Monitoring Electricity Demand Synchronization Using Copulas

Tobias Gebhard
Institute for the Protection of Terrestrial Infrastructures
German Aerospace Center (DLR)
Darmstadt, Germany
tobias.gebhard@dlr.de

Eva Brucherseifer
Institute for the Protection of Terrestrial Infrastructures
German Aerospace Center (DLR)
Sankt Augustin, Germany
eva.brucherseifer@dlr.de

Florian Steinke
Energy Information Networks & Systems
Technische Universität Darmstadt
Darmstadt, Germany
florian.steinke@eins.tu-darmstadt.de

***Abstract*—Synchronization of the behavior of residential consumers, for example during crises, can lead to overloads in electric power grids. This holds especially for distribution grids, where the electrical infrastructure is not designed for the simultaneous high consumption of all households. Therefore, the monitoring and detection of (upcoming) synchronization trends is important. It is the basis for any countermeasures. We propose to model the dependency structure of consumer demands with a Gaussian copula using its correlation parameter as an indicator for synchronization. We then analyze the probability distribution of the aggregated load depending on the synchronization indicator. This allows us to infer the synchronization parameter from load measurements in real-time using a Bayesian approach. In simulation experiments with realistic household consumption distributions, we show how increased synchronization can be detected.**

***Index Terms*—Synchronization, Copulas, Electricity demand, Resilience**

I. Introduction

Electrical power grids need to supply consumers with electric power at all times. The aggregated demand of many consumers typically stays within a narrow band because short-term demand peaks of individual consumers typically balance each other out. However, during crises such as extreme weather events, pandemics, or social conflicts this may no longer be the case. Crises often lead to the synchronization of people's behavior which also affects the power demand. While the average energy consumption pattern over a longer horizon may not change significantly, since the connected devices and human needs remain similar overall, the timing of the individual demands might change severely – and thus the aggregated peak load.

A temporarily high aggregated power demand can lead to local power outages, even if it can be served by the transmission grid. This is because local grid capacity limits of transformers and power lines might be exceeded. Various household devices produce short and high power peaks, e. g., electric kettles, stoves, or tankless water heaters. This explains the high maximum power ratings for individual residential buildings, e. g., 14.5 kVA without electric heating in Germany [1]. If the temporal usage of these devices synchronizes, the grid's power limits may be reached very quickly. Electric vehicle (EV) charging is another important factor to consider as it features comparably large loads for minutes to hours. If charging does not occur equally distributed over the day but concentrated in a short time window, low-voltage (LV) grids might become overloaded already for a low number of EVs [2].

Power grids are socio-technical systems since the behavior of people has a major influence on the state of the grid. If unusual events lead to spontaneous synchronization of people's behavior, the typical assumption of (almost) independent consumers becomes invalid. For example, in the United Kingdom, the TV program sometimes causes such a synchronization event, when at the end of one show, many viewers use an electric kettle. This can result in a rise of the total power demand of up to 2800 MW in only a few minutes [3]. During crises, people often change their behavior and therefore possibly their use of electrical power. These changes tend to happen in a synchronized fashion when people are affected by the same causes or supply situations. For example, in the corona crisis, a self-reinforcing increase in demand for toilet paper could be observed [4]. Another scenario is a chemical accident or a war that forces people to flee and potentially charge their electric vehicles shortly before. Further, in today's digital world, the spreading of information over social networks could lead to a synchronization of people and affect the power grid in timescales of minutes. Note that crises can lead to critical effects in power grids, even if the infrastructure itself is fully intact.

The early detection of these synchronization events and a short-term forecast of their strength would be beneficial for grid operators as it would allow them to counteract, e. g., by sending price signals [5] or regulating individual loads via smart meters [6]. This requires a model and an estimation framework that is capable of representing (time-)variable synchronization.

In this paper, we propose an electricity demand model that allows to represent the individual demands of different prosumers and a time-variable dependence structure between them. To this end, we use a Gaussian copula and define its correlation parameter as our synchronization indicator. We explore the resulting distribution of the aggregated load

This work has been performed in the context of the LOEWE center emergenCITY.

and show how Bayesian inference can be used to determine the posterior probability distribution of the synchronization indicator given real-time measurements of the aggregated load.

The presented framework can be used as a basis for real-time monitoring of demand synchronization. Grid operators can then be warned about upcoming critical load situations and counteract before blackouts happen. Thus, our approach is an important tool to increase disaster preparedness and resilience of power systems.

Section II reviews existing methods for electricity demand modeling. In Section III our demand synchronization model is introduced. A simulation study of our approach is presented in Section IV. We conclude in Section V.

II. Related Work

To describe synchronization effects caused by households, a *bottom-up model* [7] for power consumption/generation is needed to be able to simulate the individual behavior of prosumers. Electricity demand is often modeled with probability distributions to represent uncertainty in the individual demand behavior. For residential households, typical distributions used are Beta, Weibull, Log-normal or Log-logistic distributions [8]–[10]. Human time use is the most important variable for explaining the temporal variations of energy demand of households [11] and can be the cause of synchronization. Time-dependence of human practices can be analyzed with activity-based models [11], [12].

When analyzing the distribution of the aggregated load, the relative standard deviation typically reduces for a greater number of households [13]. The aggregated load is therefore often assumed as the sum of independent random variables [9]. The resulting distribution then tends towards a normal distribution according to the central limit theorem. However, analysis of aggregated residential load shows that it is not represented satisfactorily by a normal distribution [13], indicating that consumers cannot be assumed to be independent.

In the design of LV grids, *diversity factors* are commonly used to describe the simultaneousness of power demand. The diversity factor is defined as the ratio of the sum of the individual peak loads and the maximum of the aggregated load [14], [15]. They can be estimated for different types of devices and are typically derived from empirical knowledge.

Various models consider the influence of day time, weekday and season on demand. These variations are commonly captured in standard load profiles (SLPs) and describe an "external synchronization" that leads to a partial dependence of consumers (e. g., at night, all consumers tend to have a lower demand). For every hour of the day, day of the week and month of the year, a different distribution of individual demand can be assumed [9]. Despite this temporal interdependence, individual consumers are often assumed independent from each other.

In the field of load forecasting, the timeframe "very short term" (less than an hour) is rarely investigated [16]. Little is known about high load peaks in LV grids in intervals shorter than 15 minutes [10], [15]. Energy demand models are often fitted with real-time data from smart meters [9]–[11]. In order to assess the temporal variance of power consumption, the considered timescale is crucial. Existing models often use data with hourly intervals which is able to reflect the rough daily variations, but minute resolution is needed to capture peaks created by devices that require high power only for a short time.

Copulas have recently been investigated for EV charging models [17] as well as short-term load forecasting regarding external dependencies of power load on electricity price and temperature [18]. However, a (time-)variable internal dependence between individual consumers and its real-time monitoring are not examined.

III. Demand Synchronization Model

Our prosumer synchronization model consists of two parts. First, a demand model for the individual households is needed. Second, a dependence model based on copulas describes the (partial) synchronization between prosumers.

A. Individual Demand Model

We model the individual power demand of household i as a random variable X_i with a corresponding distribution that has a probability density function (PDF) $f_i(x)$ and cumulative distribution function (CDF) $F_i(x)$. As we consider only consumers without local generation, the distribution is 0 for $x < 0$. The shape of the distribution and its parameters are deduced from empirical household data in Section IV-A.

We use one fixed distribution for the individual demand. Daily or seasonal variations could be incorporated by adjusting the mean (and variance) of known SLPs to the distribution.

B. Dependence Modeling with Copulas

Copulas provide a way of modeling the dependency structure between random variables in a unit space. The advantage of copulas for the modeling of multivariate distributions is that the dependency structure can be modeled separately from the marginal distributions (which can be continuous or discrete). There are several variants of copulas, e.g. Gaussian copulas, Archimedean copulas or t-copulas [19].

Let unit variables U_i with uniform marginals be defined as

$$U_i = F_i(X_i). \tag{1}$$

Then, the copula is the multivariate joint CDF of these unit variables, i.e.,

$$C(u_1, \ldots, u_n) = P(U_1 \leq u_1, \ldots, U_n \leq u_n). \tag{2}$$

It captures the dependency structure between the X_i in a normalized way (here, P denotes the joint probability).

The Gaussian copula is a widely used choice in many applications. Its copula function

$$C_R^{\text{Gauss}}(u) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right) \tag{3}$$

is defined by the CDF Φ_R of a multivariate normal random variable with mean zero and correlation matrix $R \in [-1, 1]^{n \times n}$ and Φ^{-1} the inverse CDF of the univariate standard normal

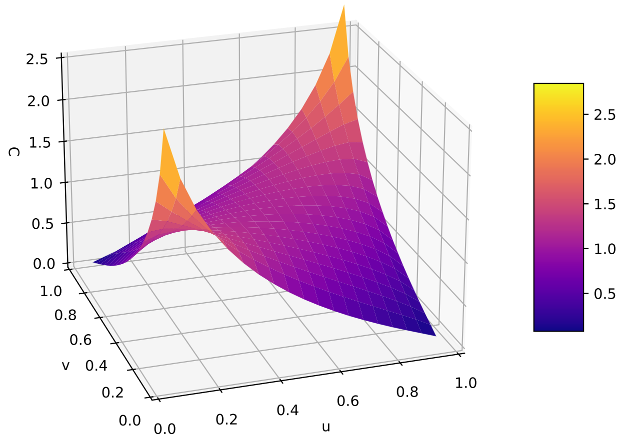


Fig. 1. Two-dimensional Gaussian copula density function with $r = 0.5$

distribution. A correlation between all variables can be modeled with a single parameter r by setting

$$R = \begin{pmatrix} 1 & r & \cdots & r \\ r & 1 & & r \\ \vdots & & \ddots & \vdots \\ r & r & \cdots & 1 \end{pmatrix} . \tag{4}$$

A value of $r = 0$ means all random variables are independent from each other. The other extreme case $r = 1$ indicates a maximal dependence of all variables, i. e. $u_1 = \cdots = u_n$. The density function of a two-dimensional Gaussian copula with correlation $r = 0.5$ is shown in Fig. 1.

From now on, we call $r \in [0, 1]$ the synchronization (indicator).

C. Aggregated Load

Next, we evaluate the aggregated load of many consumers,

$$Y_r = \sum_{i=1}^{n} X_i \tag{5}$$

and its resulting distribution. In a first step, we consider identically distributed marginals (all X_i are distributed as X).

There are two extreme cases. If all consumers are independent from each other, it is very unlikely that the total consumption is very high or low. According to the central limit theorem, the sum converges to a normal distribution for $n \to \infty$ for any distribution of X if the variance of X is finite [20]. It has the following expectation value and standard deviation:

$$\mathrm{E}(Y_0) = n\,\mathrm{E}(X) \tag{6a}$$

$$\mathrm{Std}(Y_0) = \sqrt{n}\,\mathrm{Std}(X) \tag{6b}$$

On the other hand, if all consumers are fully synchronized, the distribution of the sum is the same as the marginal distribution, but scaled with n

$$Y_1 = nX . \tag{7}$$

For $0 < r < 1$ the distribution of Y_r is somewhere between those extreme cases. Note that Y_r has the same mean regardless of r.

D. Inference

We aim to infer the synchronization parameter r from measured values of the aggregated load y. Such measurements could be conducted at a (secondary) substation.

The copula model yields the probability distribution of the aggregated load $p(y|r)$ for a fixed value r. If a prior $p(r)$ is chosen, we can calculate the probability of synchronization r for given y using Bayes' rule as

$$p(r|y) \propto p(y|r) \cdot p(r) . \tag{8}$$

IV. Simulation Results

The presented copula model is simulated with distributions that are derived from typical residential consumption data. Additionally, an EV charging scenario is investigated. Before we analyze synchronization for many consumers, the joint distribution is visualized in a minimal example for the bivariate case.

A. Individual Household Demand

For the derivation of a probability distribution describing household power demand, a publicly available dataset of residential household data [21] is investigated. This dataset contains measured timeseries data from six residential buildings in southern Germany. The measurements were conducted in 1-minute intervals over a period of several years ranging from 2015 to 2019.

We analyze the consumption of `Residential2`, according to the description a "residential building, located in the suburban area". The average consumption is 2493 kWh/year, which corresponds to 0.285 kW on average. The standard deviation is 0.47 kW. Because of the high temporal resolution of the data, power peaks of up to 14 kW can be detected.

Three different distributions, fitted via maximum likelihood estimation (MLE), were compared with the histogram of power consumption (see Fig. 2). The Beta distribution is capable of modeling the part at low power (0-1 kW) which entails most of the distribution, but decreases very fast as it does not have a *fat tail* and therefore cannot represent high values of power. The Weibull distribution better represents data at high values, however, it is not suitable for low values as it tends to infinity for zero power. The Log-normal distribution can adapt to both parts well and is therefore suitable to represent the power demand of a household.

The PDF of the Log-normal distribution is defined as

$$f_{\text{Lognormal}}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) . \tag{9}$$

To adapt the parameters of the distribution to different kinds of consumers, the parameters are calculated such that the mean and the variance of the data and the Log-normal distribution match.

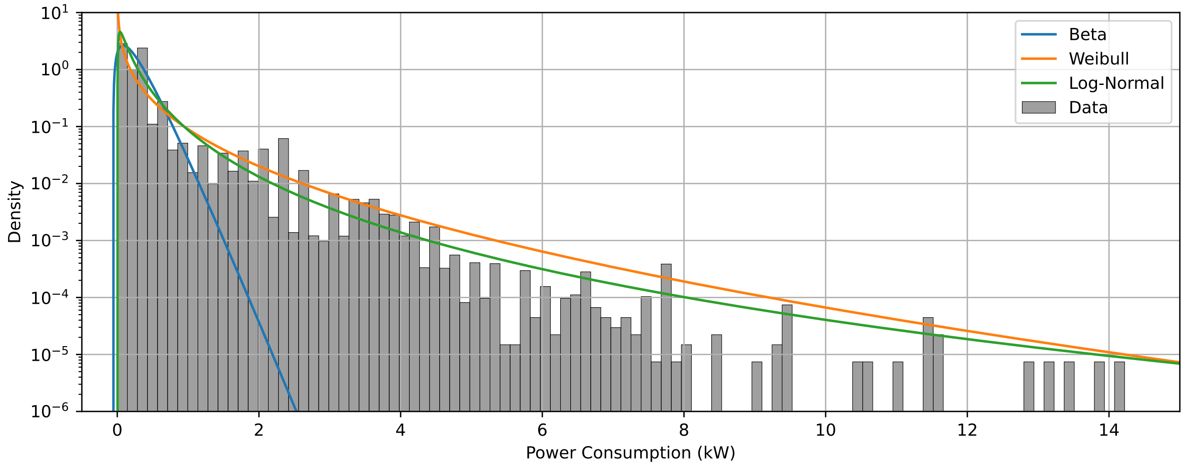


Fig. 2. Histogram of power consumption of a residential household and fitted distributions. The density in logarithmic scale reveals rarely occurring high demands.

B. Electric Vehicle Charging

In another scenario, we assume that 20 % of the households use EV home-charging. Every owner of an EV is assumed to charge one hour per day with a charging power of $11\,\mathrm{kW}$. This can be modeled with a discrete probability distribution:

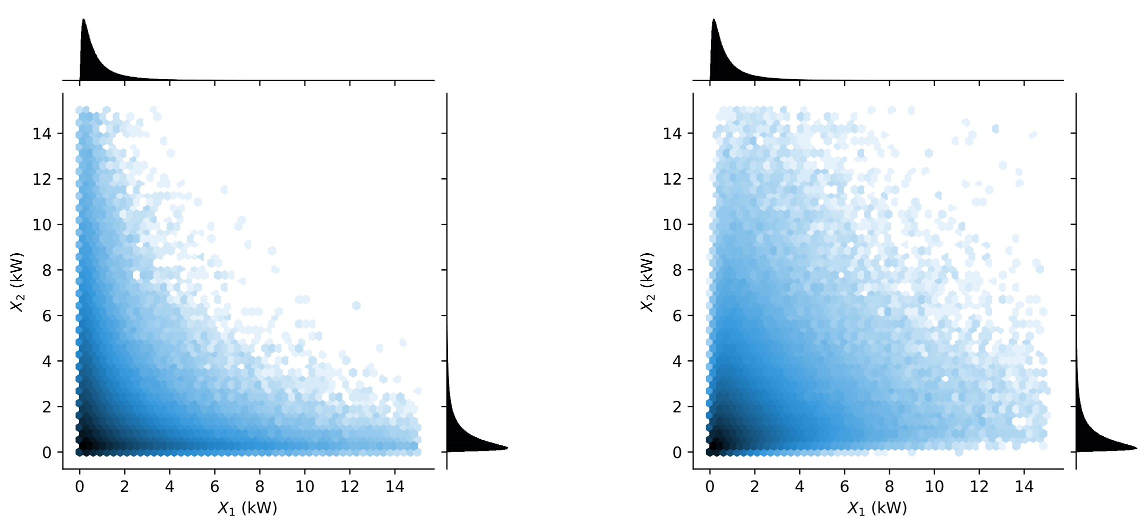
$$P(X = 0\,\mathrm{kW}) = \frac{23}{24}\ ,\ P(X = 11\,\mathrm{kW}) = \frac{1}{24}\ . \tag{10}$$

C. Joint Distribution for 2 Consumers

As a minimal example, the Gaussian copula described in Section III-B is applied for $n = 2$ consumers. Their power consumption X_1 and X_2 is assumed to be distributed according to the described Log-normal distribution. Fig. 3 shows a histogram of the resulting joint distribution. It can be concluded that the probability of both power consumptions being high, i. e., total power being unusually high, increases with higher synchronization r.

D. Aggregated Load for n Consumers

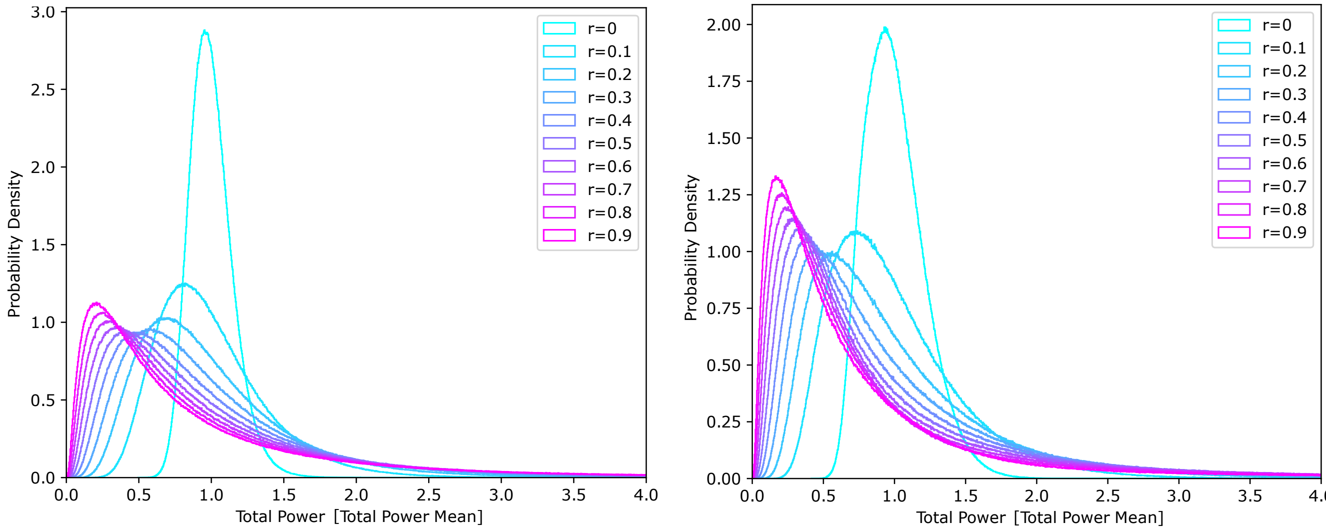
Now we analyze the distribution of the aggregated load for higher n. The power consumption of all $n = 100$ households is assumed to be identically distributed with the Log-normal distribution (9) with $0.5\,\mathrm{kW}$ mean and $0.75\,\mathrm{kW}$ standard deviation. The PDF f_{Y_r} of total power consumption Y_r (5) is calculated in dependence of r with Monte-Carlo-Simulations (10^7 samples). Fig. 4a shows a continuous transition from independence ($r = 0$), resulting in an approximate normal distribution (mean $50\,\mathrm{kW}$), to full synchronization ($r = 1$),



(a) $r = 0$ (Independence)

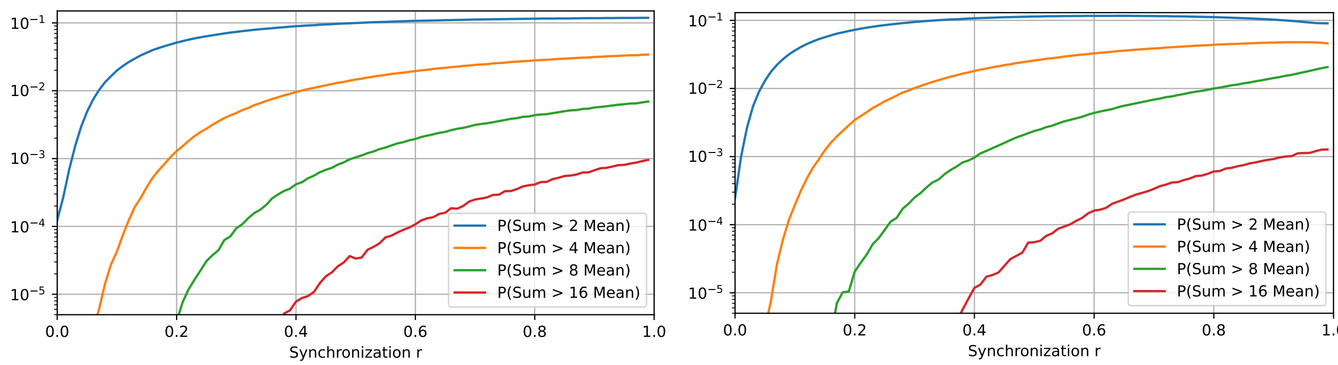
(b) $r = 0.5$

Fig. 3. Histogram of Gaussian copula with $n = 2$ Log-normal distributions ($1\,\mathrm{kW}$ mean and $1.5\,\mathrm{kW}$ standard deviation) as marginals. The color of the cells represents the frequency in a logarithmic scale. (a) If X_1 and X_2 are independent, it can be observed that the frequency of X_1 and X_2 being high at the same time is very low. (b) If there exists a synchronization between X_1 and X_2 with $r = 0.5$, the upper right area that represents total consumption being high is more densely filled. Further, it can be seen that the occurrence of X_1 being low and X_2 being high at the same time (and vice versa) decreases.



(a) 100 Households (b) 100 Households + 20 EVs

Fig. 4. Probability distribution of the aggregated load Y_r with synchronization $r \in [0, 1)$. The total power is normalized to the mean.



(a) 100 Households (b) 100 Households + 20 EVs

Fig. 5. Probabilities of total power exceeding the mean by a factor of 2, 4, 8 and 16 in dependence of r

resulting in a scaled Log-normal distribution, as outlined in Section III-C.

Next, we inquire the probability of exceeding a certain limit

$$P(Y_r > y_{\text{limit}}) = \int_{y_{\text{limit}}}^{\infty} f_{Y_r}(x)\mathrm{d}x \tag{11}$$

which could represent physical power line limits or fuses in a transformer. The probability of exceeding different limits is shown in Fig. 5. For $r = 0$ the (approximate) normal distribution tends to zero very quickly, implying that very high values virtually never occur. The distributions for higher r tend towards zero much more slowly because they converge to a Log-normal distribution for $r \to 1$ which exhibits a fat tail. This means that unusually high loads of e. g., 4-8 times the average load (200-400 kW) cannot be excluded. To set the values into perspective, a probability of 10^{-5} corresponds to an average of 5.3 minutes per year.

The results are compared with the EV charging scenario (Section IV-B). Here, the mean amounts to $59.2\,\text{kW}$. Fig. 4b reveals higher variances of the aggregated load for the same values of r. It can be recognized that for e. g. 2 times the average demand, a lower synchronization r is most likely. In Fig. 5b it can be observed that the probabilities of exceeding multiples of the mean are higher overall for the EV scenario. As r is getting closer to 1, the curves are approaching each other.

E. Inference of synchronization indicator

As outlined in Section III-D, measurements of the aggregated load can be used for inference of the synchronization indicator. In case of just one sample y, MLE would return the r of the PDF of Y_r with the highest value at y. As shown in Fig. 4, a bi-modal behavior emerges: For values of y that are around the mean, $r = 0$ (independence) is most likely. However, if y exceeds this area below or above, PDFs of $r > 0$ are maximal and thus more likely. At both ends of the distribution, $r = 1$ (full synchronization) seems most likely.

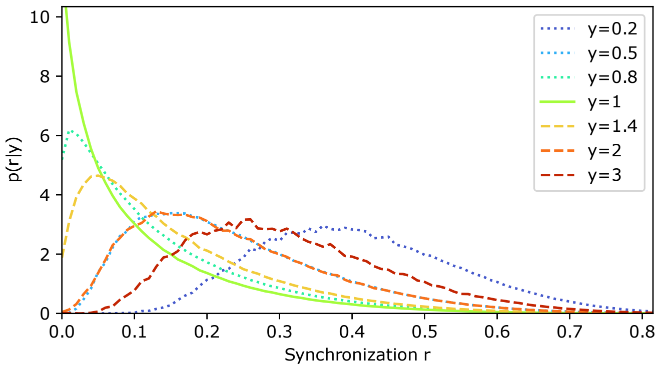


Fig. 6. 100 Households, Posterior probability of r for different measurements of the aggregated load y (measured in normalized units as in Fig. 4)

Via Bayesian inference we can include prior knowledge, since low synchronization is most common and high values of r occur rarely. We therefore assume a Beta distribution $\mathrm{Beta}(1, 5)$ for the prior $p(r)$. Subsequently, we can assess the posterior probability of all $r \in [0, 1]$ for a measurement y, namely $p(r|y)$ (8). Fig. 6 shows the posterior probability of r for different measurements y. The mentioned bi-modal behavior can be detected here as well. While for $y = 1$ the maximum of the posterior yields $r = 0$, for $y < 1$ and $y > 1$ the maximum moves towards higher values of r. Interestingly, $p(r|y)$ yields about the same curve for y and $\frac{1}{y}$.

V. Conclusion

In this work, a novel approach for modeling the synchronization of the electricity demands of residential households is presented. By explicitly modeling the coupling between individual consumers via copulas, internal synchronization effects as well as synchronization induced by external factors can be examined. The strength of synchronization can be encoded in only one parameter. In future, other kinds of correlation matrices or other copulas than the Gaussian could be analyzed.

The advantage of copulas pays off as the synchronization model is separated from the marginal demands – the individual demand can be replaced by any distribution, allowing to include e. g. different types of households, a fraction of EV home-charging, heat pumps, but also photo-voltaic as distributed power generation.

Transferring this theoretical study into industrial tools, the model could be parameterized from standard load profiles (SLPs) or recorded data of the aggregated load. The resulting distributions for the aggregated load can be precalculated in the required resolution. The framework can then be used in a real-time environment to provide estimates for the current state of synchronization. The estimated synchronization indicator could be continuously monitored and in its normalized form serve as a simple early warning indicator.

References

[1] *DIN 18015-1:2007-09 Elektrische Anlagen in Wohngebäuden - Teil 1: Planungsgrundlagen*, Std.

[2] L. Held, A. Märtz, D. Krohn, J. Wirth, M. Zimmerlin, M. R. Suriyah, T. Leibfried, P. Jochem, and W. Fichtner, "The Influence of Electric Vehicle Charging on Low Voltage Grids with Characteristics Typical for Germany," *World Electric Vehicle Journal*, vol. 10, no. 4, p. 88, 2019.

[3] Can you have a big 'switch off'? (visited on 2022-04-14). [Online]. Available: http://news.bbc.co.uk/2/hi/uk_news/magazine/6981356.stm

[4] Coronavirus crisis: experimental data reflect current buying behaviour. (visited on 2022-04-14). [Online]. Available: https://www.destatis.de/EN/Press/2020/04/PE20_146_61.html

[5] J. Carr, A. Brissette, E. Ragaini, and L. Omati, "Managing Smart Grids Using Price Responsive Smart Buildings," *Energy Procedia*, vol. 134, pp. 21–28, Oct. 2017.

[6] S. Bruno, S. Lamonaca, M. L. Scala, G. Rotondo, and U. Stecchi, "Load control through smart-metering on distribution networks," in *2009 IEEE Bucharest PowerTech*, Jun. 2009, pp. 1–8.

[7] H. K. Jacobsen, "Integrating the bottom-up and top-down approach to energy–economy modelling: the case of Denmark," *Energy Economics*, vol. 20, no. 4, pp. 443–461, Sep. 1998.

[8] S. Heunis and R. Herman, "A probabilistic model for residential consumer loads," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 621–625, 2002.

[9] J. Munkhammar, J. Rydén, and J. Widén, "Characterizing probability density distributions for household electricity load profiles from high-resolution electricity use data," *Applied Energy*, vol. 135, pp. 382–390, 2014.

[10] P. Kadurek, W. L. Kling, P. F. Ribeiro, and J. F. G. Cobben, "Electricity demand characterization for analyzing residential LV distribution networks," in *2013 IEEE Grenoble Conference*, 2013, pp. 1–5.

[11] J. Torriti, "A review of time use models of residential electricity demand," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 265–272, 2014.

[12] ——, "Understanding the timing of energy demand through time use data: Time of the day dependence of social practices," *Energy Research & Social Science*, vol. 25, pp. 37–47, 2017.

[13] E. Carpaneto and G. Chicco, "Probability distributions of the aggregated residential load," in *International Conference on Probabilistic Methods Applied to Power Systems*, 2006, pp. 1–6.

[14] A. Sargent, R. Broadwater, J. Thompson, and J. Nazarko, "Estimation of diversity and kwhr-to-peak-kw factors from load research data," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1450–1456, 1994.

[15] F. Provoost and M. van Lumig, "The use of smart meters to improve customer load models," in *International Conference and Exhibition on Electricity Distribution*, 2011.

[16] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustainable Cities and Society*, vol. 35, pp. 257–270, 2017.

[17] J. Einolander and R. Lahdelma, "Multivariate copula procedure for electric vehicle charging event simulation," *Energy*, vol. 238, p. 121718, Jan. 2022.

[18] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 127–136, Apr. 2019.

[19] H. Joe, *Dependence modeling with copulas*. Chapman and Hall/CRC, 2014.

[20] Y. Dodge, *The concise encyclopedia of statistics*. Springer, New York, NY, 2008.

[21] "Data package household data. version 2020-04-15," Open Power System Data, 2020, (Primary data from various sources, for a complete list see URL). [Online]. Available: https://data.open-power-system-data.org/household_data/2020-04-15