Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 68/ 2020

Analysing the relationship between district heating demand and weather conditions through conditional mixture copula

F. Marta L. Di Lascio, Andrea Menapace,
Maurizio Righetti

Noname manuscript No.
(will be inserted by the editor)

Analysing the relationship between district heating demand and weather conditions through conditional mixture copula

F. Marta L. Di Lascio · Andrea Menapace · Maurizio Righetti

Received: date / Accepted: date

Abstract Efficient energy production and distribution systems are urgently needed to reduce world climate change. Since modern district heating systems are sustainable energy distribution services that exploit renewable sources and avoid energy waste, in-depth knowledge of thermal energy demand, which is mainly affected by weather conditions, is essential to enhance heat production schedules. We hence propose a mixture copula-based approach to investigate the complex relationship between meteorological variables, such as outdoor temperature and solar radiation, and thermal energy demand in the district heating system of the Italian city Bozen-Bolzano. We analyse data collected from 2014 to 2017, and estimate copulas after removing serial dependence in each time series using autoregressive integrated moving average models. Due to complex relationships, a mixture of an unstructured Student-t and a flipped Clayton copula is deemed the best model, as it allows differentiating the magnitude of dependence in each tail and exhibiting both heavy-tailed and asymmetric dependence. We derive the conditional copula-based probability function of thermal energy demand given meteorological variables, and provide useful insight on the production management phase of local energy utilities.

Keywords Conditional probability; Copula function; District heating; Mixture copula; Flipped copula; Thermal energy demand
JEL codes: C10, C32, P28

F. Marta L. Di Lascio
Faculty of Economics and Management, University of Bozen-Bolzano, Italy
Tel.: +39-0471-013285
Fax: +39-0471-013009
E-mail: marta.dilascio@unibz.it

Andrea Menapace
Faculty of Science and Technology, University of Bozen-Bolzano, Italy

Maurizio Righetti
Faculty of Science and Technology, University of Bozen-Bolzano, Italy

1 Introduction

Efficient energy production and distribution systems are urgently needed as crucial elements of the climate change response in Europe and many other regions (Mathiesen et al., 2011; Eriksson et al., 2007; Lund et al., 1999). By 2020, the European Commission's strategy (European Commission, 2011) for competitive, sustainable, and secure "Energy 2020" aims to "reduce its greenhouse gas emissions by at least 20%, increase the share of renewable energy to at least 20% of consumption, and achieve energy savings of 20% or more". To achieve these targets, the EU is making an effort to identify strategies to decrease recourse to fossil fuels, limit environmental pollution, and reduce climate change. Amongst others, the implementation of the Strategic Energy Technology Plan (European Commission, 2015) should accelerate the development of low-carbon energy systems. These systems aim to exploit the full potential of renewable energy sources, such as geothermal and solar, to satisfy heating and cooling demand.

Lund (2014) describes the technical characteristics of sustainable energy systems and identifies the challenges for developing a system entirely based on a renewable non-fossil heat supply. Smart networks for energy transition, such as district heating and cooling, should be designed to distribute heat through networks with low grid losses to existing buildings, energy-renovated buildings, and new low-energy buildings (Lund et al., 2014). Specifically, district heating systems (DHSs hereafter) are thermal energy distribution services that allow matching local energy sources with urban heat requirements through pipeline networks and heat exchangers (Frederiksen and Werner, 2013). The recent development of these systems is typically based on a combination of fluctuating renewable energy sources (e.g., wind, geothermal, and solar power), with residual resources (e.g., waste and biomass). The intermittent and unpredictable nature of renewable sources entails designing DHSs in such a way that thermal energy demand can be satisfied even when renewable sources are limited (Lund et al., 2018). Hence, an efficient energy production and distribution system requires in-depth knowledge of thermal energy demand.

As well-known, thermal energy demand depends on meteorological variables, including outdoor temperature, wind speed, solar radiation, humidity, and precipitation, which can have a strong impact on demand (Soutullo et al., 2016). Variations in weather conditions affect energy consumption whose management is often enormously challenging. Moreover, increasingly likely extreme climatic events are a problem for the energy production process, which must be able to adequately deal with unexpected energy demand. Hence, efficiently planning and managing a heating system, especially in the case of DHS, urgently requires in-depth knowledge of the complex relationship between thermal energy demand and meteorological variables. Knowledge of thermal energy demand and the complex relationship with the weather is also crucial to obtain reliable forecasts of such demand. These challenging needs induce resorting to copula theory.

Copula (Sklar, 1959) is a distribution function that potentially enables describing any kind of complex multivariate dependence structure of the data generating process without information on margins. Copulas have shown their considerable usefulness due to their theoretical and practical advantages. Compared with standard approaches, such as multivariate normal distribution, copula models, or copulas, (Trivedi and Zimmer, 2005; Durante and Sempi, 2015) are indeed a powerful tool to describe the dependence structure in a set of random variables that exhibit non-linear and non-Gaussian relations, heavy tails, and asymmetries. Copulas have been successfully used in several multivariate statistics fields, e.g., time series analysis (Chen and Fan, 2006; Zhao and Zhang, 2018) and clustering (Di Lascio and Giannerini, 2016; Prenen et al., 2017), and applied in a wide range of domains, e.g., environmental sciences (Belzunce et al., 2016; Baey et al., 2017; Grler et al., 2017), genetics and medicine (Shi and Zhang, 2015; Behrouzi and Wit, 2018), and engineering (Genest and Favre, 2007; Zhu and Zhao, 2018). Moreover, to further enhance the flexibility of multivariate distribution modelling and generate dependence structures not belonging to existing copula families, a finite mixture of heterogeneous parametric copulas can be used. A number of studies adopt the finite mixture of parametric copula densities modelling approach, see, e.g., Hu (2006); Zimmer (2012); Liu et al. (2019); Qu and Lu (2019).

In this study, we aim to investigate the complex relationship between thermal energy demand and meteorological variables and quantify the effect of extreme climatic events on district heating demand. To provide evidence of the complex relationship between meteorological variables and thermal energy demand, we adopt an approach based on copulas and combinations of copulas, looking at the conditional distribution function of thermal energy demand given certain weather conditions. This function enables obtaining relevant information mainly on the production schedule and the management of thermal storage. The copula function has several advantages in this context. First, the copula model ensures flexibility in jointly modelling variables with a different and complex univariate density function. Second, the copula model easily accommodates the asymmetry and heavy tails of thermal energy demand and meteorological variables. Third, the effect of weather can be identified through a functional form by specifying the conditional copula of thermal energy demand. Fourth, copula mixtures allow differentiating the contribution to the dependence in each tail, capturing diverse patterns of dependence in the data, and hence to offer a less restrictive parametric modeling. Moreover, to the best of our knowledge, a copula mixture-based approach has never been applied to energy and meteorological variables in a multivariate perspective. The only study based on copulas and analysing the impact of temperature on energy demand is that of Di Lascio et al. (2019) using a bivariate copula with symmetric dependencies and focusing only on peak thermal power demand. Conversely, we here *i*) investigate the joint impact of a multivariate set of meteorological variables on daily thermal energy demand, *ii*) attempt to shed new light on how extreme climatic phenomena affect energy demand by also exploiting the advantages of copula mixtures, and *iii*) discuss the implications of the selected copula model on energy production.

In the empirical analysis, we apply copula theory to a dataset from the Italian city Bozen-Bolzano related to the heating season in a 3-year period (from 15-10 to 15-04 for each year). The dataset contains meteorological data, e.g., solar radiation and outdoor temperature, collected by a weather station, and thermal energy demand provided by intelligent substations equipped with smart meter devices. We find that the unstructured Student-t–flipped Clayton copula mixture, which exhibits heavy-tailed and asymmetric dependence, best suits our data. Moreover, the conditional version of the selected copula enables investigating the probability of exceeding energy production under extreme weather conditions. Our findings are expected to improve our understanding of thermal energy demand according to real and simulated weather scenarios to aid the development of appropriate production management strategies.

The remainder of the paper is organized as follows. Section 2 describes a typical DHS and its role in a sustainable energy system. Section 3 illustrates the collected data concerning the DHS of the Italian city Bozen-Bolzano. Section 4 presents the statistical methodology used and applied in Section 5. Section 6 discusses the findings and the managerial implications. Section 7 presents concluding remarks.

2 District heating system and sustainable energy

District heating systems are thermal energy distribution services that provide heat through pipeline networks and heat exchanger substations. The first DHS was introduced in the United States at the end of $19th$ century using steam as the heat carrier (Woods and Overgaard, 2016) with the main aim of increasing security and comfort with respect to individual boilers. From the $1930s$ to $1970s$, the second generation of DHSs was developed, usually using pressurised hot water over $100°C$ to distribute heat to achieve fuel savings, and introducing combined heat and power (CHP hereafter). The two oil crises of 1973 and 1979 resulted in upgrading and deploying DHSs aimed at increasing energy efficiency and introducing alternative non-fossil fuel-based sources, such as biomass and waste. The third generation of DHSs use low temperature hot water (below $100°C$) as the heat carrier, and better performing materials, such as pre-insulated pipes, and compact heat exchanger substations (Lund et al., 2014; Woods and Overgaard, 2016; Werner, 2017).

Recently, a new concept of DHS based on a sustainable energy system consisting of 100% renewable sources has been introduced (Connolly et al., 2012, 2013). The so-called 4th DHS generation (Lund et al., 2014) has to support energy transition from a traditional fossil system to a sustainable energy system in which renewable resources are preponderant, and particular attention is paid to energy conservation and efficiency measures (Lund, 2007; Lund and Mathiesen, 2009; Mathiesen et al., 2014). Thus, DHSs are based on a pool of heat sources: fluctuating renewable sources, such as geothermal and solar power, residual resources, such as biomass and waste, energy recovered from waste and industrial surplus, and energy from efficient technologies such as CHP and heat pumps (Frederiksen and Werner, 2013).

Due to the urgent need to increase the sustainability and efficiency of energy systems and reduce their environmental impact, DHSs assume an important role in urban areas for several reasons. First, central management of the heat supply replaces less efficient domestic boilers of single or groups of apartments. Second, DHSs enable using scarce local heat sources and recovering heat with the reduced use of higher quality fuels, such electricity power and gas. Third, DHSs allow managing fluctuating renewable sources using thermal storage with heat capacity compensation from the daily to the seasonal. Fourth, DHSs enhance the resilience and reliability of the supply system, as well as decreasing maintenance costs. Fifth, the smart energy system of the future entails coordinating the gas, thermal, electricity, and water grid for the optimal management of the entire energy system (Lund et al., 2018; Werner, 2017; Ilic et al., 2014; Lund et al., 2014). Nevertheless, DHSs also have some weaknesses, requiring an initial high capital investment and a complex institutional framework.

Today, DHSs have a significant role in heat supply with 11.5 EJ of heat supplied worldwide in 2014 (IEA International Energy Agency, 2016). However, the exploitation of DHSs varies considerably from country to country, with the most developed systems in Denmark and Iceland where the majority of residential buildings located in urban areas are supplied by DHSs. While DHSs are deemed important in supporting future energy and environmental targets, such as the reduction of carbon dioxide emissions, the potentialities of DHSs need further identification and implementation (Werner, 2017).

3 Dataset of the district heating system of Bozen-Bolzano

Residential and industrial heat demand in the city of Bozen-Bolzano is partially fulfilled by a DHS that is in continuous expansion. Currently, the DHS feeds around 3.500 apartments and 100 industrial-commercial buildings through a pipe network of 23 Km. The DHS is characterised by three heat production facilities consisting of a waste-to-energy plant (WtE hereafter), two combined heat and power (CHP hereafter) engines, and six traditional natural gas-fed boilers with a total thermal capacity of 79.2 MW. In October 2016, a new thermal storage tank was added to improve the production management of the DHS.

The dataset gathered includes both the thermal energy demand of consumers connected to the DHS of Bozen-Bolzano and the meteorological data of the weather station of S. Maurizio observed from 2014 to 2017. We select data of the heating season in Bozen-Bolzano that starts on October 15th and ends on April 15th according to DPR Decreto del Presidente della Repubblica n. 74 (16 aprile 2013), Art. 4. The DHS of Bozen-Bolzano is characterized by an intelligent energy network endowed with smart heat meters, that provide real-time information and high quality monitoring (Sun et al., 2016). These innovative devices, integrated in each substation, are fundamental in this study, as they yield high quality and high frequency data. Hence, as for thermal energy demand, each substation of the DHS provides observations of 15 minutes each. To avoid truncated time series, we select the 110 substations that provide time series with no null interval of data. Next, we apply a filter to detect outliers, i.e., values that exceed the maximum exchangeable power of each substation. These values are treated as missing values due to measurement errors of the substation heat meters. We obtain the daily thermal energy demand (TED hereafter) by aggregating the observations of each day over the substations. The filtering procedure leads to 36 missing data, that is, 6.55% of the total number (550). Hence, we impute the thermal energy demand time series using a seasonal Kalman filter (Moritz and Bartz-Beielstein, 2017; Moritz et al., 2015). Fig. 1 presents the TED time series with imputed data. Note that, forthwith, we indicate with TED the imputed time series of thermal energy demand.

As for the meteorological variables, we observe the outdoor temperature (OT hereafter), solar radiation (SR hereafter), humidity, wind speed, and precipitation detected every 5 or 10 minutes by the weather station. We average the observations of each day to obtain the daily value of each meteorological variable.

First, we look at the relationship between thermal energy demand and all the observed meteorological variables. The estimated Spearman's correlation coefficient $\hat{\rho}_s$ and associated p-value are shown in Tab. 1. These results together with

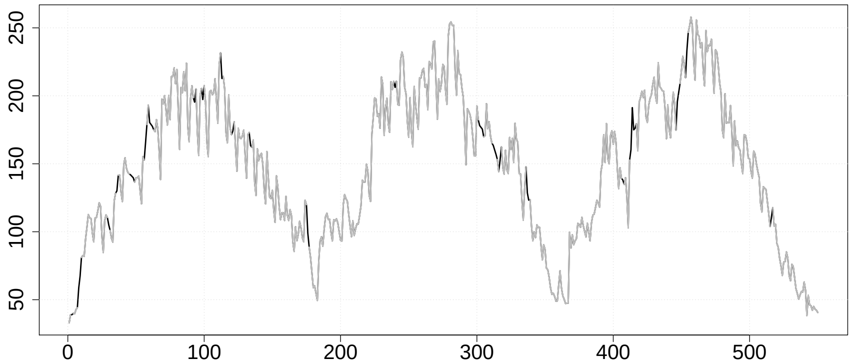

Fig. 1 TED (in MWh) time series: black segments for imputed values, grey segments for observed values.

Table 1 Estimated Spearman's correlation coefficient $\hat{\rho}_s$ between TED and each of the observed meteorological variables, and the p-value of the corresponding test.

	Outdoor temperature	Solar radiation	Humidity	Wind speed	Precipitation
$\hat{\rho}_s$	-0.926	-0.523	0.077	-0.273	-0.087
p-value	< 0.001	< 0.001	0.07	< 0.001	0.040

the scatter plot of TED versus humidity, wind speed, and precipitation shown in Fig. 2 lead us to consider in the subsequent analysis only the OT and the SR variables. The preliminary correlation analysis indicates strong negative and significant dependence between TED and OT, and moderate negative significant dependence between TED and SR (see Tab. 1). On the contrary, the Spearman's correlation coefficient estimated on OT and SR is $\hat{\rho}_s = 0.52$ (p-value < 0.001), showing significant and positive dependence between the two meteorological variables. Fig. 3 shows the three-dimensional scatter plot of the selected time series on the left, and the pair scatter plot on the right. These plots highlight a non-linear and asymmetric dependence between the variables of interest.

4 Methodology

The statistical methodology using copula function theory enables investigating the dependence structure between daily thermal energy demand and the meteorological variables. Since we have serial dependent data, we first model each time series separately through the seasonal autoregressive integrated moving average (SARIMA hereafter) models according to the well-known Box-Jenkins procedure (Box and Jenkins, 1970). Second, we obtain the residual time series and analyse the multivariate cross-dependence relationship through the copula. Finally, we derive the conditional probability function of thermal energy demand given the outdoor temperature and solar radiation, analysing it according to extreme climatic scenarios. We assume that the functional form of the copula remains fixed over time, and the sample and dependence parameter is not allowed to vary according to an evolution equation. This choice is motivated by the fact that we

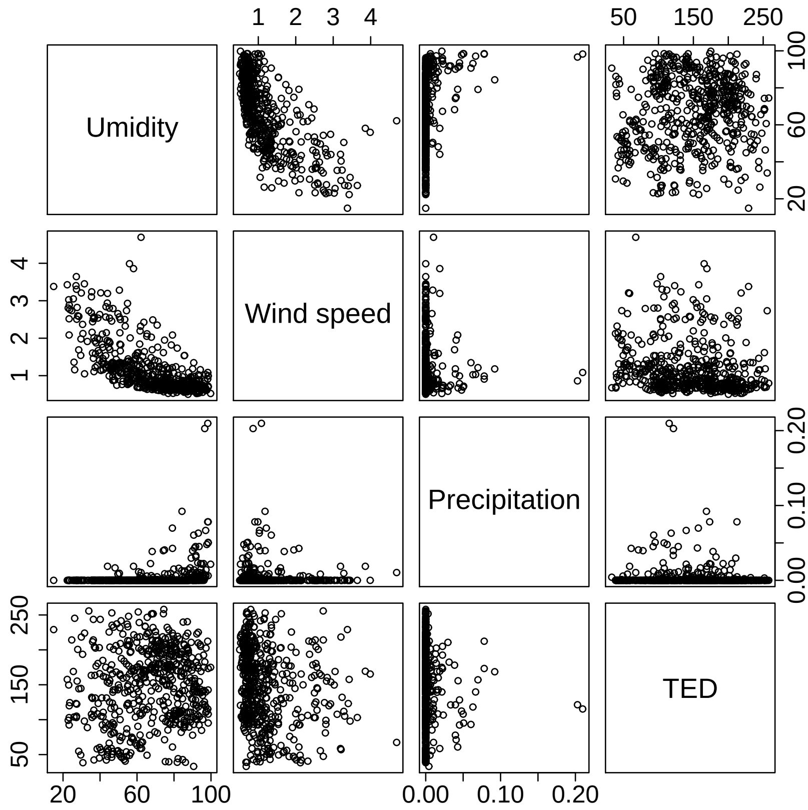

Fig. 2 Scatter plot of TED in MWh, humidity in %, wind speed in m/s, and precipitation in mm.

focus the analysis only on daily data of thermal energy demand in the heating season whose dependence is not expected to change over time.

4.1 Marginal modelling using ARIMA

To analyse the serial dependence of outdoor temperature, solar radiation, and thermal energy demand, we use the well-known SARIMA models that enable capturing both the trend (seasonal and non) and autocorrelation. The general SARIMA$(p, d, q)(P, D, Q)_s$ model for a generic univariate stochastic process Z_t is as follows:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{1}$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ is the classic white noise process with variance σ_ε^2, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ ($\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps}$) is the autoregressive (seasonal autoregressive) polynomial in B of grade p (P), $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ ($\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$) is the

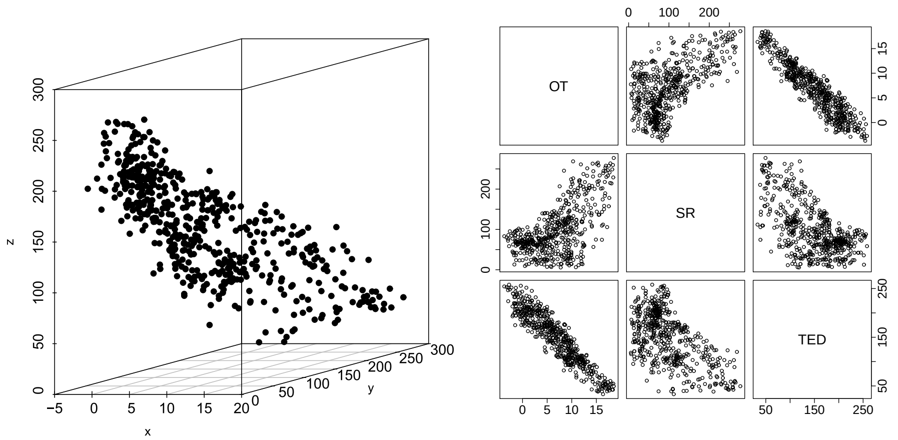

Fig. 3 Three-dimensional scatter plot of OT in $°C$ (x-axis), SR in W/m^2 (y-axis), and TED in MWh (z-axis) (left), and pair scatter plot of OT, SR, and TED (right).

moving average (seasonal moving average) polynomial in B of grade q (Q), and $\nabla^d = (1-B)^d$ ($\nabla_s^D = (1-B^s)^D$) is the difference (seasonal difference) operator of order d (D). Hence, the identification of the values p, P, q, Q, d, and D enables developing a specific model and proceed with its estimation. We analyse the three time series separately according to the Box-Jenkins procedure (Box and Jenkins, 1970), which due to its simplicity and generality is the most frequently used. Being a classic approach, we do not go into details of the procedure and recall only its main steps: 1. analysis of the stationarity of the time series and selection of the potential difference operator and transformation to obtain stationarity, 2. identification of the SARIMA model based on the autocorrelation and partial autocorrelation function of the stationary time series, 3. check for the goodness of the selected SARIMA model, mainly through the t-test of the estimated coefficients and the Ljung-Box test for the autocorrelation of residuals.

4.2 Joint and conditional dependence modelling using copulas

To investigate the multivariate and complex relationship between outdoor temperature, solar radiation, and thermal energy demand, we use copula theory. The concept of 'copula' or 'copula function' originated in the context of probabilistic metric spaces through the well-known Sklar's theorem (Sklar, 1959) whose probabilistic interpretation states that every joint distribution function $F(\cdot)$ can be expressed in terms of i) p marginal distribution functions $F_j(X_j)$, with X_j continuous random variable and $j = 1, \dots, p$, and ii) the copula distribution function $C(\cdot) : [0,1]^p \to [0,1]$ as follows:

$$F(X_1, \dots, X_j, \dots, X_p) = C(F_1(X_1), \dots, F_j(X_j), \dots, F_p(X_p)). \quad (2)$$

Hence, copula is a p-dimensional cumulative distribution function (CDF hereafter) with standard uniform margins $F_j(\cdot) \sim U(0,1)$ (Trivedi and Zimmer, 2005; Nelsen, 2006; Durante and Sempi, 2015, for details). The great theoretical and practical potentiality of the copula is evident when the joint density function is derived as follows:

$$f(X_1, \ldots, X_j, \ldots, X_p) = c(F_1(X_1), \ldots, F_j(X_j), \ldots, F_p(X_p)) \prod_{j=1}^{p} f_j(X_j) \tag{3}$$

where $c(U_1, \ldots, U_j, \ldots, U_p) = \frac{\partial^p C(U_1,\ldots,U_j,\ldots,U_p)}{\partial U_1 \ldots \partial U_j \ldots \partial U_p}$ is the copula density associated with a copula $C(U_1, \ldots, U_j, \ldots, U_p)$ and $f_j(\cdot)$ with $j = 1, \ldots, p$ are the univariate margins. According to Sklar's theorem, any joint probability function $f(\cdot)$ can be split into two parts that are independent of each other: *i*) the copula $c(\cdot)$, which expresses the association among variables, e.g., the multivariate dependence structure of a joint density function, and *ii*) the univariate marginal densities $f_j(\cdot)$, with $j = 1, \ldots, p$. Such separation affects the log-likelihood function of $f(\cdot)$, which is composed of two positive terms as follows:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log c\left(F_1\left(x_{1i}\right), \ldots, F_j(x_{ji}), \ldots, F_p\left(x_{pi}\right); \boldsymbol{\theta}\right) + \sum_{i=1}^{n}\sum_{j=1}^{p} \log f_j\left(x_{ji}\right) \tag{4}$$

for all real vectors $(x_{j1}, \ldots, x_{ji}, \ldots, x_{jn})$ with $j = 1, \ldots, p$. In Eq. (4) the first term includes the copula density and its parameters $\boldsymbol{\theta}$, and the second the marginal densities and their parameters and it is valid. Hence, it is possible to decompose the estimation problem into two steps and combine different estimation methods, gaining high flexibility in the modelling. To estimate the copula function, we use the two-step sequential maximum likelihood estimation method (Joe and Xu, 1996) in its semi-parametric version (Genest et al., 1995). Thus, the marginal parameters are estimated in the first step and then used to estimate the dependence parameter of the copula function in the second step. Here we model margins through the empirical CDF $\hat{F}_j\left(x_{ji}\right)$ computed from $x_{j1}, \ldots, x_{jn}$, with $j = 1, \ldots, p$ and multiplied by $n/(n+1)$ to avoid problems at the boundary of $[0,1]^p$. Then, we estimate the copula parameter through the pseudo-likelihood approach as follows:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log c\left\{u_{1i}, \ldots, u_{ji}, \ldots, u_{pi}; \boldsymbol{\theta}\right\} \tag{5}$$

where $\hat{u}_{ji} = n\hat{F}_j(x_{ji})/(n+1)$ and is equivalent to $\hat{u}_{ji} = r_{ji}/(n+1)$ where r_{ji} is the rank of x_{ji} among $x_{j1}, \ldots, x_{jn}$, with $j = 1, \ldots, p$.

As for the kind of copula model, we focus on the one hand on copulas with a single parameter such as the Clayton and Gumbel copulas that capture asymmetric (respectively, left or right tail) dependence, and the Joe, Gaussian, and Student-t copulas that capture symmetric dependence with different tail heaviness. On the other hand, we use the Gaussian and Student-t copulas with unstructured correlation matrix to differently model the strength of each bivariate relationship. We also take into consideration the flipped versions of the Clayton, Gumbel, and Joe copulas (see Brechmann and Schepsmeier (2013)) to correctly model the relationship among the phenomena under investigation. Specifically, in accordance with the analysed data, we work with the following three-dimensional flipped copulas obtained by replacing some of the original random variables with a countermonotonic counterpart:

$$C_{001}(U_1, U_2, 1-U_3) = C(U_1, U_2) - C(U_1, U_2, 1-U_3) \tag{6}$$
$$C_{110}(1-U_1, 1-U_2, U_3) = U_3 - C(1-U_1, 1-U_2, U_3) \tag{7}$$

where $C_{000}(\cdot) = C(\cdot)$ is the copula given in Eq. (2) with $p = 3$ and the zero (one) subscript indicates a non-flipped (flipped) axis.

Moreover, given the difficulty of capturing diverse and more complex patterns of dependence structures through a single copula model, copula mixtures can be used. A mixture copula $C_{\text{mix}}(\cdot)$ is a finite linear combination of several copulas (Vrac et al., 2012) that can be written as follows:

$$C_{\text{mix}}(\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k C_k(\mathbf{U}; \boldsymbol{\theta}_k) \tag{8}$$

where $\{C_1(\cdot), \ldots, C_k(\cdot), \ldots, C_K(\cdot)\}$ is the set of candidate copulas with a vector of unknown dependence parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k, \ldots, \boldsymbol{\theta}_K)$ and the p-dimensional marginal distribution $\mathbf{U} = (U_1, \ldots, U_j, \ldots, U_p)$, and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k, \ldots, \pi_K)$ is the vector of weights, called mixing coefficients, such that $0 \leq \pi_k \leq 1, \forall k$ with $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_k = 1$. In Eq. (8), both the copula parameters $\boldsymbol{\theta}$ and the weights $\boldsymbol{\pi}$ control the shape of the mixture copula's dependence structure. As for the estimation of the mixture copula model in Eq. (8), following Hu (2006), we use the two-step semi-parametric estimation method described above into the EM algorithm (Dempster et al., 1977), which makes it possible to maximize the log-likelihood of the copula in Eq. (8). Specifically, the EM algorithm is an iterative procedure and each iteration is guaranteed to increase the log-likelihood and to converge to a local maximum. For our mixture copula model, the EM algorithm can be summarized as follows:

- initialize $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ by setting the $\hat{\boldsymbol{\pi}}^{\mathbf{0}} = (\hat{\pi}_1^0, \ldots, \hat{\pi}_k^0, \ldots, \hat{\pi}_K^0)$ and $\hat{\boldsymbol{\theta}}^{\mathbf{0}} = (\hat{\boldsymbol{\theta}}_1^{\mathbf{0}}, \ldots, \hat{\boldsymbol{\theta}}_k^{\mathbf{0}}, \ldots, \hat{\boldsymbol{\theta}}_K^{\mathbf{0}})$ values respectively, and evaluate the initial value of the log-likelihood in Eq. (4) as written for the copula in Eq. (8);
- E-step: evaluate the mixing coefficients, i.e., the posterior probabilities, using the current parameter values:

$$\hat{\pi}_k^{\text{new}} = \sum_{i=1}^{n} \frac{\hat{\pi}_k^0 c_k \left\{u_{1i}, \ldots, u_{ji}, \ldots, u_{pi}; \hat{\boldsymbol{\theta}}_k^0\right\}}{\sum_{k=1}^{K} \hat{\pi}_k^0 c_k \left\{u_{1i}, \ldots, u_{ji}, \ldots, u_{pi}; \hat{\boldsymbol{\theta}}_k^0\right\}}$$

  by varying k in $\{1, \ldots, K\}$;
- M-step: re-estimate the copula dependence parameters using the current mixing coefficients

$$\hat{\boldsymbol{\theta}}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log \sum_{k=1}^{K} \hat{\pi}_k^{\text{new}} c_k \left\{u_{1i}, \ldots, u_{ji}, \ldots, u_{pi}; \boldsymbol{\theta}_k\right\}$$

  where $\hat{\boldsymbol{\theta}}^{\text{new}} = (\hat{\boldsymbol{\theta}}_1^{\text{new}}, \ldots, \hat{\boldsymbol{\theta}}_k^{\text{new}}, \ldots, \hat{\boldsymbol{\theta}}_K^{\text{new}})$;

– evaluate the log-likelihood in Eq. (4) for the mixture model in Eq. (8) and check for convergence of the log-likelihood. If the convergence criterion is not satisfied, i.e., the change of the maximized log-likelihood of the mixture copula in the last iteration is greater than $1e-08$, return to the E-step and store $\hat{\boldsymbol{\theta}}^{\text{new}} \to \hat{\boldsymbol{\theta}}^{0}$ and $\hat{\pi}_k^{\text{new}} \to \hat{\pi}_k^0, \forall k = 1, \ldots, K$.

The components of a mixture copula model can potentially vary among any kind of copula model, providing huge flexibility and enabling defining new multivariate copula models. For a review of copula models see Nelsen (2006); Zimmer and Trivedi (2006); Durante and Sempi (2015).

According to the copula model, the value of the dependence parameter has a specific meaning. However, for all the copulas considered here, there is an analytical relationship between the copula dependence parameter and Kendall's rank correlation coefficient τ, which is a scale-invariant measure of association and useful to obtain a normalized "degree of association" between two continuous random variables with a specified copula (Durante and Sempi, 2015). Hence, the greater the value of Kendall's τ, the stronger the association among the margins. To specify and select the 'best' copula model, following Patton (2012), we perform the leave-one-out cross-validation method (CV hereafter) for copulas (Grønneberg and Hjort, 2014), and then employ the well-known Akaike information criterion (AIC hereafter) as well as the logarithm of the maximized likelihood function of the estimated copula model (LL hereafter).

Since we are interested in investigating the complex relationship between thermal energy demand and meteorological variables with the final aim of investigating demand given weather scenarios, we introduce the following conditional distribution function:

$$
\begin{aligned}
&P(X_j > x_j | X_1 < x_1, \ldots, X_{j-1} < x_{j-1}, X_{j+1} < x_{j+1}, \ldots, X_p < x_p) = \\
&= 1 - \frac{C(F_1(x_1), \ldots, F_j(x_j) \ldots, F_p(x_p))}{C(F_1(x_1), \ldots, F_{j-1}(x_{j-1}), F_{j+1}(x_{j+1}) \ldots, F_p(x_p))} \\
&= 1 - C(F_j(x_j) | F(x_1), \ldots, F(x_{j-1}), F(x_{j+1}), \ldots, F(x_p))
\end{aligned}
\tag{9}
$$

where $(x_1, \ldots, x_j, \ldots, x_p)$ is a vector of values in the domain of the variables $(X_1, \ldots, X_j, \ldots, X_p)$, $C(F_j(x_j)|F(x_1), \ldots, F(x_{j-1}), F(x_{j+1}), \ldots, F(x_p))$ is the value of the conditional copula defined using Bayes' rule (see Zimmer and Trivedi (2006)), and the copula $C(\cdot)$ can be any kind of copula model, e.g. flipped copula, mixture copula etc.

5 Empirical analysis

Here we study the pre-processed dataset described in Section 3 using the methodology explained in Section 4. Hence, we model thermal energy demand, outdoor temperature, and solar radiation separately using the SARIMA model. We then analyse the dependence among the residual of the marginal models through the three-dimensional copula families. Given the meteorological variables, we investigate TED through a conditional copula function.

5.1 Marginal modelling

Since the aim of this section is to obtain uncorrelated residual time series, we separately model the three time series using the Box-Jenkins procedure. First, we check the non-stationarity of the time series with the autocorrelation and partial autocorrelation functions at lag $1, \cdots, 60$. The middle column of Fig. 4 shows the autocorrelation functions, suggesting the non-stationarity in mean of the three series. In particular, the OT and SR show a negative linear trend (upper and

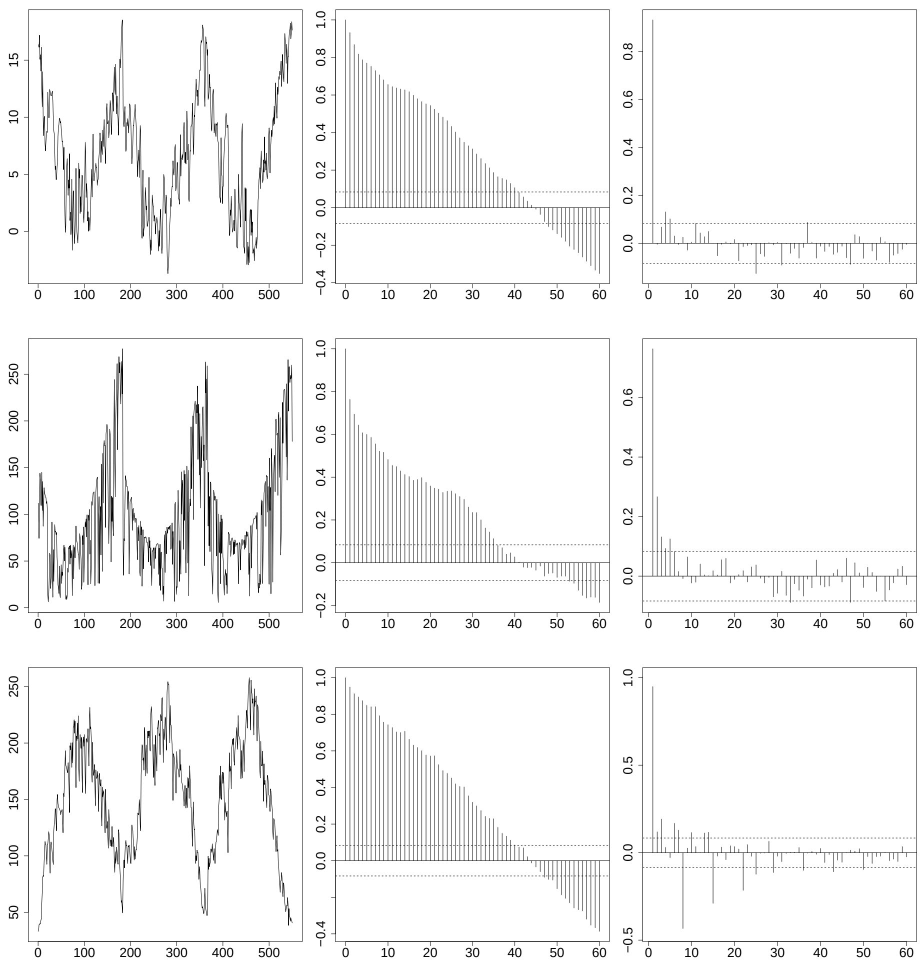

Fig. 4 Plot of time series, autocorrelation function and partial autocorrelation function of OT in $^{\circ}C$ (upper panel), SR in W/m^2 (middle panel) and TED in MWh (lower panel).

middle panel of Fig. 4), whereas TED presents a negative linear trend and a seasonal weekly trend (lower panel of Fig. 4). To remove non-stationarity, we apply the difference operator with $d = 1$ to OT and SR, and both the difference

Table 2 Estimated Spearman's correlation coefficient $\hat{\rho}_s$ between each pair of the OT, SR, and TED residual time series, and the p-value of the corresponding test.

	OT and TED	SR and TED	OT and SR
$\hat{\rho}_s$	-0.569	-0.201	0.159
p-value	< 0.0001	< 0.0001	0.0002

operator with $d = 1$ and the seasonal difference operator with $D = 1$, $s = 7$ to TED, respectively.

Based on the autocorrelation and partial autocorrelation of stationary time series (see Fig. 5), we identify the following $ARIMA(0,1,4)$, $ARIMA(1,1,1)$, and $SARIMA(0,1,1)(0,1,1)_7$, for OT, SR, and TED, respectively:

$$\nabla^1 Z_t^{\mathrm{OT}} = (1 + 0.093B + 0.139B^2 + 0.162B^3 + 0.110B^4)\varepsilon_t \quad (10)$$

$$(1 - 0.277B)\nabla^1 Z_t^{\mathrm{SR}} = (1 + 0.791B)\varepsilon_t \quad (11)$$

$$\nabla^1 \nabla_7^1 Z_t^{\mathrm{TED}} = (1 + 0.1B)(1 + 0.771B^7)\varepsilon_t \quad (12)$$

whose coefficients and residuals are successfully tested through the Student-t and the Ljung-Box test, respectively. The uncorrelated residuals of the estimated models enable capturing the serial dependence of the original correlated time series. Before jointly analysing the considered residual time series, we investigate their pairwise relationship. Tab. 2 and Fig. 6 present the estimated Spearman's correlation coefficient $\hat{\rho}_s$ of each pair of variables and their scatter plot, respectively, confirming the presence of association between the variables and, in particular, asymmetry and non linearity behaviours.

5.2 Joint dependence modelling

As reported in Section 4, the multivariate dependence among the OT, SR, and TED series of residuals is analysed through the copula function. U_1, U_2, and U_3 are uniform variables obtained from the probability integral transform of the residual OT, SR, and TED, respectively. The extreme value copula function is tested using the Kojadinovic et al. (2011) and Rémillard and Scaillet (2009) tests, both rejecting the null hypothesis of the extreme value copula. Thus, we work on the Elliptical and Archimedean families and Joe copulas. We thus estimate symmetric copulas (Gaussian and Student-t) and asymmetric copulas (Clayton, Gumbel, and Joe) taking into account the negative association between residuals of TED and both residuals of OT and SR, and the positive association between OT and SR. As noted by a referee, the different behaviours of each pair of considered variables can be captured when a more flexible copula model is used. Hence, we estimate unstructured Gaussian and Student-t copulas as well as copula mixtures.

Specifically, we estimate the 10 previously introduced copula models with the pseudo-maximum likelihood method, and select the best one based on the selection model criteria LL, AIC, and CV defined in Section 4. Tab. 3 shows the estimation results of the three-dimensional copula models considered. According to all the selection model criteria, the best copula model is the Student-t with an unstructured correlation matrix and 13 degree of freedom. Due to the presence of asymmetric

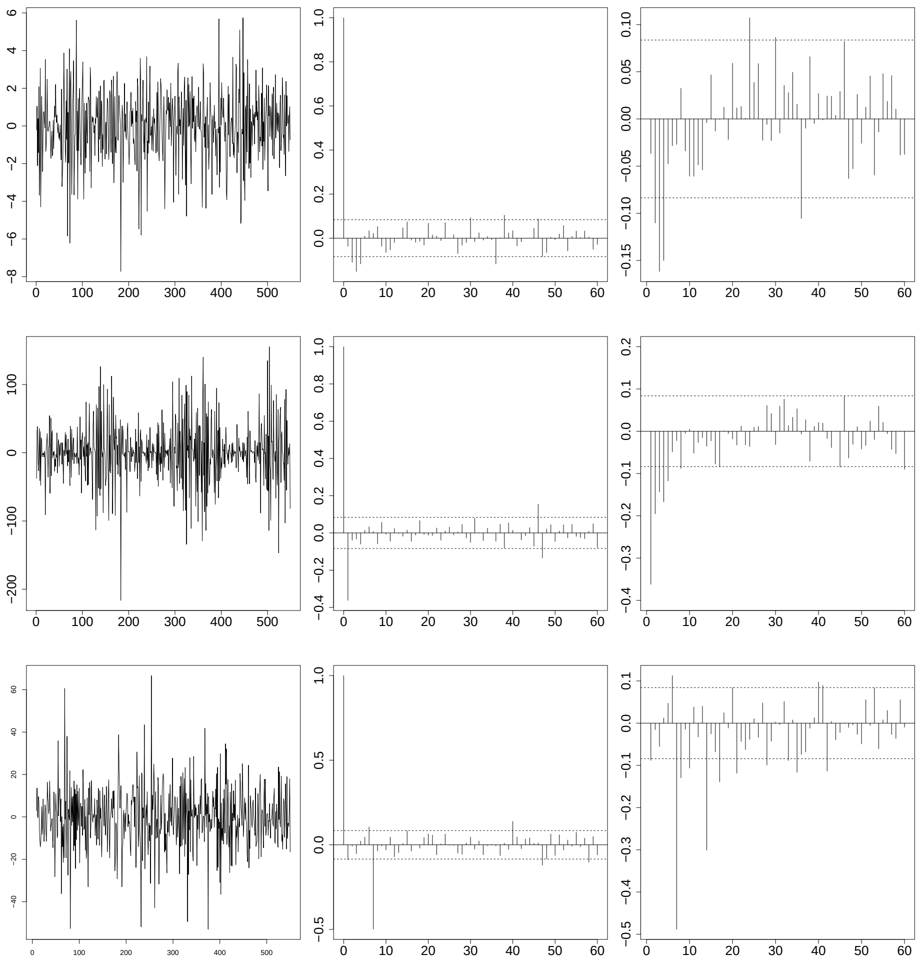

Fig. 5 Plot of stationary time series and the corresponding autocorrelation and partial autocorrelation function: OT in $°C$ (upper panel), SR in W/m^2 (middle panel), and TED in MWh (lower panel).

relationships in our data (see Fig. 3 (right) and Fig. 6), we combine the selected unstructured Student-t copula with the best asymmetric single parameter copula model, which is the Clayton with the third axis flipped as given in Eq. (6), i.e., Clayton$_{(0,0,1)}$. We estimate a mixture copula with two components that, based on AIC and CV (see Tab. 3), appears to be very competitive with respect to the unstructured Student-t copula. Using Eq. (8), the CDF of the selected copula is as follows:

$$\begin{aligned} C^{t\text{-Clay}}_{\text{mix}}(\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\pi}) &= \pi_1 C^t(U_1, U_2, U_3) + \pi_2 C^{\text{Clay}}_{001}(U_1, U_2, 1 - U_3) \\ &= \pi_1(t_{3,\nu}(t_\nu^{-1}(U_1), t_\nu^{-1}(U_2), t_\nu^{-1}(U_3); \boldsymbol{\theta}_t)) + \\ &\quad \pi_2((U_1^{-\theta_{\text{Clay}}} + U_2^{-\theta_{\text{Clay}}} - 1)^{-\frac{1}{\theta_{\text{Clay}}}} - (U_1^{-\theta_{\text{Clay}}} + U_2^{-\theta_{\text{Clay}}} + (1 - U_3)^{-\theta_{\text{Clay}}} - 2)^{-\frac{1}{\theta_{\text{Clay}}}}) \end{aligned} \tag{13}$$

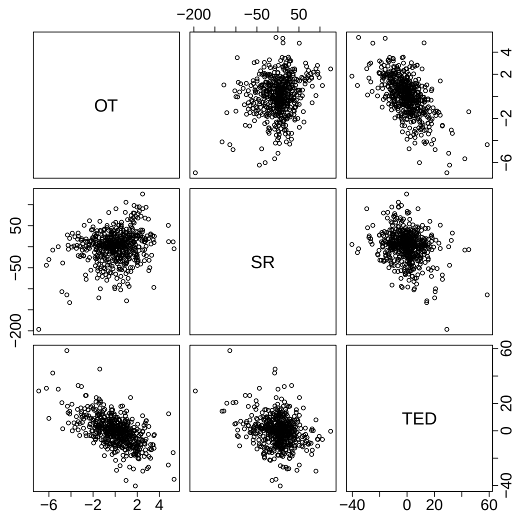

Fig. 6 Pair scatter plots of the OT, SR, and TED residual time series.

Table 3 Estimation results: three-dimensional copula models, estimated dependence parameters $\hat{\boldsymbol{\theta}}$, and the values of the selection model criteria LL, AIC, and CV.

Copula	$\hat{\boldsymbol{\theta}}$	LL	AIC	CV
Gaussian	-0.208	42.63	-83.26	41.47
Unstructured Gaussian	(0.199, -0.593, -0.227)	132.24	-258.47	129.04
Student-t	-0.206	60.97	-117.94	59.03
Unstructured Student-t	(0.185, -0.598, -0.218)	136.28	-264.55	132.19
$\text{Clayton}_{(0,0,1)}$	0.434	74.51	-147.03	73.72
$\text{Clayton}_{(1,1,0)}$	0.389	63.36	-124.73	60.95
$\text{Gumbel}_{(0,0,1)}$	1.234	63.64	-125.27	61.78
$\text{Gumbel}_{(1,1,0)}$	1.230	70.75	-139.50	69.23
$\text{Joe}_{(0,0,1)}$	1.285	42.62	-83.24	40.13
$\text{Joe}_{(1,1,0)}$	1.284	51.41	-100.82	49.50
Student-t–$\text{Clayton}_{(0,0,1)}$ mixture		137.49	-260.98	132.01
Student-t component	(0.159, -0.647, -0.204)			
$\text{Clayton}_{(0,0,1)}$ component	0.493			
Weights	(0.847, 0.153)			

where $C^t(\cdot)$ is the unstructured three-dimensional Student-t copula and $C_{001}^{\text{Clay}}(\cdot)$ is the three-dimensional flipped Clayton copula. Moreover, $t_{p,\nu}(\cdot)$ is the standard p-variate Student-t distribution with ν degrees of freedom (recall that ν controls the heaviness of the tails), $t_\nu^{-1}(\cdot)$ denotes the inverse univariate Student-t distribution function, $\boldsymbol{\theta}_t$ is the unstructured correlation matrix, and $C_{001}^{\text{Clay}}(U_1, U_2, 1-U_3) = C^{\text{Clay}}(U_1, U_2) - C^{\text{Clay}}(U_1, U_2, 1-U_3)$ where $C^{\text{Clay}}(U_1, U_2) = (U_1^{-\theta_{\text{Clay}}} + U_2^{-\theta_{\text{Clay}}} - 1)^{\frac{1}{\theta_{\text{Clay}}}}$ and $C^{\text{Clay}}(U_1, U_2, U_3) = (U_1^{-\theta_{\text{Clay}}} + U_2^{-\theta_{\text{Clay}}} + U_3^{-\theta_{\text{Clay}}} - 2)^{\frac{1}{\theta_{\text{Clay}}}}$ are the bivariate and three-dimensional Clayton copula density with dependence parameter θ_{Clay} (Cherubini et al., 2004), respectively. The selected mixture model has the following estimated parameters: $\hat{\boldsymbol{\theta}}_t = (0.159, -0.647, -0.204)$, number of degrees of freedom $\nu = 15$, $\hat{\theta}_{\text{Clay}} = 0.493$ and $(\pi_1 = 0.847, \pi_2 = 0.153)$, and is thus able to capture both the symmetric and asymmetric behaviour and take into account different bivariate relationships. The two-component mixture copula model attributes about 15% of the joint distribution to the flipped Clayton component of the mixture indicating that upper-left-tail dependence, that is the dependence in the corner (0,0,1), is both stronger and more prevalent. The unstructured Student-t copula and the selected mixture copula appear somewhat similar, at least at a cursory glance. However, they treat tail dependence differently, which can lead to substantially different probability calculations of tail events that are our main endpoint, as we will show in the next section (Fig. 10). Fig. 7 presents the contour plot of the bivariate marginal copula densities of the selected mixture, which as expected shows the diverse relationships between the considered variables. Fig. 8 on the left shows the scatter plot of the selected copula model.

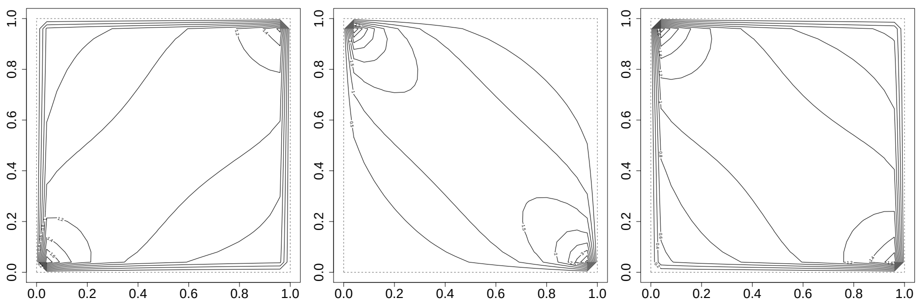

Fig. 7 Contour plot of the bivariate marginal copula densities of the selected unstructured Student-t–Clayton$_{(0,0,1)}$ mixture copula: OT-SR residual time series (left), OT-TED residual time series (middle), and SR-TED residual time series (right).

5.3 Conditional dependence modelling

Here we study the conditional copula function of thermal energy demand given the meteorological variables, i.e., outdoor temperature and solar radiation, applying Eq. (9) to the selected unstructured Student-t–Clayton$_{(0,0,1)}$ mixture copula model:

$$\begin{aligned} P(X_3 > x_3 | X_1 < x_1, X_2 < x_2) &= 1 - P(X_3 < x_3 | X_1 < x_1, X_2 < x_2) \\ &= 1 - \frac{C(F_1(x_1), F_2(x_2), F_3(x_3))}{C(F_1(x_1), F_2(x_2))} \\ &= 1 - \frac{C_{\text{mix}}^{t\text{-Clay}}((u_1, u_2, u_3), \boldsymbol{\theta}, \boldsymbol{\pi})}{C_{\text{mix}}^{t\text{-Clay}}((u_1, u_2), \boldsymbol{\theta}, \boldsymbol{\pi})} \end{aligned} \tag{14}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{t}}, \theta_{\text{Clay}})$, $\boldsymbol{\pi} = (\pi_1, \pi_2)$, and

$$C_{\text{mix}}^{t\text{-Clay}}((u_1, u_2), \boldsymbol{\theta}, \boldsymbol{\pi}) = \pi_1 (t_{2,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2); \theta_t)) + \pi_2 (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}.$$

Fig. 8 on the right presents the conditional probability in Eq. (14), where $u_3 \in [0,1]$ and $u_1, u_2 = (0.01, 0.05, 0.15, 0.3)$. Thermal energy demand shows a negative exponential shape. Noteworthy is that the probability of high thermal energy demand steeply increases when a small change in the meteorological variables is detected, especially for extreme weather scenarios. As an example, the probability that TED is greater than its 75th percentile (193.6 MWh) conditional on the 5th percentile of OT and SR ($-1.32°C$ and 21 W/m^2, respectively) is 0.83, while when the probability is conditioned on the 1st percentile of OT and SR, which corresponds to a $-2.65°C$ and 10 W/m^2 of radiation, it increases to 0.92. In addition, when both the meteorological variables assume values smaller than 0.3 (i.e., OT$= 3.44°C$ and SR$= 63.3W/m^2$), the conditional probability distribution of interest appears to be weakly exponential. Based on the graph in the left part of Fig. 9, we can assert

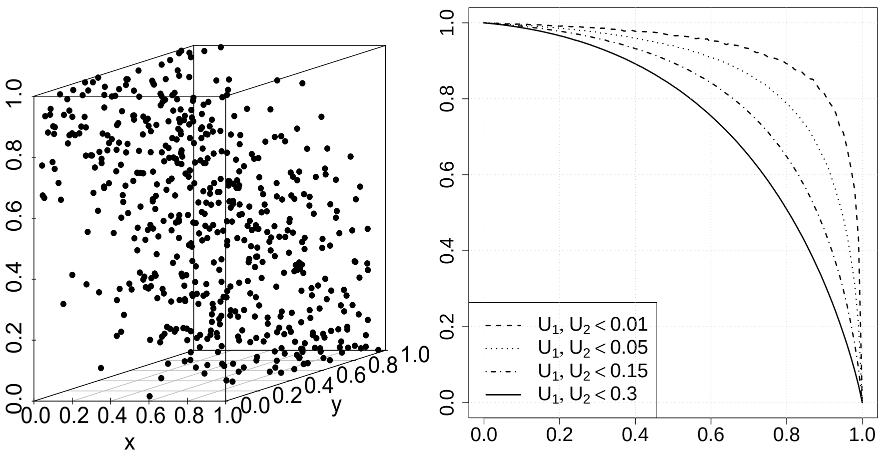

Fig. 8 Scatter plot of the selected unstructured Student-t–Clayton$_{(0,0,1)}$ mixture copula: OT (x-axis), SR (y-axis) and TED (z-axis) (left) residual time series. Copula-based conditional probability function in Eq. (14): residual TED quantile (x-axis) (right).

that for extreme values of outdoor temperature ($U_1 < 0.15$, $0.797°C$), the effect of decreasing values of solar radiation on the probability of high thermal demand appears to be moderate. On the contrary, outdoor temperature, when solar radiation is lower than 0.15, i.e., lower than $39.457W/m^2$, shows a strong impact of the probability of high thermal demand (Fig. 9, right part). For example, $P(X_3 > x_3|X_1 < x_1, X_2 < x_2) = P(U_3 > 0.75|U_1 < 0.15, U_2 < 0.01) = 0.80$, whereas $P(X_3 > x_3|X_1 < x_1, X_2 < x_2) = P(U_3 > 0.75|U_1 < 0.01, U_2 < 0.15) = 0.90$. Hence, despite the exchangeability of the selected components of the mixture model, the flexibility of the selected mixture copula enables assessing the different contribution of the considered meteorological variables. In Fig. 10 we show the divergence between the conditional copula probability function in Eq. (9) computed for the two best estimated models (unstructured Student-t and unstructured Student-t–Clayton$_{(0,0,1)}$ mixture copula) by varying the quantiles of TED residual time series and for a certain values of U_1 and U_2. It is evident that the as soon as the meteorological events become extreme, the behaviour of the two models diverges and the

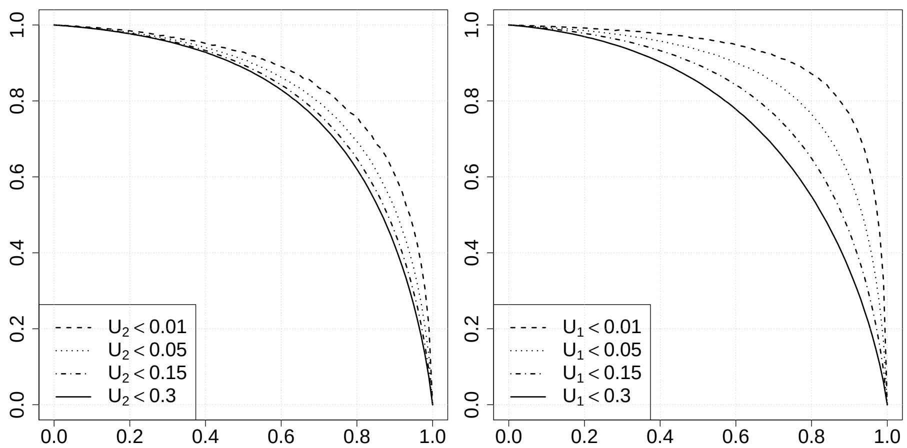

Fig. 9 Mixture copula-based conditional probability function in Eq. (14) with $U_1 < 0.15$ (left) and $U_2 < 0.15$ (right): residual TED quantile (x-axis).

relevance of the flipped Clayton component of the mixture copula clearly emerges.

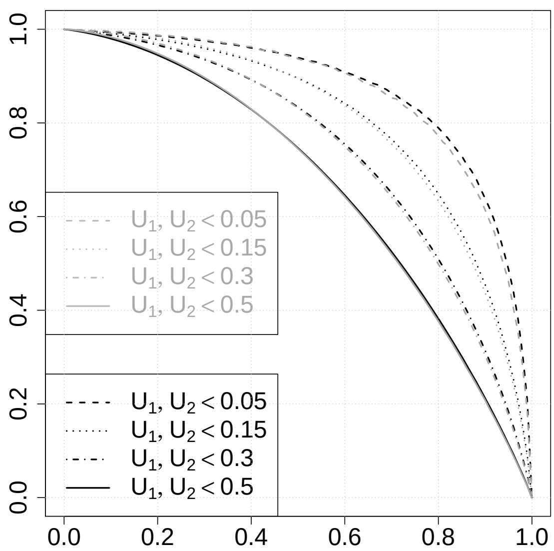

Fig. 10 Conditional probability function in Eq. (9) as written for the mixture copula unstructured Student-t–Clayton$_{(0,0,1)}$ (black lines) and unstructured Student-t copula (gray lines): residual TED quantile (x-axis).

6 Discussion

Here we argue the potential usefulness and implications of the statistical modelling procedure proposed in Section 4 and applied to the case of the Italian city Bozen-Bolzano in Section 5. To do so, we discuss the daily thermal energy production (TEP hereafter) of the DHS of Bozen-Bolzano instead of the TED. In particular, we study the production schedule of the three different production facilities supplying the DHS of Bozen-Bolzano, i.e., waste to energy (WtE) plant, combined heat and power (CHP) engines, and traditional natural gas-fed boilers, taking into consideration their thermal power limits. Since our endpoint is the conditional probability of exceeding the power limits (see Eq. (14)), we derive thermal energy demand (X_3 in Eq. (14)) from thermal energy production by subtracting the distribution heat losses of 9.4% according to Dalvit (2017). Moreover, due to missing information on the DHS facilities in the years $2015-2017$, we perform the analysis only on the heating season $2014 - 2015$.

The WtE plant has a nominal thermal capacity of 32 MW, which is, however, limited to 14 MW due to the technical characteristics of connection piping between the pumping station of the DHS and the considered facility. The WtE plant allows disposing municipal solid waste producing electrical and thermal energy. Two methane-fed engines instead produce heat and power co-generation with a total maximum thermal production limit of 3.7 MW. A traditional series of gas-fed boilers has a capacity limit of 35.5 MW covering the remaining thermal energy demand. According to the efficiency and sustainability principles, the order of activation of the three facilities in the DHS of Bozen-Bolzano is as follows: WtE plant, CHP engines, and traditional boilers.

Fig. 11 shows the production shares of the DHS facilities in the selected period of observation. Here we consider the total number of substations (155) instead of the sample of 110 substations used in the previous empirical analysis, since we are interested in the performance of the whole DHS of Bozen-Bolzano. Of note is that WtE shows inconstant behaviour that reflects the intermittent availability of municipal solid waste. When there is a lack of WtE, boilers compensate production. Indeed, boilers are a very important facility, since they cover thermal energy demand when peaks in thermal demand occur and alternative facilities are lacking. Important to underline is that in October 2016 a new thermal storage tank was added to the DHS, and the contribution of boilers became less relevant thanks to the so-called “peak shave” effect. As for CHPs, we note that their contribution is almost constant over time, and they have a smaller capacity than the other facilities.

In Fig. 12, we show if and when thermal energy production overcomes the heat capacity limit of the WtE and CHPs facilities (upper panel), and the corresponding weather conditions (two plots in the middle). Our final aim here is to provide information on the probability of exceeding the capacity limit of the considered facilities to obtain information on the management of the energy facilities. Note that, to compute Eq. (14) at the production level, the thermal energy demand model is rescaled to the selected sample of 110 substations, that is, 52% of total thermal demand. Looking at the days with heat production that exceeds the WtE plant plus CHPs capacity limit, the corresponding OT and SR values are, as expected, very low and the corresponding conditional probabilities are high (grey bands in Fig. 12). Specifically, in those days when OT reaches its minimum

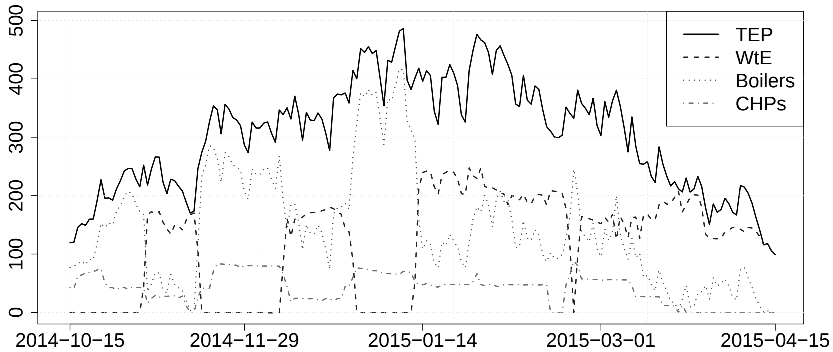

Fig. 11 Thermal energy production time series according to the type of facility: WtE plant, CHP engines, traditional boilers, and TEP (in MWh).

value, which is $-1.1°C$ corresponding to an SR of $64.1W/m^2$, the probability of exceeding the production limit of the most sustainable facilities, WtE plant and CHPs, given in Eq. (14) is 0.73. By comparing this value with that obtained considering the minimum value of SR, which is $24.6W/m^2$ corresponding to an OT of $1.71°C$ and leading to a conditional probability equals to 0.63, we can confirm the different contributions of the two considered meteorological variables on thermal energy production. Interestingly, the two highest peaks of the computed probabilities (0.83 and 0.76 observed on December 27th and December 13th, respectively) correspond to extreme values of solar radiation ($13W/m^2$ and $9.27W/m^2$) and outdoor temperature ($-1.65°C$ and $0.49°$), as expected, but production values not exceeding the considered capacity limits, potentially due to the thermal inertia of the buildings and the so-called calendar effect. As for all the remaining time periods considered, the probability of exceeding the facility limit of the WtE plant and CHPs is coherent with the weather conditions and useful to indicate how thermal energy production should be scheduled.

7 Conclusion

In this study, we have provided a copula-based analysis of thermal energy demand and meteorological variables concerning the DHS of the Italian city Bozen-Bolzano. More precisely, using a mixture copula model, we have analysed the complex relationship between energy demand, outdoor temperature, and solar radiation in a 3-year period, and have derived the conditional probability distribution of energy demand given the weather scenarios. Our main purpose was to provide insights on the impact of weather on thermal energy demand, especially when unexpected and unusual climatic events imply sudden and high energy demand. In-depth knowledge of the conditional probability law allows drawing important implications for planning and efficiently managing the production of energy in a DHS. DHSs play an important role in sustainable energy systems with a relevant contribution to reducing climate change.

Our proposed methodology could be extended in several ways. First, a dynamic copula-based approach could provide further information on the investigated de-

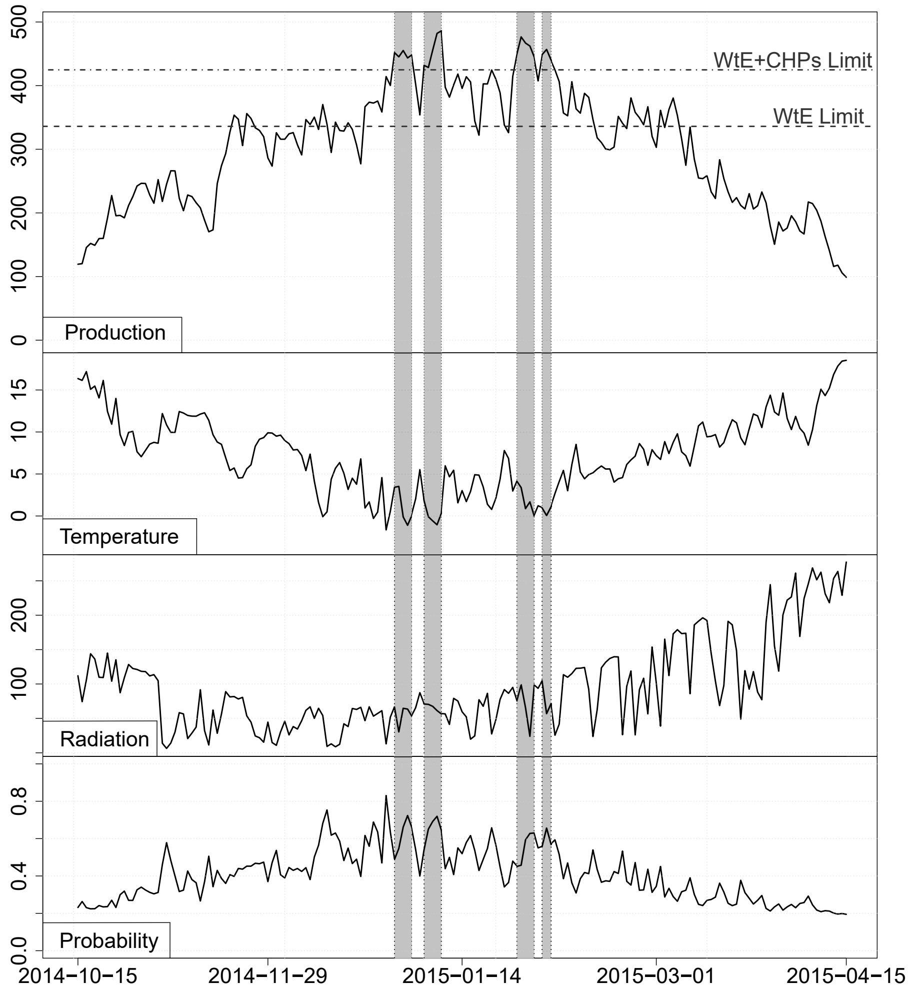

Fig. 12 TEP (in MWh) time series with facility limits of the WtE plant and the WtE plant plus CHPs (upper panel) and the corresponding meteorological time series: OT in $°C$ and SR in W/m^2 (middle panel). Conditional probability in Eq. (14) computed on original data (lower panel). The grey bands highlight the time period exceeding the facility limit of the WtE plant and CHPs.

pendence. Second, it would be interesting to exploit the proposed approach with weather forecasts to obtain information on the next days production schedule. Finally, an extension of our research to other countries, which is completely not trivial, could be useful to provide evidence on the robustness of our results.

Acknowledgements

We thank the referee for comments that greatly improved the development of this work. The first author (corresponding author F. Marta L. Di Lascio) acknowledges the support of the Free University of Bozen-Bolzano, Faculty of Economics and Management, via the project "The use of Copula for the Analysis of Complex and Extreme Energy and Climate data" (CACEEC). The second author (Andrea Menapace) acknowledges Alperia and the Bozen-Bolzano province for providing the analysed data. The third author (Maurizio Righetti) acknowledges the support via the project "Thermo Fluid Dynamics, infrastructures for applied research to business and industry in South Tyrol" (LTFD), ERDF (European Regional Development Fund) 2014-2020 Programme, CUP I52F16000850005.

References

Baey C, Sahlin U, Clough Y, Smith HG (2017) A model to account for data dependency when estimating floral cover in different land use types over a season. Environmental and ecological statistics 24(4):505–527

Behrouzi P, Wit EC (2018) Detecting epistatic selection with partially observed genotype data by using copula graphical models. Journal of the Royal Statistical Society Series C: Applied Statistics p In press

Belzunce F, Mulero J, Ruiz JM, Suárez-Llorens A (2016) On conditional skewness with applications to environmental data. Environmental and ecological statistics 23(4):491–512

Box GE, Jenkins GM (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco

Brechmann EC, Schepsmeier U (2013) Modeling dependence with c- and d-vine copulas: The r package cdvine. Journal of Statistical Software 52(3):543–552

Chen X, Fan Y (2006) Estimation of copula-based semiparametric time series models. Journal of Econometrics 130(2):307–335

Cherubini U, Luciano E, Vecchiato W (2004) Copula methods in finance. John Wiley & Sons Inc., Chichester, West Sussex

Connolly D, Mathiesen B, Østergaard P, Møller B, Nielsen S, Lund H, Trier D, Persson U, Nilsson D, Werner S (2012) Heat Roadmap Europe 1: First Pre-Study for the EU27. Tech. rep.

Connolly D, Mathiesen BV, Østergaard PA, Nielsen S, Persson U, Werner S (2013) Heat Roadmap Europe 2050: Second Pre-study for the EU27. Energy Engineering p 236

Dalvit G (2017) Modelling and Optimisation of a District Heating System. Master's thesis, Free University of Bozen-Bolzano, Italy

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B 39(1):1–38

Di Lascio FML, Giannerini S (2016) Clustering dependent observations with copula functions. Statistical Papers pp 1–17, DOI 10.1007/s00362-016-0822-3

Di Lascio FML, Menapace A, Righetti M (2019) Joint and conditional dependence modeling of peak district heating demand adn outdoor temperature: a copula-base approach. Statistical Methods & Applications pp 1–23, DOI 10.1007/s10260-019-00488-4

DPR Decreto del Presidente della Repubblica n 74 (16 aprile 2013) Regolamento recante definizione dei criteri generali in materia desercizio, conduzione, controllo, manutenzione e ispezione degli impianti termici per la climatizzazione invernale ed estiva degli edifici e per la preparazione dell'acqua calda per usi igienici sanitari, a norma dell'articolo 4, comma 1, lettere a) e c), del d.lgs. 19 agosto 2005, n. 192 della repubblica n. 74

Durante F, Sempi C (2015) Principles of Copula Theory. CRC Press, Boca Raton

Eriksson O, Finnveden G, Ekvall T, Björklund A (2007) Life cycle assessment of fuels for district heating: A comparison of waste incineration, biomass- and natural gas combustion. Energy Policy 35(2):1346–1362, DOI 10.1016/j.enpol.2006.04.005

European Commission (2011) Directorate General for Energy. Energy 2020 e a strategy for competitive, sustainable and secure energy. Publications Office of the European Union, Luxembourg

European Commission (2015) Directorate General for Energy. Towards an Integrated Strategic Energy Technology (SET) Plan: Accelerating the European Energy System Transformation. Publications Office of the European Union, Brussels

Frederiksen S, Werner S (2013) District heating and cooling. Studentlitteratur

Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12:347–368

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

Grler B, Ayyad C, Mateu J (2017) Modelling count data based on weakly dependent spatial covariates using a copula approach: application to rat sightings. Environmental and Ecological Statistics 24(3):433–448

Grønneberg S, Hjort NL (2014) The copula information criteria. Scandinavian Journal of Statistics 41:436459

Hu L (2006) Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics 16(10):717–729

IEA International Energy Agency (2016) World energy balances until 2014. paris

Ilic DD, Dotzauer E, Trygg L, Broman G (2014) Introduction of large-scale biofuel production in a district heating system–an opportunity for reduction of global greenhouse gas emissions. Journal of Cleaner Production 64:552–561

Joe H, Xu JJ (1996) The estimation method of inference functions for margins for multivariate models. DOI http://dx.doi.org/10.14288/1.0225985, URL `https://open.library.ubc.ca/collections/facultyresearchandpublications/52383/items/1.0225985`

Kojadinovic I, Segers J, Yan J (2011) Large-sample tests of extreme-value dependence for multivariate copulas. Canadian Journal of Statistics 39(4):703–720

Liu G, Long W, Zhang X, Li Q (2019) Detecting financial data dependence structure by averaging mixture copulas. Econometric Theory 35(4):777815

Lund H (2007) Renewable energy strategies for sustainable development. Energy 32(6):912–919, DOI 10.1016/j.energy.2006.10.017

Lund H (2014) Renewable energy systems: a smart energy systems approach to the choice and modeling of 100% renewable solutions. Academic Press

Lund H, Mathiesen BV (2009) Energy system analysis of 100% renewable energy systems-The case of Denmark in years 2030 and 2050. Energy 34(5):524–531, DOI 10.1016/j.energy.2008.04.003

Lund H, Hvelplund F, Kass I, Dukalskis E, Blumberga D (1999) District heating and market economy in latvia. Energy 24(7):549–559, DOI 10.1016/

S0360-5442(99)00017-1

Lund H, Werner S, Wiltshire R, Svendsen S, Thorsen JE, Hvelplund F, Mathiesen BV (2014) 4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems. Energy 68:1–11, DOI 10.1016/j.energy.2014.02.089

Lund H, Østeraard PA, Chang M, Werner S, Svendsen S, Sorknæs P, Thorsen JE, Hvelplund F, Mortensen BOG, Mathiesen BV, Bojesen C, Duic N, Zhang X (2018) The status of 4th generation district heating: research and results. Energy 164:147–159, DOI 10.1016/j.energy.2018.08.206

Mathiesen BV, Lund H, Karlsson K (2011) 100mitigation and economic growth. Applied Energy 88(2):488–501, DOI 10.1016/j.apenergy.2010.03.001

Mathiesen V, Pagh M, Scott N, Skougaard P, Scott N (2014) CEESA 100% Renewable Energy Transport Scenarios Towards 2050: Technical Background Report Part 2. DOI ISBN:978-87-91404-35-1

Moritz S, Bartz-Beielstein T (2017) imputets: time series missing value imputation in r. The R Journal 9(1):207–218

Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J (2015) Comparison of different methods for univariate time series imputation in r. arXiv preprint arXiv:151003924

Nelsen RB (2006) Introduction to copulas. New York: Springer

Patton AJ (2012) A review of copula models for economic time series. Journal of Multivariate Analysis 110:4–18

Prenen L, Braekers R, Duchateau L (2017) Extending the archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. Journal of the Royal Statistical Society Series B: Statistical Methodology 79(2):483–505

Qu L, Lu Y (2019) Copula density estimation by finite mixture of parametric copula densities. Communications in Statistics - Simulation and Computation pp 1–23, DOI 10.1080/03610918.2019.1622720

Rémillard B, Scaillet O (2009) Testing for equality between two copulas. Journal of Multivariate Analysis 100(3):377–386

Shi P, Zhang W (2015) Private information in healthcare utilization: Specification of a copula-based hurdle model. Journal of the Royal Statistical Society Series A: Statistics in Society 178(2):337–361

Sklar A (1959) Fonctions de répartition à n dimensions et leures marges. Publications de l'Institut de Statistique de L'Université de Paris 8:229–231

Soutullo S, Bujedo L, Samaniego J, Borge D, Ferrer J, Carazo R, Heras M (2016) Energy performance assessment of a polygeneration plant in different weather conditions through simulation tools. Energy and Buildings 124:7–18, DOI 10.1016/j.enbuild.2016.04.031

Sun Q, Li H, Ma Z, Wang C, Campillo J, Zhang Q, Wallin F, Guo J (2016) A comprehensive review of smart energy meters in intelligent energy networks. IEEE Internet of Things Journal 3(4):464–479

Trivedi PK, Zimmer DM (2005) Copula Modeling: An Introduction for Practitioners, vol 1. Foundations and Trends in Econometrics

Vrac M, Billard L, Diday E, Chédin A (2012) Copula analysis of mixture models. Computational Statistics 27(3):427457

Werner S (2017) International review of district heating and cooling. Energy 137:617–631, DOI 10.1016/j.energy.2017.04.045, URL `https://doi.org/10.`

`1016/j.energy.2017.04.045`

Woods P, Overgaard J (2016) Historical development of district heating and characteristics of a modern district heating system. In: Advanced District Heating and Cooling (DHC) Systems, Elsevier, pp 3–15

Zhao Z, Zhang Z (2018) Semiparametric dynamic max-copula model for multivariate time series. Journal of the Royal Statistical Society Series B: Statistical Methodology 80(2):409–432

Zhu Q, Zhao W (2018) Correcting climate model simulations in heihe river using the multivariate bias correction package. Environmental and ecological statistics 25(3):387–403

Zimmer DM (2012) The role of copulas in the housing crisis. The Review of Economics and Statistics 94(2):607–620

Zimmer DM, Trivedi PK (2006) Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. Journal of Business and Economic Statistics 24:6376