Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 53/ 2018

Joint and conditional dependence modeling of peak district heating demand and outdoor temperature: a copula-based approach

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

Joint and conditional dependence modeling of peak district heating demand and outdoor temperature: a copula-based approach

F. Marta L. Di Lascio[1], Andrea Menapace[2], and Maurizio Righetti[2]

[1]Faculty of Economics and Management, University of Bozen-Bolzano, Italy - Corresponding author. Email: marta.dilascio@unibz.it

[2]Faculty of Science and Technology, University of Bozen-Bolzano, Italy

Abstract

This paper examines the complex dependence between the peak district heating demand and the outdoor temperature. The final aim is to provide the probability law of the heat demand given extreme weather conditions and derive useful implications for the management and the production of thermal energy. We propose a copula-based approach and consider the case of the district of the city of Bozen-Bolzano. The analysed data concerns daily maxima of heat demand observed from January 2014 to November 2017 and the corresponding outdoor temperature. We find that the marginal behavior of the univariate time series of the district heating demand and the temperature is well-described by autoregressive integrated moving average models. Moreover, the selected copula model exhibits a symmetric dependence between the two investigated phenomena that tend to comove closely together during the whole heating season. Taking into account the conditional behaviour of the heat demand given the temperature leads to find that the demand is strongly affected by the temperature and, in case of extreme climatic events, the demand of thermal energy reach a peak with high probability. These findings motivate for improving the production schedule, the system design, and the operational strategies.

Keywords: ARIMA models, Copula function, Conditional probability, District heating system, Outdoor temperature, Peak heat demand.

JEL codes: C10, C32, P28.

1 Introduction

The global demand for energy has increased rapidly in the twentieth century leading to the current energy systems based on fossil fuel [47, 30]. The energy demand is predicted to increase by over 50% within 2030 considering the current energy consumption trend [46]. This situation is aggravated by correlated phenomena, such as the climate change scenario, the increase of greenhouse emissions and the limited nature of the conventional fuels, such as natural gas and oil. In response to the current scenario, there is the necessity to promote energy conservation and efficiency, and new technology introduction [48]. A future energy system, which is based on renewable sources, e.g. solar and wind, and on leftover resources, e.g. biomass and waste, has been identified as the only feasible way to overcome the serious problem of the impact of energy production on the environment [29, 28, 32, 4].

Recent studies [1, 43, 11] argue that the district heating system (DHS, hereafter), which is a system for distributing heat widely employed in many countries such as Denmark and Sweden, has a key role to achieve a sustainable energy system. Thus, the researchers increased their efforts to develop energy efficiency solutions based on DHSs, such as renewable sources integration [38], energy recovery [35], system modeling [14] and optimization [51], production schedules and operating strategies [50]. Nevertheless, the greatest efforts focused on the issues of the supply side and only little attention has been paid to the consumer side [31]. On the contrary, the accuracy of the heat demand models has a relevant influence on every step of the development of the DHSs, like investments planning, system design, operational strategies and controls. For such reason, the research interest on the heat demand analysis represents a crucial challenge and it has recently been growing as underlined by several authors [22, 23, 39, 40, 41, 26, 12, 26, 31].

The main factors that affect the heat demand are building's features, technical characteristics of DHSs (e.g., kind of pipelines, power plants, exchange substations) and meteorological and the socio-economic conditions. Among these factors, the weather, that can be taken into consideration in different ways, e.g. solar radiation, wind speed, temperature, etc., is one of the most relevant [31, 52].

A deep knowledge of the heat demand and its relationship with the weather is crucial for having a reliable forecasting of heat request. Therefore, many approaches have been proposed [47, 40, 31], and many focuses on the investigation of the association between heat demand and outdoor temperature. Dotzauer [13] presents a simple forecasting model using two functions for modeling temperature and social component, respectively, and estimated through a linear least squares method with linear constrains. Popescu et al. [37] propose an enhancement of the method in [13] analysing three different multivariate regression models by varying the set of covariates involved, like solar radiation and indoor/outdoor temperature. Amjady [3] introduces a modified autoregressive integrated moving average (ARIMA, hereafter) for a short-term forecasting of daily peak load by taking into account the temperature. Other works use a two-stage approach that, firstly, performs a functional clustering procedure with the aim to classify the heat load profiles and, secondly, a functional linear regression model for peak load forecasting [20, 16]. Another data-driven approach used especially for short-term forecasting is based on machine learning methods, such as support vector machine [53], regression tree, neural network [25] and multivariate linear regression [8].

Despite the increase of studies in heat demand forecasting, in the literature there are two important lacks: *i*) a proper investigation of the complex relationship among the heat demand and environmental factors, such as outdoor temperature, and *ii*) an in-depth analysis of pick heat demand caused from extreme climatic events. To face with these two gaps we propose a copula-based approach with the aim at providing probabilistic information on the peak heat demand in an innovative way.

The performed analysis consists into three steps. The first step concerns the analysis of the serial correlation of heat demand and outdoor temperature time series through seasonal autoregressive integrated moving average (SARIMA, hereafter) model. The second step investigates the dependence relationship between the two series of uncorrelated ARIMA residuals by using copula models. Once a specific dependence structure is found, the probability of a peak heat demand is derived in the third step of the analysis through the copula-based conditional probability function of heat demand given an outdoor temperature. This procedure is applied to the data concerning the city of Bozen-Bolzano (Italy) that have been detected through the heat exchanger substations of the DHS and one weather station. The daily peak demand and the corresponding outdoor temperature have been observed for three different groups of consumers: all types of DHS's consumers, the residential consumers, and the residential consumers of space heating (without hot water).

To the best of our knowledge, the proposed approach has never been applied to peak heat demand data. Similar approaches have been used in different contexts. Among others, [2] investigate the dependence between crude oil and natural gas prices, [36] analyse the dependence between wind power production and electricity prices, [7] analyse the climatic impact on floods, and [44] employ a copula autoregressive methodology for analysing the wind speed and direction.

The paper is organized as follows. Section 3 describes the collected and analysed data. Section 2 presents the statistical modeling strategy. Section 4 focuses on the empirical analysis and discusses the findings. In Section 5 concluding remarks are provided.

2 Statistical modeling

The used methodology consists of three steps. Firstly, the marginal probability distribution of the two variables of interest, heat demand and outdoor temperature, are estimated. We fit a SARIMA model for time series using the Box&Jenkins procedure [5]. Thus, we account for the non-stationarity of each time series and we model the serial dependence structure in the two series taken separately. Next, the residuals of the two estimated models for univariate time series are computed. Being not autocorrelated, they make it possible to resort to copula theory. Indeed, in the second step of the analysis we model the complex cross-dependence relationship between the heat demand and the temperature through an appropriate copula. Finally, we derive the conditional probability function of the heat demand given the outdoor temperature to provide the probability law of the demand according to extreme climatic events.

2.1 Marginal distribution modeling

To analyse the serial dependence of historical temperature and heat demand we employ the well-known SARIMA models that make it possible to capture both trend (seasonal and not) and autocorrelation. We analyse the two time series taken separately by applying the Box&Jenkins procedure [5], which is the most frequently used due to its simplicity and generality. The SARIMA$(p, d, q)(P, D, Q)_s$ model 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 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, D makes it possible to identify a specific model and to proceed with its estimation.

We firstly check for the non-stationarity of the time series and eventually remove the (linear and seasonal) trend over time by applying the above introduced difference and seasonal difference operators. Next, the identification of the specific SARIMA models is made by using both a graphical tool, i.e. the plot of the autocorrelation and the partial autocorrelation functions of the differentiated series, and the Student-t test on the estimated coefficients. Once a SARIMA model is selected and estimated, the analysis of residual time series is performed. Since we are interested in applying the copula theory, we check that residuals are not autocorrelated and we perform the t-Student test on the autocorrelation function by varying the lag and the Ljung-Box test.

2.2 Dependence structure modeling

Copula function is born with Sklar's theorem [45] that states that every joint distribution function $F(\cdot)$ can be expressed in terms of K marginal distribution functions F_k, with $k = 1, \ldots, K$ and the copula distribution function $C(\cdot) : [0,1]^K \to [0,1]$ as follows:

$$F(x_1, \ldots, x_k, \ldots, x_K) = C(F_1(x_1), \ldots, F_k(x_k), \ldots, F_K(x_K)) \tag{2}$$

for all $(x_1, \ldots, x_k, \ldots, x_K) \in \bar{\mathbb{R}}^K$ (where $\bar{\mathbb{R}}$ denotes the extended real line). Hence, copula is a K-dimensional cumulative distribution function (cdf, hereafter) with standard uniform margins. The above theorem is at the origin of the increasing use of copulas for modeling multivariate continuous distributions. According to the Sklar's theorem, we can split any joint probability function $f(\cdot)$ into two independent parts as follows:

$$f(x_1, \ldots, x_k, \ldots, x_K) = c(F_1(x_1), \ldots, F_k(x_k), \ldots, F_K(x_K)) \prod_{k=1}^{K} f_k(x_k) \tag{3}$$

where $f_k(\cdot)$ with $k = 1, \ldots, K$ are the margins and $c(\cdot)$ is the copula that represents the association among variables, e.g. the multivariate dependence structure of a joint density function [49, 33, 15, for details]. The log-likelihood function of $f(\cdot)$ is, thus, composed of two positive terms as follows:

$$l(\theta) = \sum_{i=1}^{n} \log c \left\{ F_1\left(X_{1i}\right), \ldots, F_k(X_{ki}), \ldots, F_K\left(X_{Ki}\right); \theta \right\} + \sum_{i=1}^{n} \sum_{k=1}^{K} \log f_k\left(X_{ki}\right) \tag{4}$$

where the first term involves the copula density and its parameter θ and the second one involves marginal densities and their parameters. Such separation determines the modeling flexibility given by the copulas. Indeed, beyond the fully maximum likelihood method [10, p.154], which estimates simultaneously both the parameters for the margins and those of the copula, it is possible to decompose the estimation problem in two steps: the first one for the identification of the marginal distributions and the second one for the specification of the appropriate copula model [49, Chapter 4]. It, thus, makes it possible to use any combination of estimation methods for the univariate distributions and the copula. In two-stage maximum likelihood estimation methods, the marginal parameters are estimated in the first step and are used to estimate the dependence parameter of the copula function in the second step. Here we focus on its semi-parametric version [17] where margins are modelled through the empirical cdf $\hat{F}_k\left(X_{ki}\right)$ computed from $X_{k1}, \ldots, X_{kn}$, with $k = 1, \ldots, K$. Then, the copula parameter is estimated through the pseudo-likelihood approach that consists of maximizing the log pseudo-likelihood as follows:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log c \left\{ \hat{U}_{1i}, \ldots, \hat{U}_{ki}, \ldots, \hat{U}_{Ki}; \theta \right\} \tag{5}$$

where $\hat{U}_{ki} = n\hat{F}_k (X_{ki}) / (n+1)$ and $n/(n+1)$ is the scaling factor classically introduced to avoid problems at the boundary of $[0,1]^K$ and it is equivalent to $\hat{U}_{ki} = R_{ki}/(n+1)$ where R_{ki} is the rank of X_{ki} among $X_{k1}, \ldots, X_{kn}$.

Since we are interested in investigating the complex relationship between heat demand and outdoor temperature with the final aim of investigating the demand given extreme weather condition, we introduce the conditional copula function. Given a two dimensional copula $C(\cdot)$ with margins U_1 and U_2, [54] demonstrate that Bayes Rule can also be used to recover conditional copula as follows:

$$C_{u_1|u_2}(u_1, u_2) = \frac{C(u_1, u_2)}{u_2}.$$

Similarly, the conditional probabilities $P(U_1 > u_1 | U_2 < u_2)$ and $P(U_1 > u_1 | u_{21} < U_2 < u_{22})$ computed through the copula, which are more interesting for our analysis, can be expressed as follows:

$$P(U_1 > u_1 | U_2 < u_2) = \frac{u_2 - C(u_1, u_2)}{u_2} \tag{6}$$

$$P(U_1 > u_1 | u_{21} < U_2 < u_{22}) = \frac{u_{22} - u_{21} - C(u_1, u_{22}) + C(u_1, u_{21})}{u_{22} - u_{21}}. \tag{7}$$

Copula models

Here the attention focuses on one-parameter bivariate copulas. In the literature, many different bivariate copula models are available. For a review see [33, 15]. The families considered here are Clayton, Frank, Gumbel, Plackett, Joe, Gaussian, and Student-t, as well as rotated versions of the Clayton, the Gumbel and the Joe copulas. The Elliptical (Gaussian and Student-t) and the Archimedean (Clayton, Frank and Gumbel) families are mostly used but Placket and Joe copulas have been widely used as alternatives to the bivariate Gaussian and Gumbel copula, respectively [24, see, e.g.]. Moreover, the rotation by 90 and 270 degrees leads to the rotated copulas can be obtained using [6]:

$$C_{90}(u_1, u_2) = u_2 - C(1 - u_1, u_2) \tag{8}$$

$$C_{270}(u_1, u_2) = u_1 - C(u_1, 1 - u_2) \tag{9}$$

where $u_k = F_k(x_k)$ with $k = 1, 2$. Note that here we apply the above rotations only to the Clayton, the Gumbel and the Joe copulas and we do not consider the rotation by

180 degrees since, coherently with the dependence structure of the empirical data, we need to capture a negative dependence, i.e. a dependence in the left upper corner and in the right bottom corner (see, for details, the Section 4). We, thus, cover a large set of multivariate features that include asymmetries and heavy tails. The families considered here are defined in Table 1 and displayed in Fig. 1. According to a copula model, the value of θ has a specific meaning. However, it is always true that the greater the value of the dependence parameter, the stronger the association among the margins.

Table 1: Definition of some classic one-parameter bivariate copula models with corresponding range of the dependence parameter θ and its relation with Kendall's τ. u_k with $k = 1, 2$ are uniformly distributed margins so that $x_k = F^{-1}(u_k) \sim F_k$. Φ is the cdf of the standard Gaussian distribution, $\Phi_G(\cdot, \cdot)$ is the standard bivariate Gaussian distribution, $t_{2,\nu}(\cdot, \cdot)$ is the standard bivariate Student-t distribution with ν degrees of freedom (recall that ν controls the heaviness of the tails), and $t_\nu^{-1}(\cdot)$ denotes the inverse univariate Student-t distribution function. $D_1(\theta)$ denotes the Debye function of order one $1/\theta \int_0^\theta t/(\exp^t - 1)dt$.

Copula	$C(u_1, u_2; \theta)$	Parameter range	Kendall's τ
Gaussian	$\Phi_G[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta]$	$\theta \in (-1, 1)$	$\frac{2}{\pi} \arcsin(\theta)$
Student-t	$t_{2,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2); \theta)$	$\theta \in (-1, 1)$ $\nu \in (2, \infty)$	$\frac{2}{\pi} \arcsin(\theta)$
Clayton	$\left[u_1^{-\theta} + u_2^{-\theta} - 1\right]^{-\frac{1}{\theta}}$	$\theta \in (0, \infty)$	$\frac{\theta}{\theta+2}$
Frank	$-\frac{1}{\theta} \ln \left\{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{(e^{-\theta} - 1)}\right\}$	$\theta \in (-\infty, \infty)$	$1 - \frac{4}{\theta}[1 - D_1(\theta)]$
Gumbel	$e^{\left[-(-\log u_1 - \log u_2)^{1/\theta}\right]}$	$\theta \in [1, \infty)$	$1 - \frac{1}{\theta}$
Plackett	$\frac{1 + (\theta-1)(u_1+u_2) - \sqrt{[1+(\theta-1)(u_1+u_2)]^2 - 4\theta(\theta-1)u_1 u_2}}{2(\theta-1)}$	$\theta > 0$	$\frac{\theta+1}{\theta-1} - \frac{2\theta \log \theta}{(\theta-1)^2}$ [1]
Joe	$1 - \left[1 - \left\{\left[1 - (1-u_1)^\theta\right]^{1/\theta} + \left[1 - (1-u_2)^\theta\right]^{1/\theta} - 1\right\}^\theta\right]^{1/\theta}$	$\theta \in [1, \infty)$	$1 + \frac{4}{\theta^2} \int_0^1 t \ln(t)(1-t)^{2(1-\theta)/\theta} dt$

[1] Spearman's ρ_S instead of the Kendall's τ, since it cannot be computed analytically for the Plackett copula.

Copula model specification and quantile dependence

An important issue is the specification and the selection of the copula model. Nowadays, there is a large number of procedures proposed in the literature to test a copula. See [9, 18] for extensive reviews.

To compare different models, following [34], we perform the goodness-of-fit test (GOF, hereafter) by [18] and the leave-one-out cross-validation method (CV, hereafter) developed for copulas [21]. Moreover, we employ the well-known following Akaike information criterion (AIC, hereafter):

$$AIC(h) = -2\text{LL} + 2h \tag{10}$$

where LL is the logarithm of the maximized likelihood function of the estimated copula model, h is the total number of parameters, which is different from one only for the Student-t copula, where $h = 2$. Moreover, in order to test of multivariate extreme-value dependence we employ two tests. One is defined in [19] and it is based on the bivariate probability integral transformation. The other one has been proposed by [27, 42] and it is based on the empirical copula and max-stability, where the approximate p-value is obtained by means of a multiplier technique.

Another useful exploratory method is based on the comparison between the quantile dependence of the empirical data with that of parametric models. The quantile dependence [34, 36] for negatively dependent variables is defined as follows:

$$\lambda_L^{\tilde{q}} = \text{P}(U_1 \leq \tilde{q} | U_2 \geq 1 - \tilde{q}) \quad \text{for } \tilde{q} \in (0, 0.5] \tag{11}$$

$$\lambda_U^{\tilde{q}} = \text{P}(U_1 > \tilde{q} | U_2 < 1 - \tilde{q}) \quad \text{for } \tilde{q} \in (0.5, 1) \tag{12}$$

where $\lambda_L^{\tilde{q}}$ is the lower quantile dependence and $\lambda_U^{\tilde{q}}$ is the upper quantile dependence that, computed at different quantiles $\tilde{q}$, provide a richer description of the dependence structure within the data. This can help narrow down the set of possible parametric copulas to a collection of models that are able to capture some of the characteristics we observe in the data.

3 The data sets

The data set collected in this work concerns the heat demand of consumers connected to the DHS of the city of Bozen-Bolzano (Italy) and the outdoor air temperature as detected by the S. Maurizio weather station. The period of observation goes from September 2014 to August 2017.

Bozen-Bolzano is a city of about one hundred thousand residents, located in the northern part of Italy in a region characterised by alpine weather conditions. The actual DHS

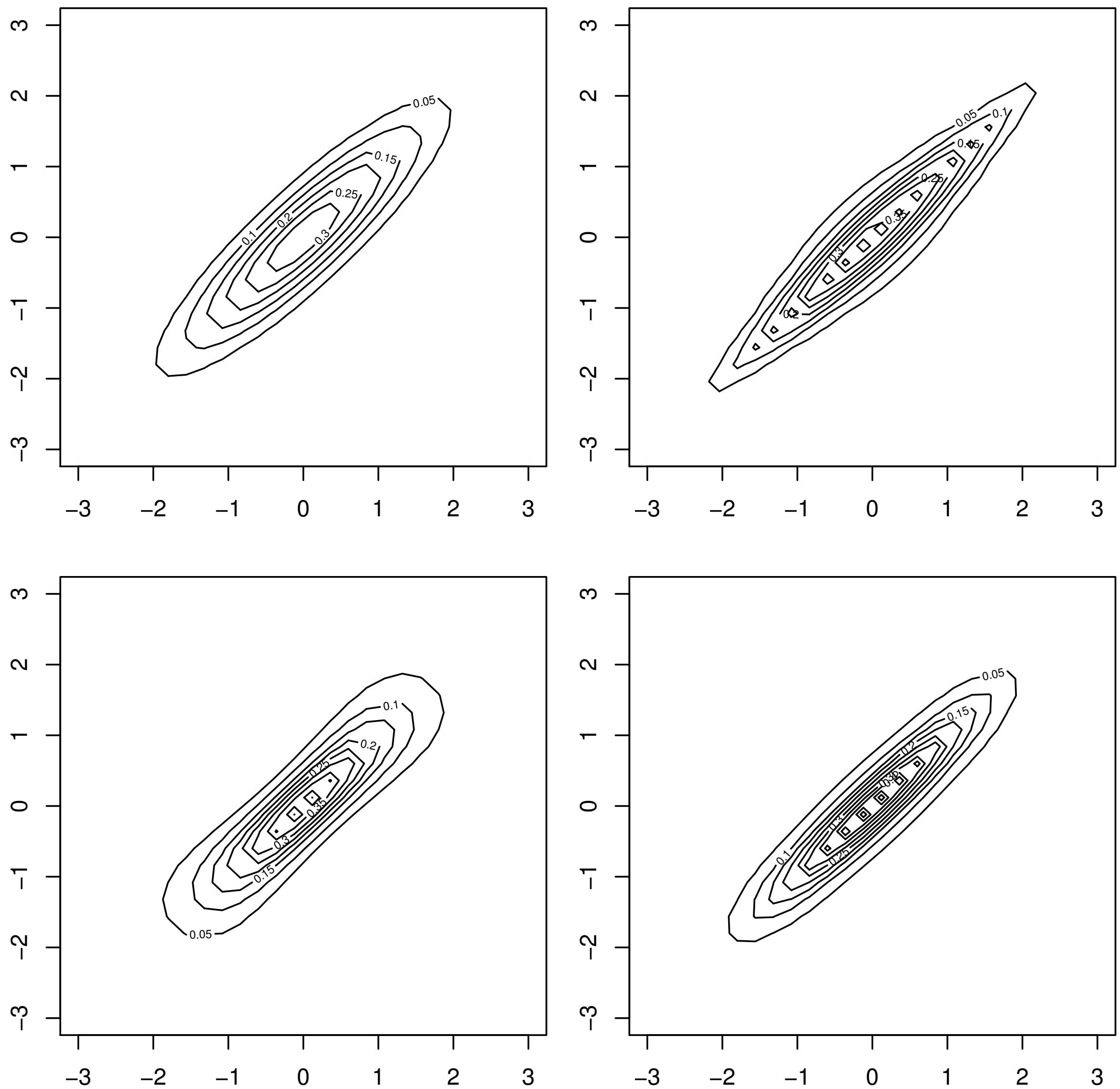

Figure 1: Contour plots of the (bivariate) copula models defined in Table 1 with standard Gaussian margins and a Kendall's τ coefficient of 0.7. Upper panel (from left to right): Gaussian and Student-t copula models. Lower panel (from left to right): Frank and Plackett copula models.

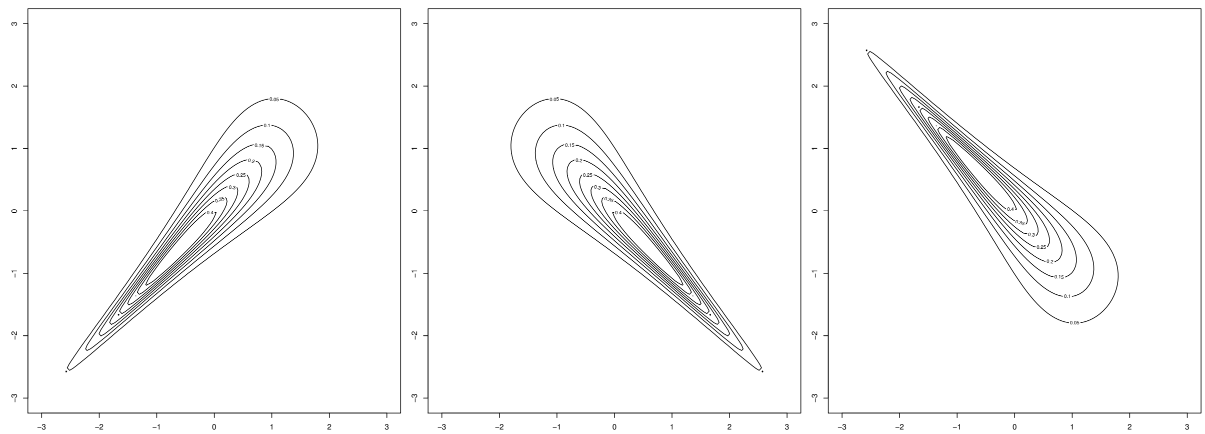

Figure 2: Contour plots of the rotated versions of (bivariate) Clayton copula models defined in Table 1 with standard Gaussian margins and a Kendall's τ coefficient of 0.7 or -0.7, according to the rotation. From left to right: Clayton, rotated Clayton by 90 degrees, rotated Clayton by 270 degrees.

of Bozen-Bolzano consists of a production branch that includes a waste-to-energy plant with a capacity of thermal 32 MW, a combined heat and power production with two methane combustion engines, each one having a thermal capacity of 1.85 MW, and a 6 backup boilers with a total thermal capacity of 11.5 MW. Moreover, a thermal storage with capacity of 5850 m^3 has been installed from October 2016 and the distribution network of about 18 Km is continuously in expansion. This DHS feeds about 3.500 flats and 100 industrial-commercial outlets exchanging heat power from the distribution network to consumers through about 200 thermal exchange substations.

Given the gradual increase in the number of connections of the DHS, starting from 155 substations in 2014 to 194 in 2017, the analysis performed in this paper focuses on the sample of 110 substations with a complete time series. Each substation detects the heat demand every 15 minutes. The selected 110 time series of heat demand have been filtered out of the data affected by measurement's error, i.e. values greater 1.1 times than the maximum thermal power for each exchanger substation. In addition, 5 days of observations (2014-11-30, 2014-12-13, 2016-02-12, 2016-02-22, and 2017-01-03) have been undetected and, therefore, deleted from the time series. Next, we have opportunely aggregated, i.e. averaged, the data to obtain hourly observations. As for the meteorological data, the S. Maurizio's weather station detected the outdoor temperature every 10 minutes; thus, we computed the average of data to obtain hourly observations.

Since the focus of this paper is the relationship between extreme temperature and peak heat demand, we perform the analysis on the maximum hourly total heat demand per day and the corresponding outdoor temperature. The first data set we analyse, called Total

Demand (TD, hereafter), contains data of 110 substations concerning all the types of consumers. Additionally, other two data sets have been analysed. One, called Residential Demand (RD, hereafter), is a subset of the TD data set and contains the data detected by 59 exchanger substations of residential users only. The other one, called Space Heating Residential Demand (SHRD, hereafter), is a subset of the RD data set and it is composed by the data coming from the 19 substations devoted exclusively for the space heating demand of residential users.

Finally, we have split the data in heating season and non-heating season, due to the relevant differences of impact factors driving the heat demand in the two periods. Indeed, the heating seasonal consumption is principally affected by buildings' characteristics and microclimate features, such as outdoor temperature, humidity, wind speed and solar radiation. On the other side, the non-heating seasonal consumption is mainly affected by the occupant behaviour (indoor condition). Hence, we focus the attention on the heating season, when the consumption of thermal energy increases and the highest peaks of heat demand are observed. To identify the heating season we select the threshold corresponding to the point where a sharp change in the quantile function of the heat demand occurs. Fig. 3 shows the quantile plot of the heat demand in the TD (left), the RD (middle), and the SHRD (right) data set. For all the three data sets, the selected threshold corresponds to the 45*th* percentile. The heating season data is, thus, identified by selecting the values greater or equal to the threshold 2860.4 *kW* for the TD, 2404.9 *kW* for the RD, and 797.5 *kW* for the SHRD data set, respectively. Note that from now on the acronyms TD, RD, and SHRD will be used to indicate the three data sets defined above but referred only to the heating season data.

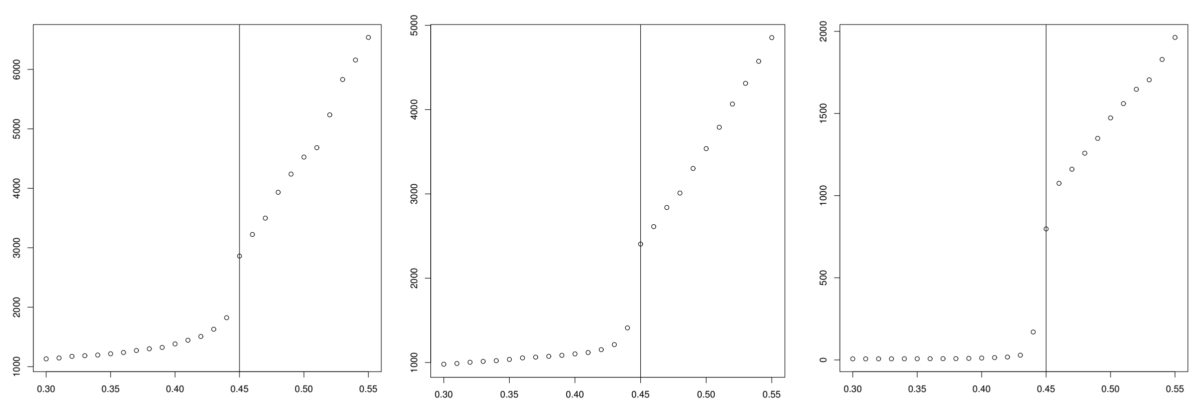

Figure 3: Quantile plot for selecting the threshold of the heat demand (thermal power in kW) to identify the heating season in the TD data set (left), the RD data set (middle) and the SHRD data set (right).

4 Empirical illustrations

In this Section we apply the statistical procedure introduced in Section 2 for analysing the heat demand of the DHS of the city of Bozen-Bolzano.

4.1 Peak heat demand modeling for TD data set

In order to study the TD data set, we perform an analysis into three main steps as described in Section 2. First, we model the univariate time series of the peak heat demand (PHD, hereafter) and outdoor temperature (OT), taken separately, through the SARIMA modeling and, next, we obtain the corresponding residual time series. Second, the dependence analysis between PHD and OT is carried out by copula theory. Finally, the conditional dependence of PHD given OT thought copula is hereby provided to evaluate the impact of extreme climatic events on the heat demand.

Fig. 4 gives an overview of the relationship between the two variables under investigation observed in the whole TD data set (left plot) and in the heating season TD data set (right plot). The Kendall's correlation coefficient between OT and PHD observed during the heating season is $\hat{\tau} = -0.65$, p-value < 0.001 (Pearson's correlation $\rho = -0.84$, p-value < 0.001 and Spearman's correlation $\rho_S = -0.85$, p-value < 0.001), indicating a significant negative relationship between the two variables.

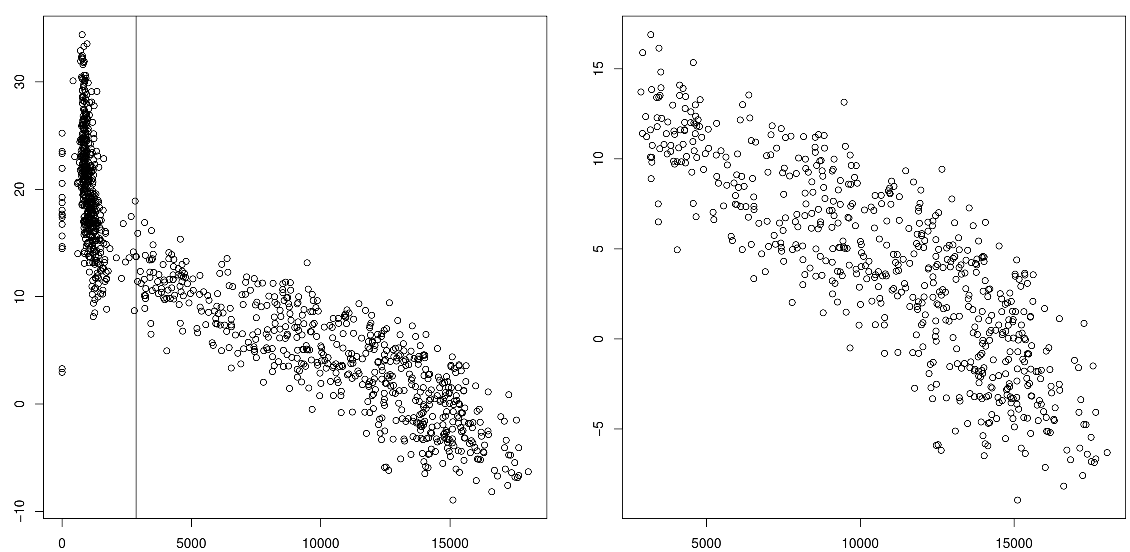

Figure 4: Scatter plots of PHD in kW (x-axis) versus OT in C° (y-axis) of the whole TD data set (left) and of the heating season TD data set (right).

4.1.1 Marginal distribution modeling

Following the Box&Jenkins procedure for both the heat demand and the outdoor temperature, we estimate a SARIMA model for each variable taken separately to obtain the corresponding uncorrelated residual time series. On the basis of the autocorrelation function and the partial autocorrelation function at lag $1,\dots,60$, we check for the non-stationariety of the two time series (see Fig. 5). The autocorrelation function of PHD presents a slight negative linear and seasonal trend meaning that there is a non-stationarity in the mean of the series, while the OT series shows only a linear trend. Both the time series are stationary in variance (range-mean plots do not shown here). We

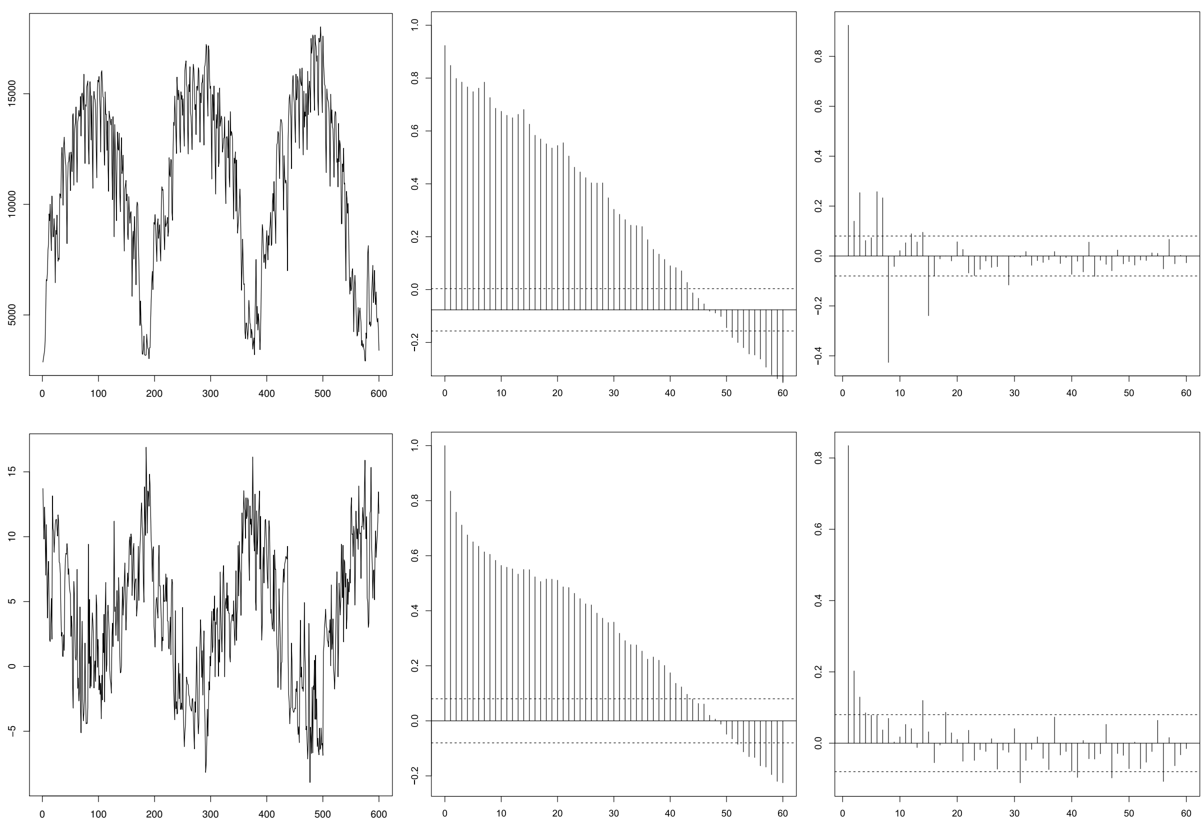

Figure 5: Preliminary time series analysis: plot of time series, autocorrelation function and partial autocorrelation function of PHD (upper panel) and OT (lower panel). Data set: TD, heating season.

remove the non-stationarity using the difference operator with $d = 1$ and the seasonal difference operator with $D = 1$, $lag = 7$ for the PHD series and the difference operator with $d = 1$ for the OT series. Fig. 6 shows the resulting stationary time series. We specify a $SARIMA(0,1,2)(0,1,1)_7$ model for the PHD series. The estimated model results:

$$\nabla^1\nabla^1_7 Z_t = (1 + 0.2915B + 0.1571B^2)(1 + 0.5673B^7)\varepsilon_t \tag{13}$$

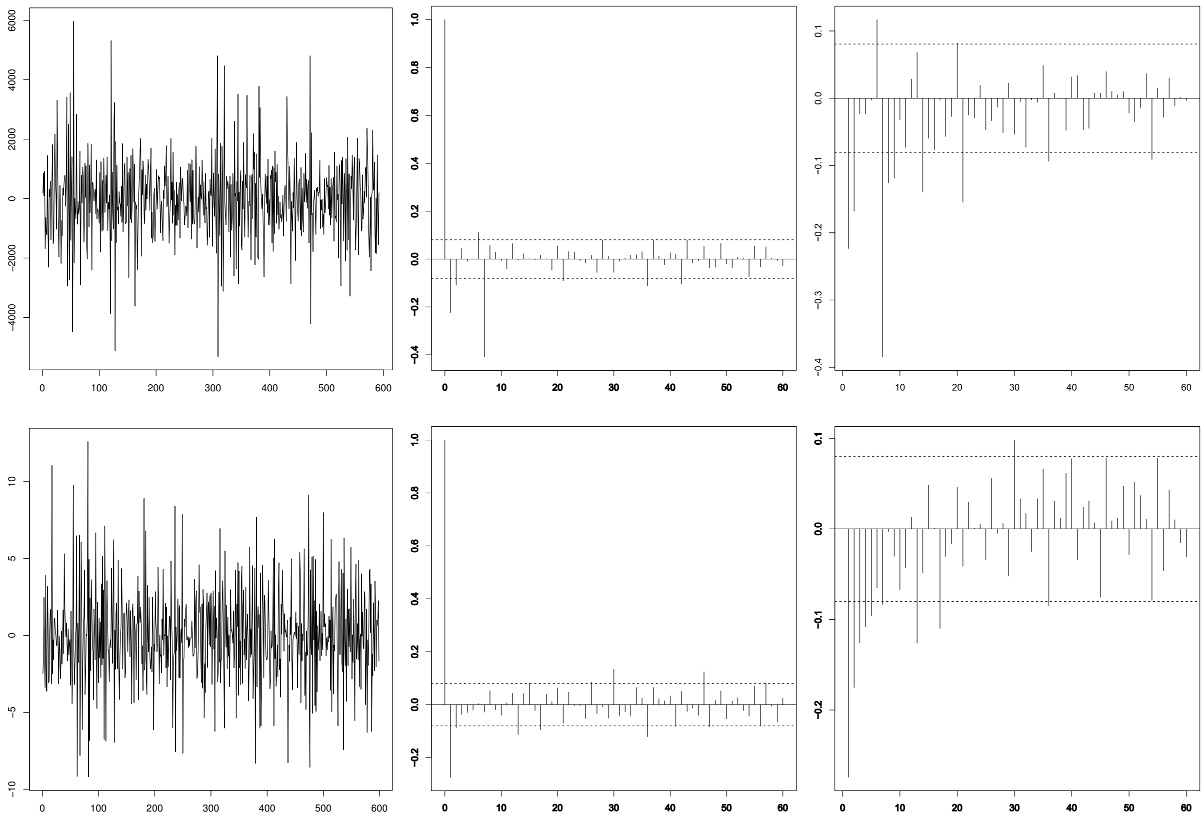

Figure 6: Plots of time series, autocorrelation function and partial autocorrelation function of the stationary PHD (upper panel) and OT (lower panel) series. Data set: TD, heating season.

As for the OT series, we specify an $ARIMA(0, 1, 2)$ model using a forward selection procedure of the coefficients of the starting model $ARIMA(5, 1, 2)$. The estimated model results:

$$\nabla^1 Z_t = (1 + 0.4013B + 0.1884B^2)\varepsilon_t \tag{14}$$

The goodness of the two estimated models is verified through the analysis of the residuals. Fig. 7 shows the residuals of the PHD and OT series, respectively on the left and on the right panel. In particular, we can see that the autocorrelation coefficient of residuals is not significantly different from zero with p-value < 0.05 at every considered lag. Thus, the estimated SARIMA models are able to catch the serial dependence since the residuals can be considered as realization of white noise stochastic processes.

The Fig. 8 shows the scatter plot of the residuals with the histogram of their univariate distribution on the left, and the scatter plot of the pseudo-observations of the two residual series on the right. The Kendall's correlation coefficient between the PHD and OT residual time series is $\hat{\tau} = -0.35$ with p-value < 0.001 (Pearson's and Spearman's correlation results, respectively, -0.49 and -0.5 both with p-value < 0.001). The observed negative

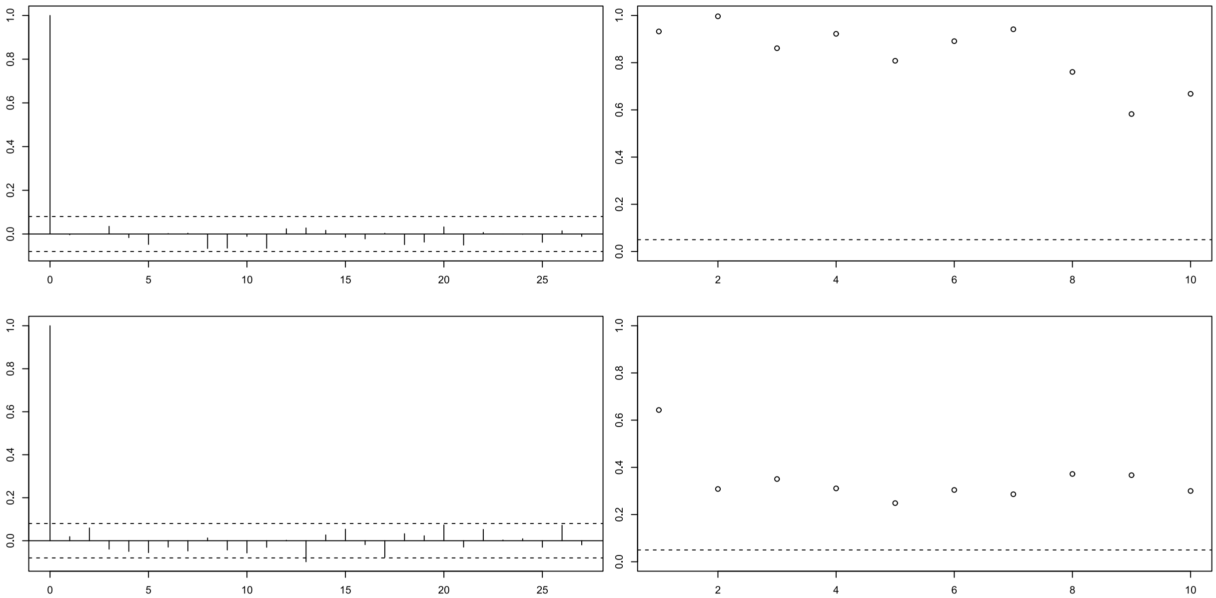

Figure 7: Autocorrelation function and Ljung-Box test on the residuals of the estimated SARIMA models for the PHD (upper) and OT (lower) series. Data set: TD, heating season.

dependence relationship between the two obtained residual series motivates and supports the use of the copula theory to investigate the complex relationship between peak heat demand and outdoor temperature.

4.1.2 Dependence structure modeling

In this section we carry out the analysis of the bivariate dependence between the PHD and OT time series of residuals. As reported in Section 2.2, only one-parameter bivariate copula models are taken into account. Precisely, we estimate the copula models belonging to the Elliptical and Archimedean families, and the Plackett and the Joe copulas (see Table 1 for technical details). We remind that we work with both symmetric copulas (the Frank, the Gaussian, the Student-t and the Plackett) and asymmetric copulas (the rotated Clayton 90°, the rotated Clayton 270°, the rotated Gumbel 90°, the rotated Gumbel 270°, the rotated Joe 90° and the rotated Joe 270°). Note that the Clayton, the Gumbel and the Joe copulas are taken into account only in theirs rotated versions due to the negative association observed between the residuals of the PHD and OT series.

The considered copula models have been estimated through the pseudo-maximum likelihood method (by using the `R` package `copula`). To select the best copula model among the ten we employ different tools: LL, AIC, GOF and CV. In addition, we test for an extreme-value copula through the test in [42] and [27], which both gives a p-value $<$

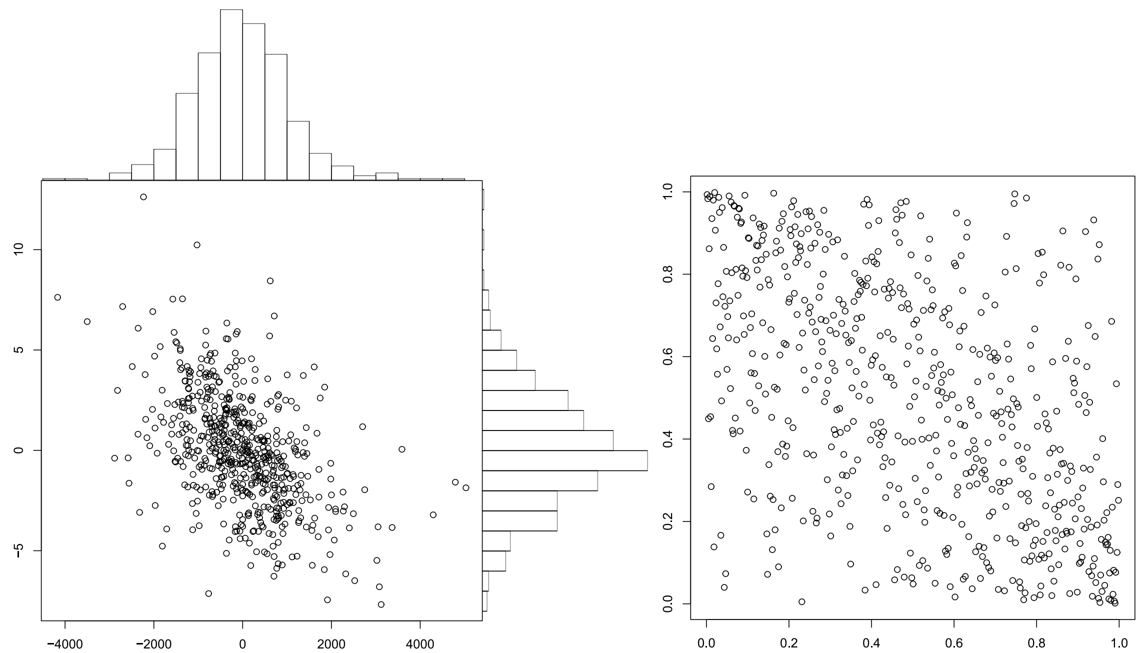

Figure 8: Scatter plot of the PHD residuals (x-axis) versus the OT residuals (y-axis) with their histogram (left) and scatter plot of the pseudo-observations of the OT residuals versus the PHD residuals (right). Data set: TD, heating season.

0.001 leading to reject the null hypothesis of a bivariate extreme-value copula. The estimation results are presented in Table 2. In the light of the obtained results, the best copulas are the Plackett (for the GOF test) and the Student-t with 6.93 degree of freedom (for all the other criteria) with the estimated value of the dependence parameter equals, respectively, $\hat{\theta} = -0.51$ (corresponding to $\tau = -0.357$) and $\hat{\theta} = 0.19$ (corresponding to $\tau = -0.344$). Due to the slight ambiguity of these results we also produce the quantile dependence plot in Fig. 9 for the empirical distribution of the data and the four copula models with the best results in terms of AIC, CV and GOF. On the basis of this plot, the Student-t copula appears to be the only one able to catch the behaviour on the tails, even though the Plackett copula better models the middle of the distribution. Overall, we select the Student-t copula. The simmetry of the selected copula tells us that the two investigated phenomena tend to comove closely together during the whole considered period with a mild dependence.

4.1.3 Conditional probability function

In this Section we exploit the findings about the dependence structure between the peak heat demand and the outdoor temperature to investigate the behaviour of the demand conditionally to a certain value of the temperature. To do that, we employ the conditional

Copula	$\hat{\theta}$	LL	AIC	p-value GOF	CV
Gaussian	-0.50	85.67	-169.34	0.141	84.32
Student-t	-0.51	91.96	-179.92	0.140	90.46
Frank	-3.58	87.21	-172.42	0.469	86.07
Plackett	0.19	90.66	-179.32	0.628	89.60
Clayton (270°)	0.72	65.96	-129.92	< 0.001	63.72
Clayton (90°)	0.74	71.40	-140.80	< 0.001	68.91
Gumbel (270°)	1.48	83.78	-165.56	0.002	81.73
Gumbel (90°)	1.47	82.30	-162.59	< 0.001	80.29
Joe (270°)	1.62	65.20	-128.39	< 0.001	62.51
Joe (90°)	1.59	60.87	-119.74	< 0.001	58.32

Table 2: Results of copula models estimated on the SARIMA residuals of the PHD and OT time series of the TD data set: estimated dependence parameter ($\hat{\theta}$) and evaluation's criteria (LL, AIC, p-value GOF, CV).

probability functions defined in eq. (6) and eq. (7) where U_1 is a uniform variable given by the probability integral transform of the residual PHD time series and U_2 is a uniform variable given by the probability integral transform of the residual OT time series. In particular, we investigate the probability of a peak demand, e.g. a demand greater than the 70th percentile of U_1, given an extreme value or a range of extreme values for the temperature. Fig. 10 shows the copula-based conditional probability function $P(U_1 > u_1|U_2 < u_2)$ in eq. (6), where $u_1 \in [0, 1]$ and $u_2 = (0.01, 0.05, 0.10, 0.15)$. Clearly, the probability that the heat demand is greater than a certain value increases as soon as the outdoor temperature becomes more and more extreme. The conditional probability function shows a negative exponential behaviour and, when $u_1 > 0.7$ ($> 13813, 75$ kW), it passes from 0.631, when the temperature is smaller than its 15th percentile ($-2.477\,C^\circ$), to 0.819 when the temperature is smaller its 1st percentile ($-6.713\,C^\circ$). The probability of having a demand greater than its 70th percentile increases of 1/3 as soon as the temperature decreases of about $4\,C^\circ$. Moreover, considering an interval value for the temperature, we find that $P(U_1 > 0.7|0 < U_2 < 0.01) = 0.819$ and, e.g., $P(U_1 > 0.7|0.10 < U_2 < 0.15) = 0.543$. Thus, the investigated probability almost halves with a slight increase of the interval for U_2, i.e. for the temperature. Coherently to what we found above, the conditional probability of interest decreases as soon as the values for U_2 increases.

To sum up, we stress that the demand of heating is strongly affected by the outdoor temperature and extreme values for the OT determine a high heat demand with high probability. Moreover, an approach based on the dependence relationship between the two

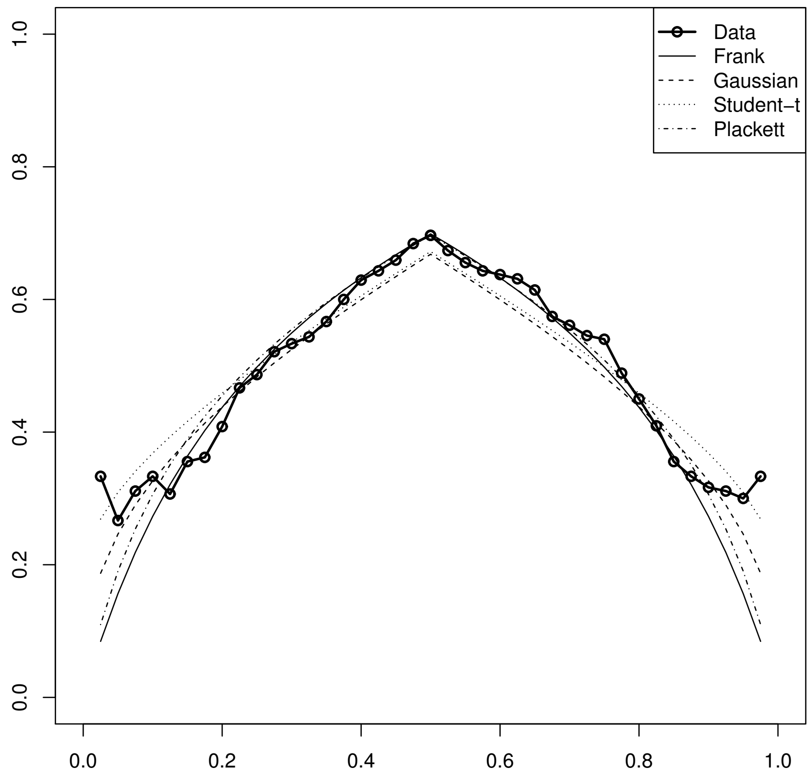

Figure 9: Estimated quantile dependence plot for quantile $\tilde{q} \in [0.025, 0.975]$ and a size step of 0.025 implied by the empirical TD data set and the four best estimated copula models among the ones in Table 1.

phenomena of itnerest is extremely important, since under independence the investigated conditional probability is strongly underestimated, e.g. $P(U_1 > 0.7|u_{21} < U_2 < u_{22}) = 0.3$ irrespectively of the climatic situation.

4.2 Analysis of the RD and SHRD data sets

In this Section we briefly present the results obtained by applying the previously described and shown procedure to the RD and SHRD data sets. Also here we focus the attention on the heating season that is identified by using the quantile plot given in Fig. 3.

The diagnostic plots of the autocorrelation and the partial autocorrelation function lead to the identification of a $SARIMA(2,1,1)(0,0,2)_7$ model for the PHD time series and an $ARIMA(4,1,1)$ model for the OT time series. Both these models shows some coefficients not significant. On the basis of a forward procedure based on the p-value of the Student-t test on coefficients and on the principle of parsimonious, the final identified models are $ARIMA(0,1,1)$ for the PHD and $ARIMA(1,1,1)$ for the OT series. In ad-

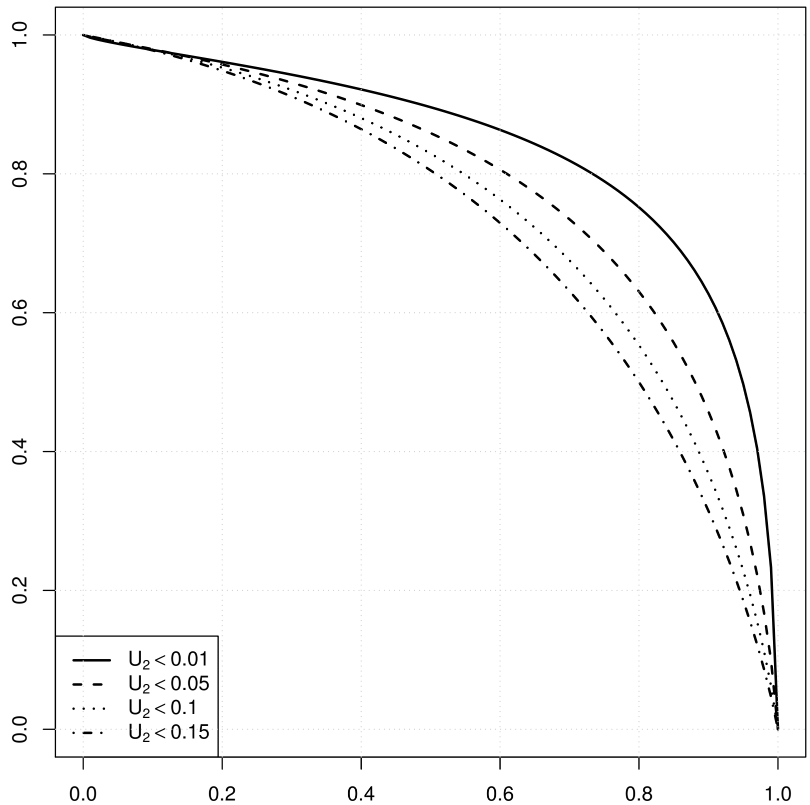

Figure 10: Copula-based conditional probability function in eq. (6) with $u_1 = 0.7$. Data set: TD, heating season.

dition, the Kendall's, Pearson's and Spearman's correlation coefficients are, respectively, -0.50, -0.39 and -0.54 (p-value < 0.001 for all the three). Table 3 shows the results of the estimation of the copula models in Table 1 for the RD data set. Fig. 11 shows in the left upper plot the quantile dependence of the four copula models that better fit the empirical RD data set. In addition, both the tests by [42] and [27] gives a p-value < 0.001 leading to reject the null hypothesis of a bivariate extreme-value copula. The analysis of the dependence structure highlights that the Plackett copula fits better than the other ones. The estimated parameter of the Plackett copula results 0.15, which corresponds to the Kendall's coefficient $\hat{\tau} = -0.402$. Interestingly, both the quantile dependence and the conditional probability function computed for the RD data set (see Fig. 11, right upper panel) exhibit a shape different from the one found for the TD data set. Here, the probability of having a high heat demand increases more slowly than that of the TD data set. Indeed, when $u_1 > 0.7$ (> 9355.09 kW), it passes from 0.669, when the temperature is smaller than its 15th percentile (-2.477) , to 0.734 when the temperature is smaller the its 1st percentile ($-6.713\,C^{\circ}$). Hence, we can argue that the heat demand for residential

Copula	$\hat{\theta}$	LL	AIC	p-value GOF	CV
Gaussian	-0.53	94.55	-187.10	0.037	92.30
Student-t	-0.56	112.51	-221.03	0.290	110.26
Frank	-4.06	106.33	-210.66	0.109	105.10
Plackett	0.15	113.60	-225.19	0.247	112.52
Clayton (270°)	0.78	71.09	-140.18	< 0.001	68.09
Clayton (90°)	0.88	88.22	-174.43	< 0.001	85.09
Gumbel (270°)	1.56	103.88	-205.75	0.003	101.49
Gumbel (90°)	1.54	93.48	-184.96	< 0.001	90.77
Joe (270°)	1.75	85.06	-168.12	< 0.001	82.02
Joe (90°)	1.66	66.80	-131.60	< 0.001	63.57

Table 3: Results of copula models estimated on the SARIMA residuals of the PHD and OT time series of the RD data set: estimated dependence parameter ($\hat{\theta}$) and evaluation's criteria (LL, AIC, p-value GOF, CV).

buildings (of both space heating and hot water) is less sensitive to a change in the outdoor temperature than the heat demand for both residential and non-residential buildings (e.g. industrial buildings, sporting centres, and so on).

Copula	$\hat{\theta}$	LL	AIC	p-value GOF	CV
Gaussian	-0.47	72.82	-143.64	0.103	71.23
Student-t	-0.49	83.84	-163.68	0.260	81.96
Frank	-3.40	79.27	-156.56	0.097	78.09
Plackett	0.20	83.66	-165.32	0.244	82.62
Clayton (270°)	0.65	54.51	-107.03	< 0.001	51.95
Clayton (90°)	0.70	65.24	-128.49	< 0.001	63.03
Gumbel (270°)	1.44	75.72	-149.44	< 0.001	73.73
Gumbel (90°)	1.44	70.60	-139.19	0.001	68.31
Joe (270°)	1.58	59.85	-117.71	< 0.001	57.43
Joe (90°)	1.54	50.47	-98.93	< 0.001	47.64

Table 4: Results of copula models estimated on the SARIMA residuals of the PHD and OT time series of the SHRD data set: estimated dependence parameter ($\hat{\theta}$) and evaluation's criteria (LL, AIC, p-value GOF, CV).

The SHRD data set contains the information on the heat demand concerning only the space heating of the residential buildings. In this case, the identified SARIMA mod-

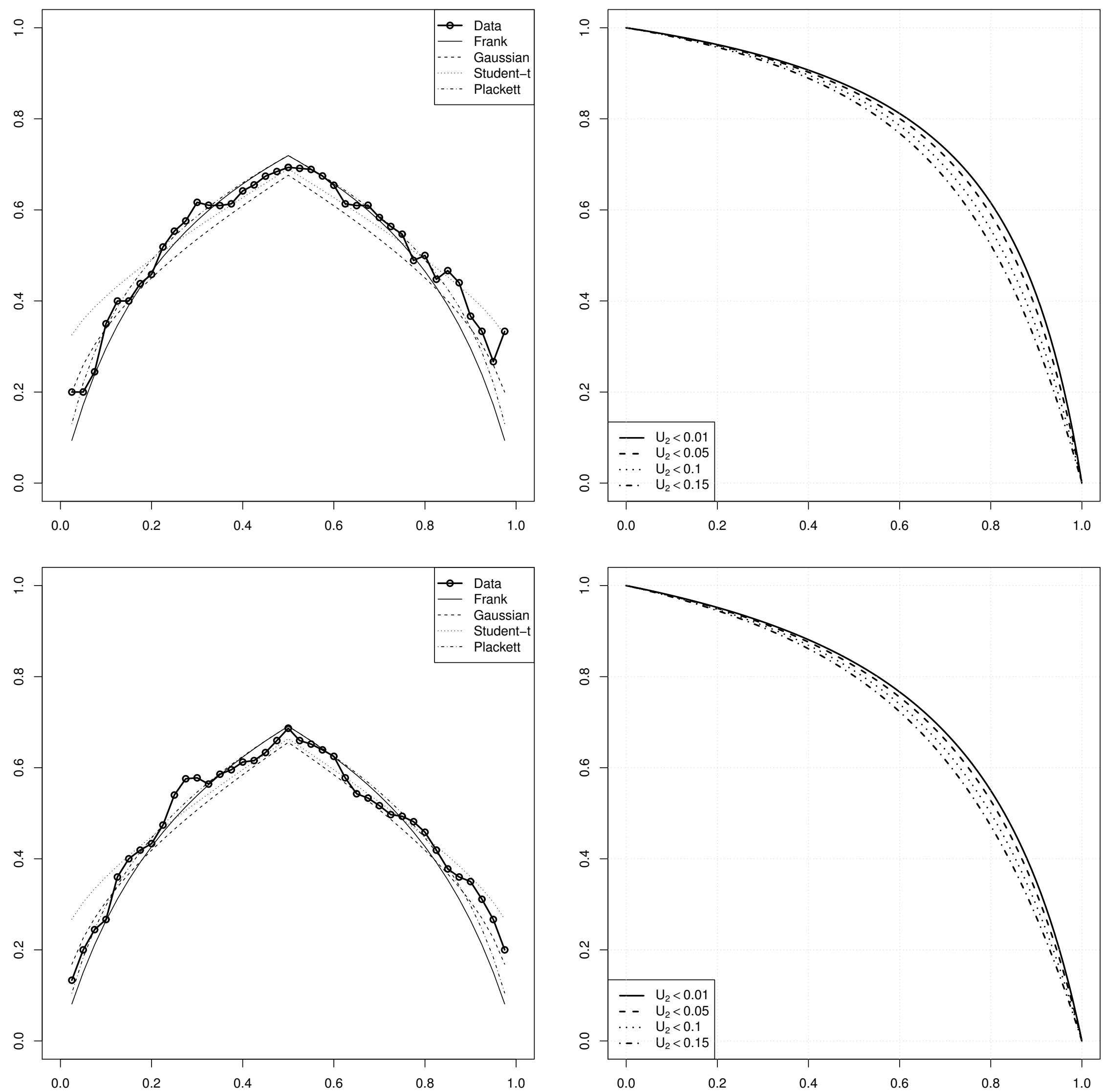

Figure 11: Quantile dependence plot (left) for quantile $\tilde{q} \in [0.025, 0.975]$ and a size step of 0.025 implied by the empirical data and the four best estimated copula models among the ones in Table 1 and conditional probability function in eq. (7) with $u_1 = 0.7$ (right) for the RD (upper) and the SHRD (lower) data sets.

els obtained applying the Box&Jenkins procedure are the $ARIMA(2, 1, 3)$ for the PHD series and $SARIMA(4, 1, 1)(1, 0, 0)_7$ for the OT series. Following a forward procedure based on the p-value of the Student-t test on coefficients, we obtain as final models an $ARIMA(0, 1, 1)$ and an $ARIMA(1, 1, 1)$ for the PHD and OT series, respectively. The Pearson's, the Kendall's and the Spearman's correlation coefficient result -0.42, -0.33 and -0.47, respectively, with a p-value < 0.001 for all the three. The results of the dependence analysis between the residuals of the two estimated ARIMA models (see Tab. 4) provide evidence for both the Student-t and the Plackett copula models with really similar values for all the three selection criteria used. On the basis of the quantile dependence (see Fig. 11, lower left plot) the Plackett copula appears to better fit the left tail dependence. The copula parameter for the Plackett results 0.2 that corresponds to a Kendall's $\hat{\tau} = -0.347$. Also here, the tests for an extreme-value copula proposed by [42] and [27] give a p-value < 0.001 leading to reject the null hypothesis of a bivariate extreme-value copula. Similarly to the results on the RD data set, the demand of heat from residential buildings is less sensitive than that from residential and non-residential buildings (TD data set). On the other hand, given a certain value for the heat demand, the conditional probability of having a pick given a certain temperature decreases very slowly by varying the value of the temperature and it is lower than that finding for the RD data set. This means that the consumption of hot water does not have an important impact on the shape of the dependence structure between OT and PHD.

5 Conclusion and discussion

We analysed the peak district heating demand and its complex relationship with the climatic situation of the Italian city Bozen-Bolzano. The final aim was providing the conditional probability of a pick heat demand given a certain extreme weather condition. The proposed approach is based on the copula theory that makes it possible to keep into consideration and model complex and nonlinear dependencies. Through several tools, e.g. the quantile dependence and the goodness-of-fit copula test, we selected the best copula model for three analysed data sets: one concerning the peak district heating demand of both residential and non-residential buildings, the second one concerns only the residential buildings and the last one concerns the demand of heat without hot water of residential buildings. A symmetric copula model has been selected for all the three cases but with slight differences in the shape and the strength of the dependence relationship since in all the three cases the dependence is mild, with a Kendall's $\hat{\tau} < 0.5$. Next, the copula-based conditional dependence of the heat demand given the outdoor temperature has been studied and, in general, we found that the district heat demand is strongly related to the outdoor temperature and the probability of a peak demand shows an exponential-type increase conditionally to a linear decrease of the temperature.

Our findings have interesting and useful implications in the energy production and market. The knowledge of the probability distribution of the heat demand conditionally to the weather can help to, first of all, improve the production schedule of thermal energy and the management of the thermal energy storage, from the integration with other energy sources to the capacity of the storages. Secondly, if renewable sources have being in use, then the conditional distribution of the demand could help in coping with their fluctuating and intermittent nature. Hence, extreme situations of heat demand could be managed in a more efficient way. Moreover, the probability distribution of peak heat demand could play a role in the optimization of the modeling of district load forecasting to guarantee the best design and operation of distributed energy system. Finally, the energy market as well as energy prices can benefit in terms of stability from a statistical method sensitive to extreme situations.

Our approach can be enhanced in a multivariate copula-based approach by including other factors. On the one hand, the total daily energy demand could be useful to improve the design and the management of the energy storage. On the other hand, further climatic factors, like solar radiation and wind speed, could be useful to provide a more complete description of the weather is given as input to the copula-based conditional probability function. Also, the apparent temperature, which can be computed by exploiting the information about the outdoor temperature, the humidity and the wind speed, could be used as a synthetic description of the weather.

The analysis of the percentiles of peak heat demand showed that the proposed approach can be a potential powerful tool for improving the management of thermal energy and its storage.

References

[1] The role of district heating in future renewable energy systems. *Energy*, 35(3):1381–1390, 2010.

[2] R. Aloui, M. Ben Aissa, S. Hammoudeh, and D.K. Nguyen. Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. *Energy Economics*, 42C:332–342, 2014.

[3] Nima Amjady. Short-Term Hourly Load Forecasting Using Time Series Modeling With Peak Load Estimation.pdf. 16(4):798–805, 2001.

[4] D Balic, O Gudmundsson, V Wilk, D Sokolov, E E Iakimetc, L Grundahl, J Duquette, A R Razani, A B Hansen, B Nastasi, G Vigants, M Kotenko, S Mohammadi, K Qvist, D S Larsen, K Hasberg, S Fritz, D Casetta, J Z Thellufsen, L Zhang, O Martin-du Pan, K Hansen, P Gilski, and C Bevilacqua. Smart Energy Systems and 4th Generation District Heating. *International Journal of Sustainable Energy Planning and Management*, 10(August):25–27, 2015.

[5] G.E. Box and G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

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

[7] Manuela Irene Brunner, Anna E. Sikorska, and Jan Seibert. Bivariate analysis of floods in climate impact assessments. *Science of the Total Environment*, 616-617:1392–1403, 2018.

[8] Tiberiu Catalina, Vlad Iordache, and Bogdan Caracaleanu. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*, 57:302–312, 2013.

[9] A. Charpentier, J.-D. Fermanian, and O. Scaillet. *Copulas: From theory to application in finance*, chapter The estimation of copulas: Theory and practice, pages 35–60. Risk books, 2007.

[10] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. John Wiley & Sons Inc., Chichester, West Sussex, 2004.

[11] David Connolly, Brian Vad Mathiesen, Poul Alberg Østergaard, Steffen Nielsen, Urban Persson, and Sven Werner. Heat Roadmap Europe 2050: Second Pre-study for the EU27. *Energy Engineering*, page 236, 2013.

[12] Magnus Dahl, Adam Brun, and Gorm B. Andresen. Using ensemble weather predictions in district heating operation and load forecasting. *Applied Energy*, 193:455–465, may 2017.

[13] Erik Dotzauer. Simple model for prediction of loads in district-heating systems. *Applied Energy*, 73(3-4):277–284, 2002.

[14] Jean Duquette, Andrew Rowe, and Peter Wild. Thermal performance of a steady state physical pipe model for simulating district heating grids with variable flow. *Applied Energy*, 178:383–393, 2016.

[15] F. Durante and C. Sempi. *Principles of Copula Theory*. CRC Press, Boca Raton, 2015.

[16] Frédéric Ferraty, Aldo Goia, Ernesto Salinelli, and Philippe Vieu. Peak-Load Forecasting Using a Functional Semi-Parametric Approach. In *Topics in Nonparametric Statistics*, volume 74, pages 105–114, 2014.

[17] C. Genest, K. Ghoudi, and L. P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552, 1995.

[18] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213, 2009.

[19] Genest-C. Ghorbal, N.B. and J. Nešlehová. On the ghoudi, khoudraji, and rivest test for extreme-value dependence. *The Canadian Journal of Statistics*, 37(4):534–552, 2009.

[20] Aldo Goia, Caterina May, and Gianluca Fusai. Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4):700–711, 2010.

[21] S. Grønneberg and N.L. Hjort. The copula information criteria. *Scandinavian Journal of Statistics*, 41:436459, 2014.

[22] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi. Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, 45:141–151, feb 2012.

[23] Zishuo Huang, Hang Yu, Zhenwei Peng, and Mei Zhao. Methods and tools for community energy planning: A review. *Renewable and Sustainable Energy Reviews*, 42:1335–1348, feb 2015.

[24] T. P. Hutchinson and Lai C. D. *Continuous Bivariate Distributions, Emphasising Applications*. Rumsby Scientific Publishing, 1990.

[25] Kosuke Kato, Masatoshi Sakawa, Keiichi Ishimaru, Satoshi Ushiro, and Toshihiro Shibano. Heat load prediction through recurrent neural network in district heating and cooling systems. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 1401–1406. IEEE, 2008.

[26] Georgios Kazas, Enrico Fabrizio, and Marco Perino. Energy demand profile generation with detailed time resolution at an urban district scale: A reference building approach and case study. *Applied Energy*, 193:243–262, may 2017.

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

[28] H. Lund and B. V. Mathiesen. Energy system analysis of 100% renewable energy systems-The case of Denmark in years 2030 and 2050. *Energy*, 34(5):524–531, 2009.

[29] Henrik Lund. Renewable energy strategies for sustainable development. *Energy*, 32(6):912–919, 2007.

[30] Henrik Lund, Sven Werner, Robin Wiltshire, Svend Svendsen, Jan Eric Thorsen, Frede Hvelplund, and Brian Vad Mathiesen. 4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68:1–11, 2014.

[31] Weiwu Ma, Song Fang, Gang Liu, and Ruoyu Zhou. Modeling of district load forecasting for distributed energy system. *Applied Energy*, 204:181–205, 2017.

[32] Vad Mathiesen, Mads Pagh, Niclas Scott, Per Skougaard, and Niclas Scott. *CEESA 100% Renewable Energy Transport Scenarios Towards 2050: Technical Background Report Part 2.* 2014.

[33] R. B. Nelsen. *Introduction to copulas.* New York: Springer, 2006.

[34] A.J. Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

[35] Urban Persson. Current and future prospects for heat recovery from waste in European district heating systems: A literature and data review. *Energy*, 110(August):25–26, 2015.

[36] A. Pircalabu, T. Hvolby, J. Jung, and E. Høg. Joint price and volumetric risk in wind power trading: A copula approach. *Energy Economics*, 62(C):139–154, 2017.

[37] Daniela Popescu, Florina Ungureanu, and Abel Hernández-Guerrero. Simulation models for the analysis of space heat consumption of buildings. *Energy*, 34(10):1447–1453, 2009.

[38] Dario Prando, Francesco Patuzzi, Giovanni Pernigotto, Andrea Gasparella, and Marco Baratieri. Biomass gasification systems for residential application: An integrated simulation approach. *Applied Thermal Engineering*, 71(1):152–160, 2014.

[39] Milan Protić, Shahaboddin Shamshirband, Dalibor Petković, Almas Abbasi, Miss Laiha Mat Kiah, Jawed Akhtar Unar, Ljiljana Živković, and Miomir Raos. Forecasting of consumers heat load in district heating systems using the support vector machine with a discrete wavelet transform algorithm. *Energy*, 87:343–351, jul 2015.

[40] Muhammad Qamar Raza and Abbas Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372, 2015.

[41] Christoph F. Reinhart and Carlos Cerezo Davila. Urban building energy modeling - A review of a nascent field. *Building and Environment*, 97:196–202, 2016.

[42] B. Rémillard and O. Scaillet. Testing for equality between two copulas. journal of multivariate analysis. *The Canadian Journal of Statistics*, 100(3):377–386, 2009.

[43] Behnaz Rezaie and Marc A. Rosen. District heating and cooling: Review of technology and potential enhancements. *Applied Energy*, 93:2–10, 2012.

[44] Carlos Sarmiento, Carlos Valencia, and Raha Akhavan-Tabatabaei. Copula autoregressive methodology for the simulation of wind speed and direction time series. *Journal of Wind Engineering and Industrial Aerodynamics*, 174(January):188–199, 2018.

[45] A. Sklar. Fonctions de répartition à n dimensions et leures marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.

[46] Michael Smith, Karlson Hargroves, Peter Stasinopoulos, Renee Stephens, and Cheryl Desha. Energy transformed: Sustainable energy solutions for climate change mitigation, 2007.

[47] L. Suganthi and Anand A. Samuel. Energy models for demand forecasting - A review. *Renewable and Sustainable Energy Reviews*, 16(2):1223–1240, 2012.

[48] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, 2009.

[49] P. K. Trivedi and D. M. Zimmer. *Copula Modeling: An Introduction for Practitioners*, volume 1. Foundations and Trends in Econometrics, 2005.

[50] Francesca Verrilli, Seshadhri Srinivasan, Giovanni Gambino, Michele Canelli, Mikko Himanka, Carmen Del Vecchio, Maurizio Sasso, and Luigi Glielmo. Model Predictive Control-Based Optimal Operations of District Heating System With Thermal Energy Storage and Flexible Loads. *IEEE Transactions on Automation Science and Engineering*, pages 1–11, 2016.

[51] Hai Wang, Haiying Wang, Haijian Zhou, and Tong Zhu. Modeling and optimization for hydraulic performance design in multi-source district heating with fluctuating renewables. *Energy Conversion and Management*, 156(July 2017):113–129, 2018.

[52] Krzysztof Wojdyga. An influence of weather conditions on heat demand in district heating systems. *energy and buildings*, 40(11):2009–2014, 2008.

[53] Leon Wu, Gail Kaiser, David Solomon, Rebecca Winter, Albert Boulanger, and Roger Anderson. Improving efficiency and reliability of building systems using machine learning and automated online evaluation. In *Systems, Applications and Technology Conference (LISAT), 2012 IEEE Long Island*, pages 1–6. IEEE, 2012.

[54] D.M. Zimmer and P.K. Trivedi. Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics*, 24:6376, 2006.