CLUSTER ANALYSIS AND CONDITIONAL COPULA: A JOINT APPROACH TO ANALYSE ENERGY DEMAND

F. Marta L. Di Lascio[1] and Roberta Pappadà[2]

[1] Faculty of Economics and Management, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy, (e-mail: marta.dilascio@unibz.it)

[2] Department of Economics, Business, Mathematics and Statistics "B. de Finetti", University of Trieste, Italy, (e-mail: rpappada@units.it)

ABSTRACT: In this work we investigate the thermal energy demand (TED) in urban areas through a copula-based approach. The proposed method enables the characterization of the probability law of TED under extreme weather conditions and for specific groups of buildings. In particular, we show how building characteristics, such as energy class and heating surface, may worsen or mitigate the impact of extreme scenarios.

KEYWORDS: Ali-Mikhail-Haq copula, cluster analysis, conditional copula, thermal energy demand

1 Introduction

The analysis of thermal consumption in urban areas is crucial to increase the sustainability and efficiency of energy systems (Menapace *et al.*, 2021) and reduce the impact of climate change. One major issue is the study of the complex dependence between thermal energy demand (TED) and weather conditions. Focusing on district heating (DH) – a heat distribution system representing a key technology to reduce waste of energy in urban areas – Di Lascio *et al.*, 2020 and Di Lascio *et al.*, 2021 analysed the temporal dynamics of TED and its relationships with meteorological variables. In particular, they assessed the effect of extreme values of solar radiation (SR, in W/m^2) and outdoor temperature (OT, in $°C$) on TED by using a conditional copula-based approach. In addition, Di Lascio *et al.*, 202X recently proposed a copula-based dissimilarity measure to group buildings according to their TED, which turns out to be strongly influenced by building characteristics, such as energy class, age class, and heating surface.

In this paper, we refine the study of the impact that meteorological variables have on TED by merging the proposals in Di Lascio *et al.*, 2021 and Di Lascio *et al.*, 202X. Sect. 2 presents the background and the proposed methodology, while Sect. 3 illustrates the application and discusses the results.

2 Methodology

Our proposal is to exploit conditional copula to describe the probability law of TED (X_3) given extreme scenarios (i.e. very low quantiles of SR (X_1) and OT (X_2)) (Di Lascio *et al.*, 2021), after performing a suitable cluster analysis to identify buildings with similar TED profile, as done in Di Lascio *et al.*, 202X.

The proposed methodology is grounded on (i) the following conditional distribution function (see Di Lascio *et al.*, 2021 for details)

$$P(X_3 > x_3 | X_1 < x_1, X_2 < x_2) = 1 - C(F_3(x_3) | F_1(x_1), F_2(x_2)) \qquad (1)$$

where $C(F_3(x_3)|F_1(x_1),F_2(x_2)) = 1 - C(U_3|U_1,U_2)$ is the conditional copula defined using Bayes' rule (Trivedi & Zimmer, 2005), and (ii) the following copula-based spatially-weighted dissimilarity measure between TED time series at different sites (see Di Lascio *et al.*, 202X for details)

$$d_{jj'} = c_{jj'} \sqrt{2(1-\theta_{jj'})} \qquad (2)$$

where $\theta_{jj'} \in [-1,1[$ is the parameter of the Ali-Mikhail-Haq copula model (Ali *et al.*, 1978) of TED time series at sites j and j' , and $c_{jj'} = \exp(g_{jj'}/\max(G))$, $\forall j \neq j'$, where $G = (g_{jj'})$ is matrix of geographical distances, is the spatial weight. The final clustering here is obtained via the hierarchical method with complete linkage rule.

In what follows, an application of the conditional-copula approach to the partition obtained via the clustering procedure based on Eq. (2) is illustrated.

3 Empirical analysis and discussion

We use hourly time series of TED of 41 residential users (i.e., one or more aggregated buildings) in Bozen-Bolzano during two intermediate weeks in January 2016. We first estimate the following dynamic panel regression model

$$\mathrm{TED}_{it} = \rho_1 \mathrm{TED}_{i(t-1)} + \rho_2 \mathrm{TED}_{i(t-24)} + \beta_1 \mathrm{SR}_{it} + \beta_2 \mathrm{OT}_{it} + \beta_3 \mathrm{OT}_{i(t-3)} + \mu_i + \varepsilon_{it}$$

where $i = 1,\ldots,41$, $t = 1,\ldots,T = 366$, $\mu_i \sim \mathrm{iidN}(0,\sigma^2_\mu)$, $\varepsilon_{it} \sim \mathrm{iidN}(0,\sigma^2_\varepsilon)$, with μ_i and ε_{it} independent. Secondly, the hierarchical clustering method with complete linkage rule and dissimilarity in Eq. (2) is applied to the 41 TED residual time series, yielding $K = 3$ clusters of users (selecting K through the average silhouette width). Fig. 1 shows the time invariant characteristics of DH users for the final clusters. Based on the obtained partition, we model the

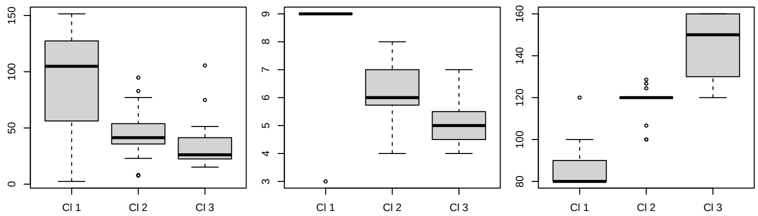

Figure 1. *Heating surface (left panel), age class (middle panel), energy class (right panel) for each cluster, from Cl 1 to Cl 3.*

temporal dynamics of TED aggregated by cluster though a suitable SARIMA model, identified via the AIC and validated by checking for residual autocorrelation ($SARIMA(2,0,0)(1,1,1)$ with a drift for SR, $SARIMA(0,1,2)(2,1,0)$ for OT, and $SARIMA(1,1,2)(0,1,1)$ for TED in the first and third cluster, $SARIMA(1,1,1)(1,1,2)$ for TED in the second cluster). Finally, we model the dependence relationship between each residual series and SR and OT via the conditional copula in Eq. (1), where a parametric copula model is chosen among the Elliptical, the Archimedean and the Joe family (Durante & Sempi, 2015). For all the three clusters, the resulting copula model is the Student-t, which is selected on the basis of the AIC and estimated via maximum likelihood. Fig. 2 shows the behaviour of TED for extreme values (low quantiles) of OT and SR for each identified cluster. While weather conditions have a clear

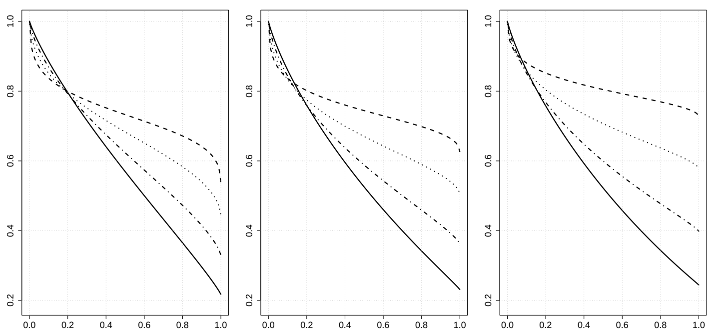

Figure 2. *Copula-based conditional probability function in Eq. (1) with* $(U_1, U_2) < 0.3$ *(solid line),* < 0.15 *(dash-dotted line),* < 0.05 *(dotted line),* < 0.01 *(dash line) for each identified cluster: residual TED quantile (x-axis) (right).*

impact regardless of the cluster considered, it is also evident that TED behaves differently in the three clusters. For instance, the probability of exceeding the 90-th percentile of TED given that SR and OT are smaller than their 0.01-th quantile is 0.671, 0.698, and 0.769 for the first, second and third cluster, respectively. Moreover, the conditional probability increases more quickly as the weather tends to a more extreme scenario for the third cluster in comparison to the other two clusters. Indeed, the first cluster, which includes large, new and efficient buildings, is characterized by the lowest impact of extreme events of TED; the third cluster shows the strongest effect, including old and small buildings, with the lowest energy performance; the second seems to be characterized by an intermediate behaviour (the buildings are small, with medium energy class and heterogeneous age).

These findings can contribute to the study of the impact of meteorological conditions on the energy needs of buildings in the urban area, thus supporting the efficient management and production of thermal energy.

References

ALI, M., MIKHAIL, N.N., & HAQ, M.S. 1978. A class of bivariate distributions including the bivariate logistic. *J. Multivar. Anal.*, **8**(3), 405–412.

DI LASCIO, F.M.L., MENAPACE, A., & RIGHETTI, M. 2020. Joint and conditional dependence modelling of peak district heating demand and outdoor temperature: a copula-based approach. *Stat. Methods Appt.*, **29**, 373–395.

DI LASCIO, F.M.L., MENAPACE, A., & RIGHETTI, M. 2021. Analysing the relationship between district heating demand and weather conditions through conditional mixture copula. *Environ. Ecol. Stat.*, **28(1)**, 53–72.

DI LASCIO, F.M.L., MENAPACE, A., & PAPPADÀ, R. 202X. A spatial AMH copula-based dissimilarity measure to cluster variables in panel data. *BEMPS wp. Under review.*, 1–18. https://econpapers.repec.org/paper/bznwpaper/bemps89.htm.

DURANTE, F., & SEMPI, C. 2015. *Principles of Copula Theory*. Boca Raton: CRC Press.

MENAPACE, ANDREA, SANTOPIETRO, SIMONE, GARGANO, RUDY, & RIGHETTI, MAURIZIO. 2021. Stochastic Generation of District Heat Load. *Energies*, **14**(17).

TRIVEDI, P. K., & ZIMMER, D. M. 2005. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, **1(1)**, 1–111.