate levels:

$$IMSE = \int_{\mathcal{Y}} \int_{\mathcal{X}} (\hat{\rho}(x, y) - \rho(x, y))^2 dxdy \approx \sum_{i=1}^{50} \sum_{\ell=1}^{k} (\hat{\rho}(x_\ell, \mathbf{y}_i) - \rho(x_\ell, \mathbf{y}_i))^2,$$

where $\rho(x_\ell, \mathbf{y}_i)$ is the true value of Spearman's ρ for covariate level x_ℓ, evaluated at the sample value $\mathbf{y}_i = (y_{1,i}, \ldots, y_{n,i})$ and $\hat{\rho}(x_\ell, \mathbf{y}_i)$ is the corresponding estimate obtained with each of the four approaches under analysis. An analogous expres-

Table 3
Results for the nonparametric analyses with Spearman's ρ as a linear function. "GP" denotes the Gaussian process model of Section 2.2; "Incons. EL" denotes the empirical likelihood approach using the inconsistent estimator (2.4), with Local Linear (LL) or Nadaraya-Watson (NW) weights and triweight or Gaussian kernel; "EL - linearised model" denotes the empirical likelihood approach of Section 2.3 using the approximation (2.11); "Bayes Splines" denotes the approach of Section 2.4.

		Clayton	Frank	Gumbel	Gaussian
GP	*IMSE*	0.009	0.003	0.004	0.006
	(Ave) CI Length	0.300	0.305	0.270	0.293
	(Ave) CI coverage	0.910	1.000	0.978	0.891
Incons. EL - LL, triweight	*IMSE*	0.260	0.297	0.276	0.267
	(Ave) CI Length	1.326	1.298	1.309	1.328
	(Ave) CI coverage	0.717	0.671	0.725	0.760
Incons. EL - NW, triweight	*IMSE*	0.277	0.282	0.262	0.260
	(Ave) CI Length	1.318	1.316	1.304	1.313
	(Ave) CI coverage	0.728	0.727	0.730	0.739
Incons. EL - LL, Gaussian	*IMSE*	0.298	0.283	0.274	0.261
	(Ave) CI Length	1.299	1.314	1.299	1.317
	(Ave) CI coverage	0.723	0.718	0.704	0.775
Incons. EL - NW, Gaussian	*IMSE*	0.280	0.290	0.302	0.230
	(Ave) CI Length	1.321	1.295	1.306	1.328
	(Ave) CI coverage	0.000	0.729	0.721	0.734
EL - linearised model	*IMSE*	1.097	1.039	1.065	0.926
	(Ave) CI Length	2.000	2.000	2.000	2.000
	(Ave) CI coverage	1.000	1.000	1.000	1.000
Bayes Splines	*IMSE*	0.003	0.003	0.004	0.002
	(Ave) CI Length	0.166	0.169	0.155	0.164
	(Ave) CI coverage	0.975	0.985	0.880	0.970

sion for the IMSE can be obtained for Kendall's τ. In addition to the IMSE values, Tables 3, 4, 5 and 6 list the average length and average coverage results for the credible intervals of level 95%.

Table 3 displays the results with Spearman's ρ as linear function $\rho(x) = 0.8x - 2$, while Table 4 lists the results with Kendall's τ as linear function $\tau(x) = 0.8x - 2$. The best performance is obtained using the GPs (Section 2.2) and Bayesian splines (Section 2.4) based approaches, where in all cases, the coverage is close to the expected one and the IMSE is lower than 0.01. The results based on empirical likelihood approaches are always worse than those based on GPs and Bayesian splines. When using the inconsistent estimator (2.4), the results are relatively similar, independently of the particular choice of the kernel (Gaussian or triweight) or weights (local-linear or Nadaraya-Watson): the interval coverage is constantly lower than the expected one (around 70%) and the IMSE is around 0.3, definitely larger than the previous cases; this behaviour may highlight the bias induced by the use of an inconsistent estimator in the empirical likelihood approach. Finally, when using the empirical likelihood approach based on the Taylor expansion (2.11), with empirical likelihood weights on the β coefficients, the variance of the estimates increases and the credible intervals cover all the parameter space.

Similarly to Tables 3 and 4, Tables 5 and 6 report the performance study of the different methods when the true functions are $\rho(x) = \sin(x)$ (Table 5) and $\tau(x) = \sin(x)$ (Table 6) respectively. Again, the best performance is achieved by the GPs and the Bayesian splines approaches. The Bayesian splines show a coverage which is similar or higher than the one obtained with the GPs approach, with only slightly longer intervals on average. The methods based on empirical likelihood approximations seem inconsistent, with larger intervals lengths and reduced coverage.

Since the methods based on Gaussian processes and Bayesian splines require replications for each level of the covariate, they tend to underperform with respect to the method based on empirical likelihood when there are only few replications available. Table 7 shows that when the number of replications are low the method based on the empirical likelihood beats the other two methods (similar results are obtained with different choices of weights), while as the number of replicates increases GP and Bayesian splines tend to outperform EL. Howers, when the covariate is continuous and there are no replicates Bayesian splines and GP cannot be obtained, being based on a first computation of an estimate of the dependence index. Similar results are obtained when the true underlying function is a sin function and when observations are simulated from different copulas.

3.2. Two covariates

We have performed simulations also considering functionals depending on two covariates, x_1 and x_2. Recall that some of the methods investigated in this paper require replications for each combination of the covariate levels, therefore they are not applicable when using continuous covariates, unless values are grouped into classes. Here, we compare the results with

Table 4
Results for the nonparametric analyses with Kendall's τ as a linear function. "GP" denotes the Gaussian process model of Section 2.2; "Incons. EL" denotes the empirical likelihood approach using the inconsistent estimator (2.4), with Local Linear (LL) or Nadaraya-Watson (NW) weights and triweight or Gaussian kernel; "EL - linearised model" denotes the empirical likelihood approach of Section 2.3 using the approximation (2.11); "Bayes Splines" denotes the approach of Section 2.4.

		Clayton	Frank	Gumbel	Gaussian
GP	*IMSE*	0.008	0.000	0.009	0.009
	(Ave) CI Length	0.187	0.174	0.217	0.217
	(Ave) CI coverage	0.623	1.000	0.771	0.771
Incons. EL - LL, triweight	*IMSE*	0.264	0.255	0.242	0.242
	(Ave) CI Length	1.269	1.303	1.337	1.337
	(Ave) CI coverage	0.703	0.737	0.773	0.773
Incons. EL - NW, triweight	*IMSE*	0.322	0.257	0.244	0.244
	(Ave) CI Length	1.340	1.320	1.319	1.319
	(Ave) CI coverage	0.698	0.715	0.751	0.751
Incons. EL - LL, Gaussian	*IMSE*	0.297	0.283	0.240	0.240
	(Ave) CI Length	1.291	1.324	1.311	1.311
	(Ave) CI coverage	0.711	0.753	0.743	0.743
Incons. EL - NW, Gaussian	*IMSE*	0.326	0.294	0.301	0.301
	(Ave) CI Length	1.369	1.278	1.243	1.243
	(Ave) CI coverage	0.000	0.000	0.000	0.000
EL - linearised model	*IMSE*	0.986	0.735	0.966	0.966
	(Ave) CI Length	2.000	2.000	2.000	2.000
	(Ave) CI coverage	1.000	1.000	1.000	1.000
Bayes Splines	*IMSE*	0.000	0.001	0.003	0.003
	(Ave) CI Length	0.141	0.137	0.153	0.153
	(Ave) CI coverage	1.000	0.967	0.917	0.917

Table 5
Results for the nonparametric analyses with Spearman's ρ as a sine function. "GP" denotes the Gaussian process model of Section 2.2; "Incons. EL" denotes the empirical likelihood approach using the inconsistent estimator (2.4), with Local Linear (LL) or Nadaraya-Watson (NW) weights and triweight or Gaussian kernel; "EL - linearised model" denotes the empirical likelihood approach of Section 2.3 using the approximation (2.11); "Bayes Splines" denotes the approach of Section 2.4.

		Clayton	Frank	Gumbel	Gaussian
GP	*IMSE*	0.504	0.492	0.428	0.366
	(Ave) IC Length	1.214	1.001	1.007	1.258
	(Ave) IC coverage	0.715	0.862	0.892	0.787
Incons. EL - LL, triweight	*IMSE*	1.620	1.972	1.762	1.672
	(Ave) CI Length	1.326	1.298	1.309	1.328
	(Ave) CI coverage	0.175	0.185	0.253	0.600
Incons. EL - NW, triweight	*IMSE*	1.217	1.823	1.612	1.601
	(Ave) CI Length	1.518	1.161	1.324	1.413
	(Ave) CI coverage	0.482	0.437	0.478	0.416
Incons. EL - LL, Gaussian	*IMSE*	1.281	1.238	1.245	1.621
	(Ave) CI Length	1.229	1.314	1.229	1.316
	(Ave) CI coverage	0.623	0.628	0.654	0.685
Incons. EL - NW, Gaussian	*IMSE*	1.202	1.270	1.202	1.236
	(Ave) CI Length	1.222	1.198	1.273	1.210
	(Ave) CI coverage	0.410	0.387	0.321	0.344
EL - linearised model	*IMSE*	1.437	1.071	1.242	1.417
	(Ave) IC Length	2.000	2.000	2.000	2.000
	(Ave) IC coverage	1.000	1.000	1.000	1.000
Bayes Splines	*IMSE*	0.541	0.502	0.720	0.541
	(Ave) IC Length	1.229	1.071	1.051	1.206
	(Ave) IC coverage	0.931	1.000	0.825	0.912

Table 6
Results for the nonparametric analyses with Kendall's τ as a sine function. "GP" denotes the Gaussian process model of Section 2.2; "Incons. EL" denotes the empirical likelihood approach using the inconsistent estimator (2.4), with Local Linear (LL) or Nadaraya-Watson (NW) weights and triweight or Gaussian kernel; "EL - linearised model" denotes the empirical likelihood approach of Section 2.3 using the approximation (2.11); "Bayes Splines" denotes the approach of Section 2.4.

		Clayton	Frank	Gumbel	Gaussian
GP	*IMSE*	0.645	0.293	0.563	0.494
	(Ave) CI Length	0.391	0.732	0.949	1.577
	(Ave) CI coverage	0.518	0.549	0.398	0.792
Incons. EL - LL, triweight	*IMSE*	0.144	0.135	0.350	0.131
	(Ave) CI Length	0.809	0.831	0.747	0.849
	(Ave) CI coverage	0.320	0.380	0.260	0.420
Incons. EL - NW, triweight	*IMSE*	0.805	0.809	0.797	0.882
	(Ave) CI Length	1.134	1.320	1.141	1.102
	(Ave) CI coverage	0.320	0.340	0.400	0.340
Incons. EL - LL, Gaussian	*IMSE*	0.303	0.305	0.457	0.309
	(Ave) CI Length	0.598	0.754	0.687	0.662
	(Ave) CI coverage	0.320	0.260	0.280	0.200
Incons. EL - NW, Gaussian	*IMSE*	0.669	0.597	0.806	0.605
	(Ave) CI Length	0.673	0.544	0.767	0.579
	(Ave) CI coverage	0.340	0.320	0.360	0.320
EL - linearised model	*IMSE*	1.420	1.504	1.854	1.534
	(Ave) CI Length	2.000	2.000	2.000	2.000
	(Ave) CI coverage	1.000	1.000	1.000	1.000
Bayes Splines	*IMSE*	0.492	0.491	0.944	0.507
	(Ave) CI Length	1.039	1.003	0.517	1.066
	(Ave) CI coverage	0.400	0.330	0.100	0.350

the semi-parametric approach proposed by Levi and Craiu (2018): while such method requires the selection of the copula, it does not require repetitions for each level of the covariates. Here, following Levi and Craiu (2018), we use the conditional cross-validated pseudo marginal likelihood (CCVML) as selection tool; the CCVML considers the predictive distribution of one response given the rest of the data.

We generated again 50 repetitions of simulations of pseudo-observations from each of the four considered copulas: Clayton, Frank, Gumbel, and Gaussian. The methods can be performed for any functional of the dependence, however we focus here on the Kendall's τ without loss of generality. We fixed

$$\tau = 0.7 + 0.15 \sin(\sqrt{10}(x_1 + 3x_2)), \tag{3.1}$$

with x_1, x_2 independently and identically distributed from $\mathcal{U}\text{nif}(0,1)$. We adopted the application setting of the algorithm used in Scenario 1 of the paper by Levi and Craiu (2018). We considered 10 repetitions for each combination of the levels of x_1 and x_2. Table 8 shows the results of the selection task for the copula model, using the CCVML. Again, the identification of the correct copula model is not easy, in particular it seems that the Gaussian copula is poorly identified through CCVML.

Table 9 compares the results of the method proposed by Levi and Craiu (2018) with the methods presented in Section 2: GPs, the empirical likelihood method based on an inconsistent estimator of the copula, the empirical likelihood method based on a linearised model for the functional of the dependence and Bayesian splines. The method based on GPs shows the best performance in terms of average IMSE. In this case, Bayesian splines show slightly larger IMSE results than GPs, which may be due to the increased dimensionality of the problem.

To implement the method based on the inconsistent copula estimator, it is necessary to extend the definition of the Nadaraya-Watson weights given in Section 2. Specifically, for p covariates the weights are defined as

$$w_i(\mathbf{x}, \mathbf{h}_n) = \frac{\mathcal{K}\left(\frac{\mathbf{X}_i - \mathbf{x}}{\mathbf{h}_n}\right)}{\sum_{i=1}^n \mathcal{K}\left(\frac{\mathbf{X}_i - \mathbf{x}}{\mathbf{h}_n}\right)}, \tag{3.2}$$

where $\mathcal{K}(\mathbf{X}_i - \mathbf{x}) = \mathcal{K}_{h_1}(X_{i1} - x_1) \times \ldots \times \mathcal{K}_{h_p}(X_{ip} - x_p)$ for some choices of bandwidth $(h_1, \ldots, h_p)$. Again, we choose a triweight and a Gaussian kernel function.

Table 9 shows that the performance of the empirical likelihood method results in lower IMSE values than those yielded by the conditional method of Levi and Craiu (2018), but in larger IMSE values than those obtained with the methods based on GPs and splines. In addition, the implementation of the empirical likelihood approach is computationally intensive, and

Table 7
IMSE values obtained in the estimation process with observations simulated from a Clayton copula with 5, 10, 15, and 20 values of covariates (indicated in the columns) and between 2 and 50 replications for each covariate. GP stands for method based on Gaussian processes, EL stands for method based on empirical likelihood (values obtained with Gaussian kernel and local-linear weights), and SPL stands for method based on Bayesian splines.

n_x		5	10	15	20
2	*GP*	1.509	1.956	1.622	1.594
	EL	0.527	0.619	0.239	0.224
	SPL	1.892	1.726	1.613	1.552
3	*GP*	1.926	1.113	1.148	1.1092
	EL	0.304	0.472	0.139	0.127
	SPL	1.838	1.838	1.624	1.524
4	*GP*	0.520	0.586	0.579	0.516
	EL	0.321	0.244	0.253	0.028
	SPL	0.829	0.524	0.315	0.635
5	*GP*	0.482	0.407	0.403	0.409
	EL	0.140	0.051	0.138	0.032
	SPL	0.456	0.438	0.484	0.307
10	*GP*	0.122	0.107	0.109	0.103
	EL	0.301	0.082	0.016	0.012
	SPL	0.989	0.606	0.176	0.153
15	*GP*	0.016	0.020	0.023	0.020
	EL	0.058	0.060	0.037	0.006
	SPL	0.780	0.030	0.001	0.003
20	*GP*	0.246	0.211	0.210	0.203
	EL	0.239	0.031	0.012	0.008
	SPL	0.989	0.004	0.019	0.005
25	*GP*	0.057	0.017	0.004	0.001
	EL	0.103	0.026	0.018	0.021
	SPL	0.929	0.011	0.013	0.001
50	*GP*	0.016	0.003	0.001	0.003
	EL	0.030	0.029	0.008	0.008
	SPL	0.789	0.004	0.002	0.002

Table 8
Frequencies of correct selection of the copula model using CCVML, where τ is a function of two covariates, see Equation (3.1). The rows represent the copula the data have been simulated from, and the columns the possible models.

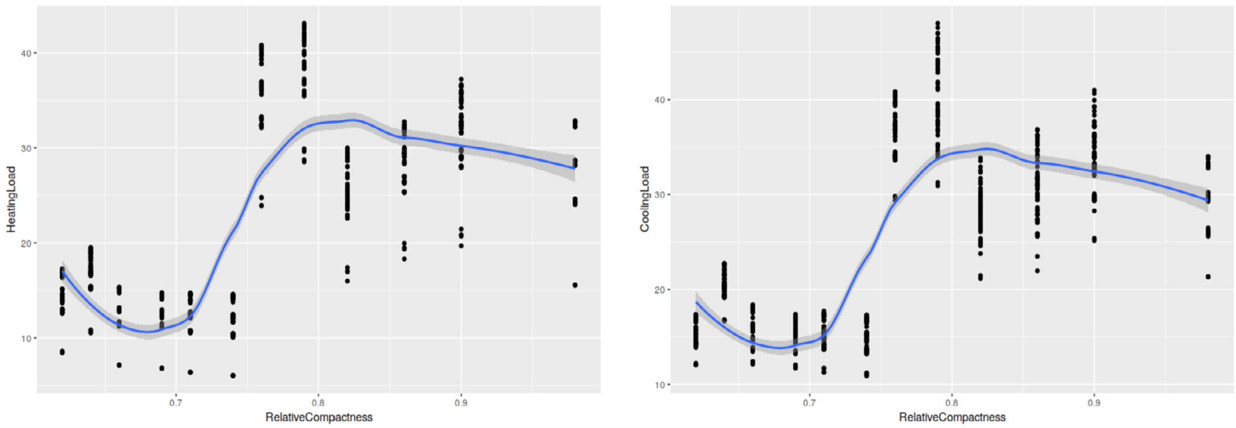
	Clayton	*Frank*	*Gaussian*	*Gumbel*
Clayton	26	11	13	0
Frank	13	15	22	0
Gumbel	0	0	10	40
Gaussian	3	10	1	36

it requires a limited number of covariates. Finally, as for Section 3.1, the performance of the method based on the empirical likelihood computed on the linearised model is the worst amongst the analysed methods.

Here recall that the approaches based on GPs and Bayesian splines require replications for each combination of covariate levels, while the semi-parametric approach of Levi and Craiu (2018) and the approaches based on the empirical likelihood can be applied when no replications are available. However, the semi-parametric conditional approach relies on the selection of the copula family (see Table 8).

Table 9
Integrated Mean Squared Errors (IMSE) for the nonparametric analyses of the Kendall's τ in presence of two covariates. "Semi-parametric (LC2018)" stands for the method proposed by Levi and Craiu (2018), while the other acronyms can be interpreted as in Table 3.

	Clayton	Frank	Gumbel	Gaussian
Semi-parametric (LC2018)	0.451	0.467	0.452	0.180
GP	0.002	0.002	0.003	0.004
Incons. EL - NW triweight	0.041	0.041	0.041	0.041
Incons. EL - NW Gaussian	0.042	0.042	0.041	0.041
EL - linearised model	1.057	1.061	1.061	1.001
Bayes Splines	0.009	0.009	0.009	0.009



(a) Heating load against relative compactness (b) Cooling load against relative compactness

Fig. 1. Scatterplots of heating and cooling load measures (plotted on the vertical axes) against relative compactness (on the horizontal axes), from the Energy Efficiency dataset. The solid line are obtained as smoothed conditional means, while the grey areas show the relevant confidence intervals.

4. Data examples

We now apply the methods described in this work to two data examples, the first in the area of civil engineering and the second in astrophysics, to compare the results of the analysis performed by the different methodologies on realistic problems.

4.1. *Energy efficiency*

The Energy Efficiency dataset includes simulations for 12 different building architectural interior environments, obtained with the Ecotect software (Roberts and Marsh, 2001). The buildings differ with respect to 8 features which, combined in different ways, lead to 768 building shapes. The response variables are the heating and the cooling loads, i.e. the amount of heat energy that would need to be added or removed to the space to maintain the temperature in the requested range, respectively. Lower thermal loads are indicators of higher energy efficiency. The data have been generated and analysed in Tsanas and Xifara (2012). The data are available at https://archive.ics.uci.edu/ml/datasets/Energy+efficiency.

An important feature when studying the energy efficiency of buildings is the relative compactness (RC), which is a measure to compare different building shapes through the surface to volume ratio (Pessenlehner and Mahdavi, 2003, Ourghi et al., 2007). Fig. 1 shows the scatterplots of the observed data of heating and cooling loads against RC. The solid line smoothed conditional mean, showing that the relationship is non-linear. We now analyse the effect on the dependence between heating and cooling loads with respect to the RC indicator.

We applied the approaches described in Section 2 to the Energy Efficiency dataset, after estimating each marginal non-parametrically.

Fig. 2 shows the approximation of the Spearman's ρ computed between heating and cooling loads as a function of RC. The blue lines show the approximation obtained with the GP method illustrated in Section 2.2 and the coral lines show results of the splines method illustrated in Section 2.4. The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. The red points are frequentist estimates for each of the levels of RC. From Fig. 2 it seems that the dependence between heating and cooling loads is strong for low and high levels of RC, respectively. For moderate levels of RC, instead, it seems that the strength of dependence is lower. This may be due to an increased variability in the data, as shown in Fig. 1. A comparison between the method based on GPs and the method based on splines shows that the approximation obtained through the GP method seems to better follow the frequentist estimates of Spearman's

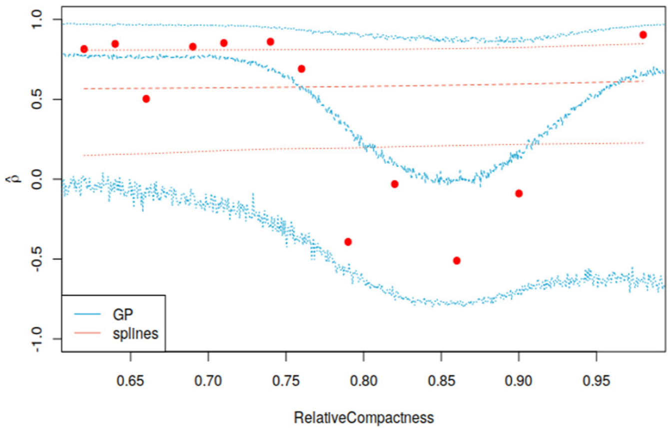


Fig. 2. Approximations of Spearman's ρ computed between heating and cooling loads, varying as a function of RC from the Energy Efficiency dataset, obtained with the GP method illustrated in Section 2.2 (blue lines) and the splines method illustrated in Section 2.4 (coral lines), respectively. The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. The red points are frequentist estimates for each of the levels of RC. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

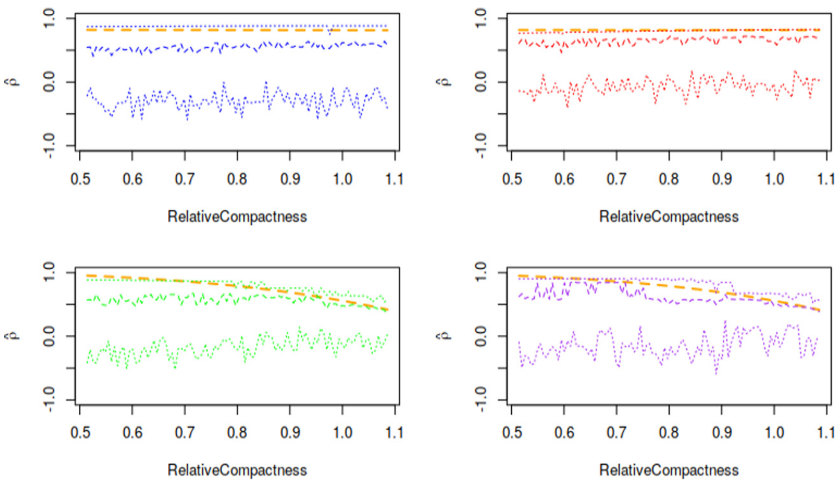


Fig. 3. Approximations of Spearman's ρ computed between heating and cooling loads, varying as a function of RC from the Energy Efficiency dataset, obtained with the method based on EL. The orange long-dashed lines denote the frequentist estimates obtained with the inconsistent estimator of the copula, the coloured dashed lines denote the posterior means obtained with the EL methods and the dotted lines denote the 95% credible intervals. Each figure represents the approximation obtained with NW weights and triweight kernel (top left), NW weights and Gaussian kernel (top right), LL weights and triweight kernel (bottom left), and LL weights and Gaussian kernel (bottom right).

ρ (red points in Fig. 2), while the approximation obtained through splines is less sensitive to changes in the value of the dependence. This may be due to the limited number of points for each level of the covariate, showing that GPs is able to follow changes in dependence with a lower number of data points.

Fig. 3 shows a similar approximation of the posterior distribution of Spearman's ρ obtained with the method based on the empirical likelihood (EL) illustrated in Section 2.3, using the inconsistent estimator of the copula. The orange long-dashed lines denote the frequentist estimates obtained with the inconsistent estimator of the copula, the coloured dashed lines denote the posterior means obtained with the EL methods and the dotted lines denote the 95% credible intervals. Each figure represents the approximation obtained with NW weights and triweight kernel (top left), NW weights and Gaussian kernel (top right), LL weights and triweight kernel (bottom left), and LL weights and Gaussian kernel (bottom right). From Fig. 3 it is evident that the approximation depends on the weights and kernel chosen in the estimator of the copula. All the approximations tend to be concentrated around large values of dependence, however methods based on local-linear weights show a decline of the dependence for large values of RC. Methods based on the EL for the parameters of the linearised model show highly variable posterior approximations of the parameters, that lead to estimates of Spearman's ρ which are very uncertain (with credible intervals including all the parameter space) and are not included here.

The presence of several building features in the Energy Efficiency dataset allows for the application of methods based on more than one covariate. In particular, we now focus the analysis on two covariates: RC and wall area (WA). Table 10 shows the frequency of the data for each combination of the RC and WA covariate levels. Fig. 4 shows the Spearman's

Table 10
Frequency table of Relative Compactness (on the rows) and Wall Area (on the columns) from the Energy Efficiency dataset.

		Wall Area						
		245.0	*269.5*	*294.0*	*318.5*	*343.0*	*367.5*	*416.5*
Relative Compactness	*0.62*						64	
	0.64					64		
	0.66				64			
	0.69			64				
	0.71		64					
	0.74	64						
	0.76							64
	0.79					64		
	0.82				64			
	0.86			64				
	0.90				64			
	0.98			64				

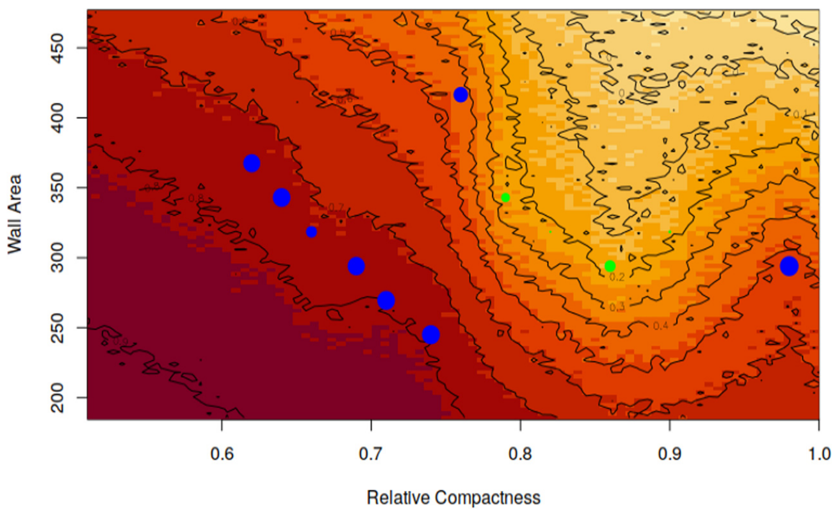


Fig. 4. Estimation of Spearman's ρ, with respect to the levels of WA (vertical axis) and RC (horizontal axis), for the Energy Efficiency dataset. Different colours denote the Spearman's ρ posterior conditional mean levels obtained by applying the GP method. Dark red denotes strong dependence, while light yellow denotes weak dependence levels. The dots represent the frequentist estimates of the Spearman's ρ computed as unconditional samples for the observations with that particular combination of the covariate levels. Blue dots denote positive ρs and green dots denote negative ρs. The size of the dots represents the scaled absolute value of the estimates.

ρ posterior conditional means obtained by applying the method based on GPs, with respect to the levels of WA and RC. Dark red denotes strong dependence, while light yellow denotes weak dependence levels. The dots represent the frequentist estimates of the Spearman's ρ computed as unconditional samples for the observations with that particular combination of the covariate levels. Blue dots denote positive ρs and green dots denote negative ρs. The size of the dots represents the scaled absolute value of the estimates.

Due to the data setting, as illustrated in Table 10, the implementation of the Bayesian splines is made more difficult for the limited amount of data points for each of the few observed combinations of covariates and the method does not converge. Similarly, the EL computed on the multivariate versions of the NW or LL weights tends to perform poorly, with many weights being close or equal to zero, which influences the overall approximation of the likelihood function.

In conclusion, in this example the only method applicable with two covariates is the one based on GPs amongst the methods presented in this work. However, it is clear that, as the number of covariates (or the number of levels in each covariate) increases, the method would require an increasing number of observations, and this reduces the applicability in high-dimensional settings.

4.2. MAGIC gamma telescope

The Cherenkov gamma telescope observes high energy gamma rays, detecting the radiation emitted by charged particles produced inside electromagnetic showers. Photons are collected in patterns forming the shower image and it is necessary to discriminate between the image caused by primary gamma rays and the one caused by other cosmic rays. Images are

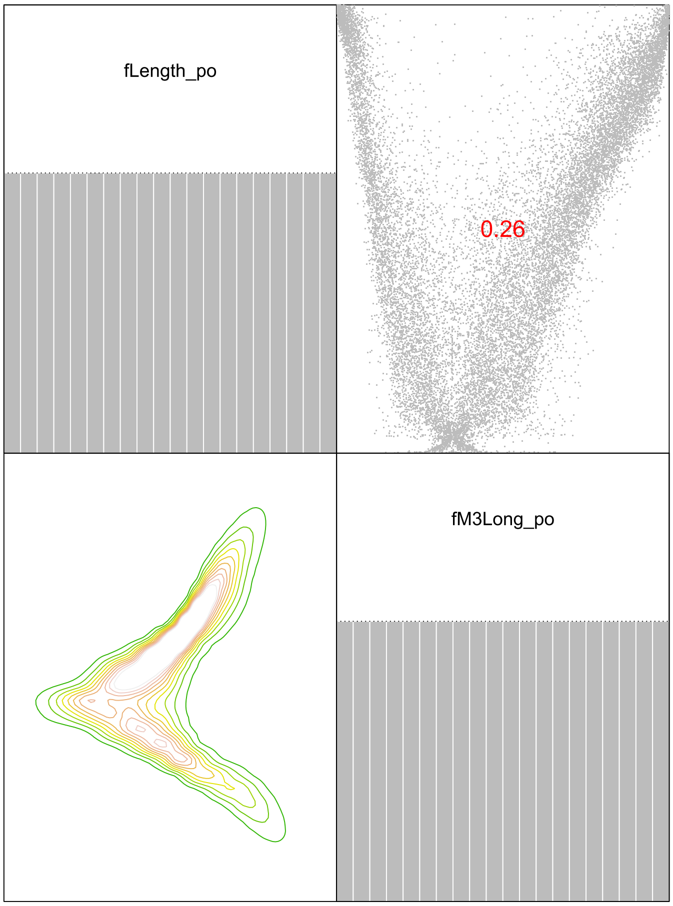


Fig. 5. Empirical normalized contour plot of the variables `Length` and `M3Long` (bottom left panel) from the MAGIC Gamma Telescope dataset; scatterplot and correlation between the pseudo-observations (top right panel); histograms of `Length` (top left panel) and `M3Long` (bottom right panel), after transforming the variables into pseudo-observations.

usually ellipses and their features, in terms of measures associated with the long and short axes, help to discriminate amongst images.

The data used in this Section are simulations of ellipses parameters generated by the Monte Carlo program Corsika (Heck et al., 1998) to simulate registrations of high energy gamma particles in a Cherenkov gamma telescope, called MAGIC (Major Atmospheric Gamma Imaging Cherenkov) telescope located on the Canary islands. The dataset is formed by 19,020 observations with 11 variables. For a full description of the dataset and the evaluation of the performance of several classification methods applied to the data, the reader is referred to Bock et al. (2004) and Dvořák and Savický (2007).[1]

The dependence between the MAGIC Gamma Telescope variables was analysed by Czado (2019) and by Nagler and Czado (2016), who pointed out the uncommon characteristics of the dependence structure between some of the variables, which do not correspond to any parametric copula families. In particular, the dependence between the variables `Length` (length of the major axis of the ellipse, in mm) and `M3Long` (third root of the third moment along the major axis, in mm) is rather peculiar, as confirmed by the empirical normalized contour plot depicted in the bottom left panel of Fig. 5.

Here we analyse how the dependence between `Length` and `M3Long`, measured by Spearman's ρ, varies with respect to the variables: `class` (which has two levels: gamma rays or background noise), `Width` (the length of the minor axis of the ellipse in mm) and `Size` (the 10-log of the sum of the content of all pixels in the image).

Fig. 6 shows the approximation of the posterior mean and credible intervals of the Spearman's ρ between `Length` and `M3Long` with respect to `Width` and `Size`, split by `class`, obtained with the GP and the spline methods. The blue lines show the results obtained with the GP method, while the coral lines show the results obtained with the splines method. The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. Red dots depict the frequentist estimates of the unconditional Spearman's ρ for observations belonging to the specific levels of the covariates. Similarly to the example in Section 4.1, splines tend to excessively smooth out the relationship between the dependence and

[1] The data are available at https://archive.ics.uci.edu/ml/datasets/magic+gamma+telescope.

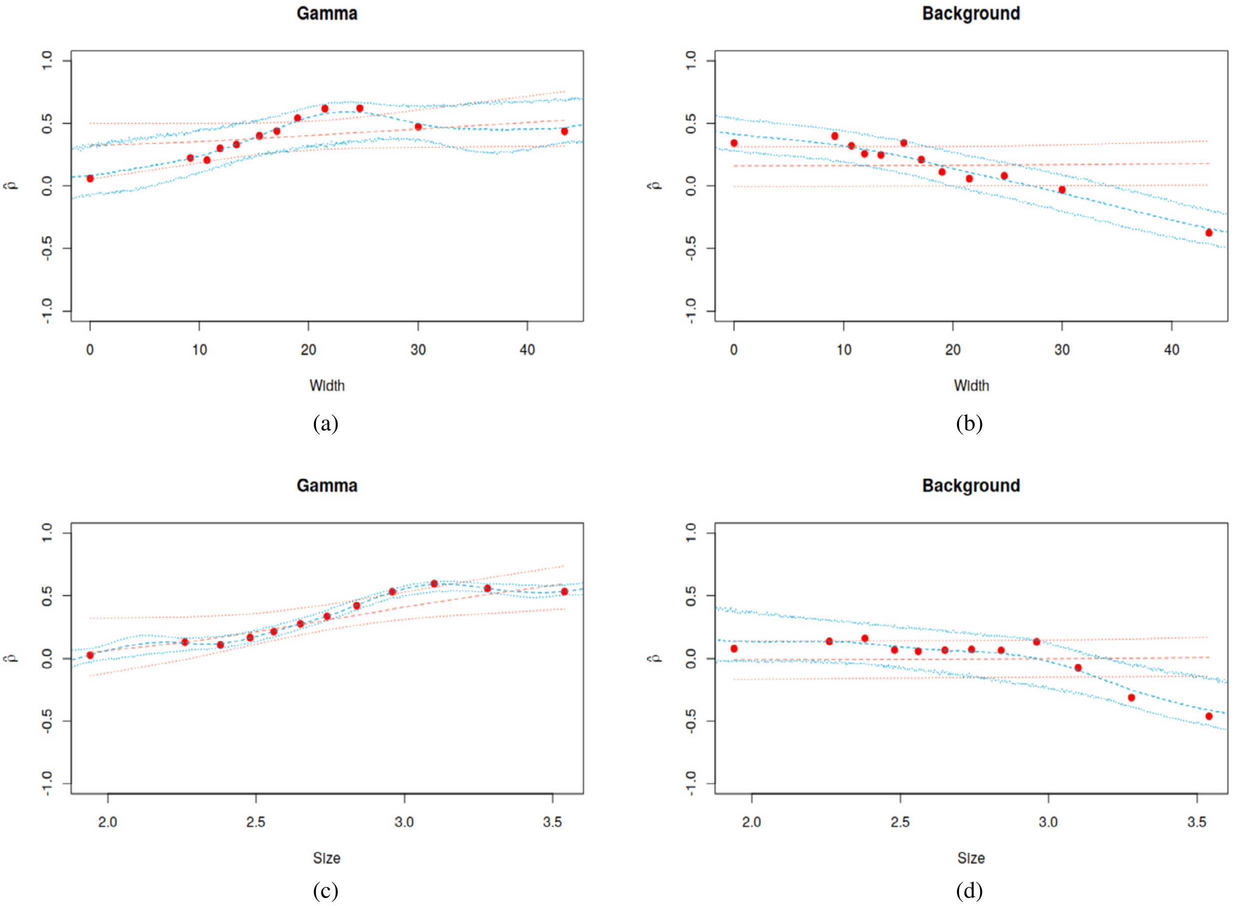


Fig. 6. Posterior approximations of Spearman's ρ between `Length` and `M3Long` as a function of `Width` (top plots) and `Size` (bottom plots), for the observations classified as gamma rays (left plots) and background noise (right plots), as defined by `class`. The blue lines show the results obtained with the GP method, while the coral lines show the results obtained with the splines method. The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. Red dots depict the frequentist estimates of the unconditional Spearman's ρ for observations belonging to the specific levels of the covariates.

the covariates, while the GPs better follow the data. Fig. 6 also shows that the dependence structure among the recorded images measurements shows different patterns for gamma rays and background noise and can be used for discriminating between these two data classes.

Fig. 7 shows the posterior approximations of Spearman's ρ between `Length` and `M3Long` with respect to `Width` and `Size`, split by `class`, calculated with the method based on the EL, with the inconsistent estimator of the copula using: NW weights with triweigth kernel (blue), NW weights with Gaussian kernel (orange), LL weights with triweight kernel (green), LL weights with Gaussian kernel (purple). The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. Red dots depict the frequentist estimates of the unconditional Spearman's ρ for observations belonging to the specific levels of the covariate. Similarly to the previous examples, the approximation strongly depends on the definition of the weights and the kernel functions within the weights, and the uncertainty associated with the estimates is larger than that obtained with methods based on splines or GPs. Moreover, the computational cost is large, where for some approximations most of the weights describing the EL associated with different values of the functional are zero or close to zero and it is not possible to obtain accurate approximations (this is the reason why some of the estimates are not shown in the plots). In general, non linear functions seem to be less well approximated than linear functions, especially if some areas of the covariate spaces are less represented.

While the uncertainty associated to the EL estimates is larger than the methods based on GP or Bayesian splines, EL is more robust: it does not require to discretize a continuous covariate to compute the frequentist estimates for each level of the covariate, which is necessary for GP and Bayesian splines. Since the bias associated with the unconditional frequentist estimates cannot be computed, GP and Bayesian splines my follow too closely these estimates, underestimating the estimation bias. Moreover, the frequentist estimates may depend on the choice of how the continuous covariate is discretized into groups.

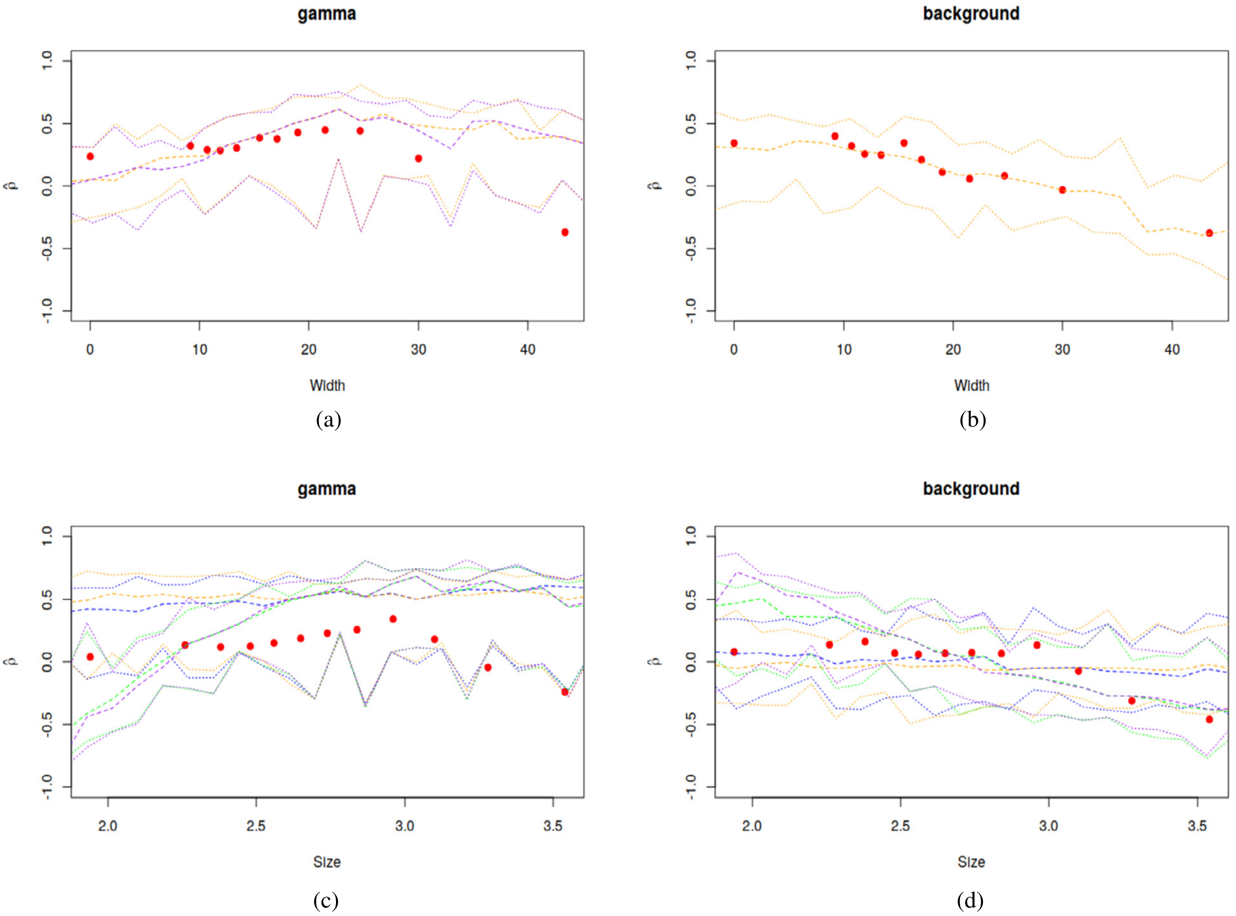


Fig. 7. Posterior approximations of Spearman's ρ between `Length` and `M3Long` as a function of `Width` (top plots) and `Size` (bottom plots), for the observations classified as gamma rays (left plots) and background noise (right plots), as defined by `class`. The lines show the results obtained with the methods based on the EL using: NW weights with triweigth kernel (blue), NW weights with Gaussian kernel (orange), LL weights with triweight kernel (green), LL weights with Gaussian kernel (purple). The inner dashed lines denote the posterior means and the dotted lines denote the 95% credible intervals. Red dots depict the frequentist estimates of the unconditional Spearman's ρ for observations belonging to the specific levels of the covariate.

5. Conclusions

In this work, we have analysed three main methodologies to approximate the posterior distribution of functionals of the dependence: Gaussian processes, methods based on the empirical likelihood, and methods based on Bayesian splines.

We have compared the methods in terms of approximation error and precision of the estimates.

The main advantage of all these methods is that they avoid the selection of the copula family. We have shown in practical examples that the selection of the copula is not an easy task. In particular, when the functional of the dependence is influenced by covariates, two main difficulties arise: the number of observations for each level of the covariates can be too limited to properly select the model and the structure of the dependence can in practice change with the level of the covariate.

Non-linear estimation procedures (like Gaussian processes and Bayesian splines) benefit from being flexible enough to adequately fit the relationship between the dependence and covariates, however they need several observations for each level of the covariates to define a noisy version of the functionals to be estimated. Such requirement can be limiting in applied contexts either because there could be only one observation for each level or because the covariate is continuous. In the latter case, groups of covariate values can be combined into discrete levels. In any case, these methods can be implemented with a limited number of covariates.

Methods based on the empirical likelihood, despite not needing replications for each covariate level, on the other hand show higher approximation errors. When using an inconsistent estimator of the copula, the approximation seems to strongly depend on the choice of the weights and the approximation error is larger than GPs- or spline-based methods. On the other hand, when using a linearised model of the functional through a Taylor's expansion, the uncertainty increases so that inference is not meaningful.

For this work, an R-package called BICC has been implemented containing code to perform the methodologies described in the article. The package is available at https://github.com/cgrazian/BICC.

References

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Abegaz, F., Gijbels, I., Veraverbeke, N., 2012. Semiparametric estimation of conditional copulas. J. Multivar. Anal. 110, 43–73.

Acar, E.F., Craiu, R.V., Yao, F., 2011. Dependence calibration in conditional copulas: a nonparametric approach. Biometrics 67 (2), 445–453.

Acar, E.F., Genest, C., Nešlehová, J., 2012. Beyond simplified pair-copula constructions. J. Multivar. Anal. 110, 74–90.

Acar, E.F., Craiu, R.V., Yao, F., et al., 2013. Statistical testing of covariate effects in conditional copula models. Electron. J. Stat. 7, 2822–2850.

Bock, R., Chilingarian, A., Gaug, M., Hakl, F., Hengstebeck, T., Jiřina, M., Klaschka, J., Kotrč, E., Savický, P., Towers, S., et al., 2004. Methods for multidimensional event classification: a case study using images from a cherenkov gamma-ray telescope. Nucl. Instrum. Methods Phys. Res., Sect. A, Accel. Spectrom. Detect. Assoc. Equip. 516 (2–3), 511–528.

Craiu, V.R., Sabeti, A., 2012. In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. J. Multivar. Anal. 110, 106–120.

Cressie, N., 1992. Statistics for Spatial Data. Wiley Online Library.

Czado, C., 2010. Pair-copula constructions of multivariate copulas. In: Copula Theory and Its Applications. Springer, pp. 93–109.

Czado, C., 2019. Analyzing Dependent Data with Vine Copulas. Springer.

Dalla Valle, L., Leisen, F., Rossini, L., 2018. Bayesian non-parametric conditional copula estimation of twin data. J. R. Stat. Soc., Ser. C, Appl. Stat. 67 (3), 523–548.

Dvořák, J., Savický, P., 2007. Softening splits in decision trees using simulated annealing. In: International Conference on Adaptive and Natural Computing Algorithms. Springer, pp. 721–729.

Geisser, S., Eddy, W.F., 1979. A predictive approach to model selection. J. Am. Stat. Assoc. 74 (365), 153–160.

Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82 (3), 543–552.

Gijbels, I., Veraverbeke, N., Omelka, M., 2011. Conditional copulas, association measures and their applications. Comput. Stat. Data Anal. 55 (5), 1919–1932.

Gijbels, I., Omelka, M., Veraverbeke, N., et al., 2012. Multivariate and functional covariates and conditional copulas. Electron. J. Stat. 6, 1273–1306.

Gijbels, I., Omelka, M., Veraverbeke, N., 2015. Estimation of a copula when a covariate affects only marginal distributions. Scand. J. Stat. 42 (4), 1109–1126.

Goegebeur, Y., Guillou, A., Le Ho, N.K., Qin, J., 2020. Robust nonparametric estimation of the conditional tail dependence coefficient. J. Multivar. Anal. 178, 104607.

Goldstein, M., Wooff, D., 2007. Bayes Linear Statistics: Theory and Methods, Vol. 716. John Wiley & Sons.

Grazian, C., Fan, Y., 2020. A review of approximate bayesian computation methods via density estimation: inference for simulator-models. Wiley Interdiscip. Rev.: Comput. Stat. 12 (4), e1486.

Grazian, C., Liseo, B., 2016. Approximate bayesian methods for multivariate and conditional copulae. In: International Conference on Soft Methods in Probability and Statistics. Springer, pp. 261–268.

Grazian, C., Liseo, B., 2017. Approximate bayesian inference in semiparametric copula models. Bayesian Anal. 12 (4), 991–1016.

Gutmann, M.U., Corander, J., 2016. Bayesian optimization for likelihood-free inference of simulator-based statistical models. J. Mach. Learn. Res. 17 (1), 4256–4302.

Haff, I.H., Aas, K., Frigessi, A., 2010. On the simplified pair-copula construction—simply useful or too simplistic? J. Multivar. Anal. 101 (5), 1296–1310.

Heck, D., Knapp, J., Capdevielle, J., Schatz, G., Thouw, T., et al., 1998. Corsika: a Monte Carlo code to simulate extensive air showers. Report FZKA 6019 (11).

Killiches, M., Kraus, D., Czado, C., 2017. Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. Aust. N. Z. J. Stat. 59 (1), 95–117.

Killiches, M., Kraus, D., Czado, C., 2016. Using model distances to investigate the simplifying assumption, model selection and truncation levels for vine copulas. arXiv preprint arXiv:1610.08795.

Klein, N., Kneib, T., 2016. Simultaneous inference in structured additive conditional copula regression models: a unifying bayesian approach. Stat. Comput. 26 (4), 841–860.

Kojadinovic, I., Yan, J., et al., 2010. Modeling multivariate distributions with continuous margins using the copula r package. J. Stat. Softw. 34 (9), 1–20.

Kurz, M.S., Spanhel, F., 2017. Testing the simplifying assumption in high-dimensional vine copulas. arXiv preprint arXiv:1706.02338.

Levi, E., Craiu, R.V., 2018. Bayesian inference for conditional copulas using gaussian process single index models. Comput. Stat. Data Anal. 122, 115–134.

Lopez-Paz, D., Hennig, P., Schölkopf, B., 2013. The randomized dependence coefficient. Adv. Neural Inf. Process. Syst. 26, 1–9.

Mengersen, K.L., Pudlo, P., Robert, C.P., 2013. Bayesian computation via empirical likelihood. Proc. Natl. Acad. Sci. 110 (4), 1321–1326.

Meyer, M.C., 2008. Inference using shape-restricted regression splines. Ann. Appl. Stat. 2 (3), 1013–1033.

Meyer, M.C., Hackstadt, A.J., Hoeting, J.A., 2011. Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. J. Nonparametr. Stat. 23 (4), 867–884.

Nagler, T., Czado, C., 2016. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. J. Multivar. Anal. 151, 69–89.

Nelsen, R.B., 2007. An Introduction to Copulas. Springer Science & Business Media.

Ourghi, R., Al-Anzi, A., Krarti, M., 2007. A simplified analysis method to predict the impact of shape on annual energy use for office buildings. Energy Convers. Manag. 48 (1), 300–305.

Owen, A.B., 2001. Empirical Likelihood. Chapman and Hall/CRC.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. Int. Econ. Rev. 47 (2), 527–556.

Paulo, R., 2005. Default priors for gaussian processes. Ann. Stat. 33 (2), 556–582.

Pessenlehner, W., Mahdavi, A., 2003. Building morphology, transparence, and energy performance. In: Proceedings of the Eighth International IBPSA Conference.

Price, L.F., Drovandi, C.C., Lee, A., Nott, D.J., 2018. Bayesian synthetic likelihood. J. Comput. Graph. Stat. 27 (1), 1–11.

Ramsay, J.O., et al., 1988. Monotone regression splines in action. Stat. Sci. 3 (4), 425–441.

Roberts, A., Marsh, A., 2001. ECOTECT: environmental prediction in architectural education. In: 19th eCAADe Conference Proceedings.

Schennach, S.M., 2005. Bayesian exponentially tilted empirical likelihood. Biometrika 92 (1), 31–46.

Schmid, F., Schmidt, R., 2007. Multivariate extensions of spearman's rho and related statistics. Stat. Probab. Lett. 77 (4), 407–416.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231.

Stander, J., Dalla Valle, L., Taglioni, C., Liseo, B., Wade, A., Cortina-Borja, M., 2019. Analysis of paediatric visual acuity using bayesian copula models with sinh-arcsinh marginal densities. Stat. Med. 38 (18), 3421–3443.

Stone, C.J., Huang, J.Z., 2002. Free knot splines in concave extended linear modeling. J. Stat. Plan. Inference 108 (1–2), 219–253.
Tsanas, A., Xifara, A., 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build. 49, 560–567.
Vatter, T., Chavez-Demoulin, V., 2015. Generalized additive models for conditional dependence structures. J. Multivar. Anal. 141, 147–167.
Vehtari, A., Ojanen, J., 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. Stat. Surv. 6, 142–228.
Veraverbeke, N., Omelka, M., Gijbels, I., 2011. Estimation of a conditional copula and association measures. Scand. J. Stat. 38 (4), 766–780.
Watanabe, S., 2013. A widely applicable Bayesian information criterion. J. Mach. Learn. Res. 14, 867–897.
Wu, J., Wang, X., Walker, S.G., 2015. Bayesian nonparametric estimation of a copula. J. Stat. Comput. Simul. 85 (1), 103–116.