Data-Based Modeling of High-Resolution Household Load Profiles

Harald Kirchsteiger and Lukas Gaisberger

ASIC Energy Research, University of Applied Sciences Upper Austria, Wels, Austria

Abstract

A novel method to generate synthetic household load profiles with a sample time of 5 minutes is presented. The method is entirely based on a dataset of measured electricity consumption over 5 consecutive years of 115 dwellings and does not require any other type of information such as household appliances in use or surveys on inhabitants' daily activities. It is shown that the collection of all synthesized signals resembles a typical seasonal variation as observed in the measurement data. At the same time, the individual load signals differ significantly from each other and therefore enable a realistic dispersion of a residential area. None of the synthesized signals is present as such in the measurement data, but all of them lie within the variability observed. A variety of different load signals of variable length can be generated which makes the method particularly interesting for large scale simulations of energy systems including large residential areas.

Keywords: load profile generator, synthetic load profile, smart grid, grid simulation

1. Introduction

Model-based development methods are state of the art in many engineering disciplines, especially also in power- and solar energy systems. Detailed mathematical models of various components in modern renewable energy systems enable sophisticated planning, design, and operation of future multi-energy grids (Martínez Ceseña et al. 2020). The significance of simulation results not only depends on the accuracy of the models, but also to a large extent on the numerical data used to feed the simulation. In this paper, the focus is on power- and energy flow simulation models for medium- to large geographic regions involving many individual consumers.

One particular requirement in such a situation is to feed the simulation with realistic electricity consumption data of individual households. For some cases, aggregating the load of a region to a single load profile (Meier et al. 1999) is a reasonable choice, for example when the interest lies in the characteristics of the transmission system. In other cases, load profiles of individual facilities have to be considered, for example when analyzing the load flow at lower grid voltage levels such as the 400V lines in Europe. This is getting even more important against the background of increasing shares of distributed solar generators at individual households, associated electricity storage devices and new consumer loads such as electric vehicles since the load variability increases (Schinke and Hirsch 2019).

Although the availability of measured individual household load profiles has kept increasing over the years, there is a need for synthetic load profile generators, which are able to reproduce realistic, original load traces in large numbers. Furthermore, a strong requirement is also the temporal resolution of the data. Only fast enough sampling rates enable the accurate consideration of energy management systems (Kirchsteiger and Steinmaurer 2020). In this paper, an algorithm to generate synthetic load profiles with a 5-minute sample time will be presented.

The problem of synthetic load data generation was considered in the literature before. Richardson et al. (2010) describe an algorithm to generate active and reactive electrical load data with a 1-minute sample time. The method relies on modeling the activities of occupants which are linked to specific appliances. Those activities, for example cooking or ironing, were gathered as part of a survey, where participants were asked what they do for how long on a typical weekend-day and weekday. This type of information is frequently called time use data (TUD). The model was validated using measurement data from 22 dwellings.

In Baetens et al. (2012) a statistical model based on similar principles as in Richardson et al. (2010) was extended with thermal demands. The combined load model was used to represent 33 dwellings in the IEEE 34 radial distribution node test feeder, designed specifically to represent residential areas.

doi:10.18086/eurosun.2022.16.06 Available at http://proceedings.ises.org

In Good et al. (2015), a demand-driven load model for both electrical and thermal demands is constructed with the aim to correctly represent the diversity observed in real dwellings and the coincidence observed within groups of buildings at the same time. Also in this work, the modeling of the electrical load relies in parts on the work of Richardson et al. (2010). For the thermal loads, the electrical analogy of R-C type of networks is used.

In Yamaguchi et al. (2019) various known approaches from literature for load generation were reviewed, re-implemented and compared against each other. A categorization of algorithms is provided, including the category of “empirical data-based time-dependent switch-on probability models”. Three other categories identified all rely on TUD.

Lombardi et al. (2019) propose a TUD based method specifically to generate load profiles of remote areas. The model is available as open-source program coded in Python.

Pflugradt and Muntwyler (2017) describe an approach to load modeling which depends on first modeling the individual psychological desires of inhabitants, for example hunger, sleepiness, entertainment. This in turn relates to activities which the persons consider to fulfill their desires, such as eat, cook a meal, sleep, take a nap, or watch TV. The method does not rely on data and follows a truly bottom-up approach. The model is made freely available by the authors.

In Paatero and Lund (2006) datasets from 702 dwellings over 1 year and 1082 dwellings over 143 days were utilized to construct new profiles on an hourly basis. The synthesized data embeds a seasonal, hourly and social variability factor which are all derived from the datasets.

Gruber et al. (2014) calculate load profiles based on a wide range of individual household appliances and utilizing a probabilistic approach. Each appliance is assigned a specific load pattern, the synthesized data has a sample time of 15 minutes.

To conclude on the literature research, most load data models follow a so-called bottom-up approach where modeling starts by defining the individual appliances in the houses and daily patterns of the inhabitants. Those choices are typically supported by large-scale public surveys where persons are asked to note which appliance they are using at which time of the day. Overall, a rather large amount of additional information is required to parameterize the models. Moreover, the surveys typically only reflect a fragment of the whole variety in usage of home appliances. Only few methods such as Paatero and Lund (2006) are known where bottom-up models are also making use of measurement data (household load data timeseries), which decreases the amount of additional information required.

The contribution and novelty of the present paper is as follows. We are presenting a novel method to synthesize individual household electrical load data with a sample time of 5 minutes, which is a finer resolution than known methods from the literature, and a requirement for accurate optimization-based load management algorithms. The proposed method works without the use of any other additional information except a collection of historic load data profiles. Specifically, no information on household appliances and inhabitants’ behavior is required. The individual profiles generated have a variation as observed in the training dataset. When large amounts of synthesized profiles are combined, the sum resembles the typical seasonal variability observed in real data. A variety of dwelling types in the training dataset, for example with and without photovoltaic system, battery storage, heat pump heating, ensures a realistic dispersion of synthesized data.

2. Problem Formulation

The problem considered is to develop an algorithm, which allows to construct new, original electrical load profiles with a high sample time. The new profiles should be different than all profiles in the training dataset, but at the same time, their combination has to agree with the combined training dataset. The problem is separated into two subproblems: (1) Data analysis: Find and extract the relevant statistical properties from the training dataset. In this step, the total daily energy (TDE) consumed is considered, which is known to exhibit a seasonal variation that has to be captured by the algorithm. (2) Data synthesis: use the discovered statistical properties to synthesize new data of household load consumption TDE. As part of this step, the new TDE profiles have to be merged with high resolution data to obtain the required 5 minute sample time signals.

3. Measurement Data

The data used in the present paper has been collected over five years by owners of photovoltaic systems with battery storage in the region of Upper Austria in the course of the accompanying research of a funding programme. The measurement data was received from the hybrid inverters which, in order to control the battery power in accordance with the consumption, are always equipped with an energy meter. This metering device is installed directly next to the meter of the power supply company such that both measure precisely the same values. Most inverters provide the power values in mean values with a temporal resolution of 5 minutes, however up to 15 minutes resolution was accepted. The various power measurement channels are shown in Fig. 1, though only P_L has been used in this analysis. The size of the households differed between two-person households to farmhouses where three generations are living, however single-family homes presented the largest proportion. The age and number of residents was diverse as well. The range varied from singles and young couples over families with 3 generations to older couples. In total, 191 houses monitored over 5 years resulted in 348.766 available days of load traces for analysis.

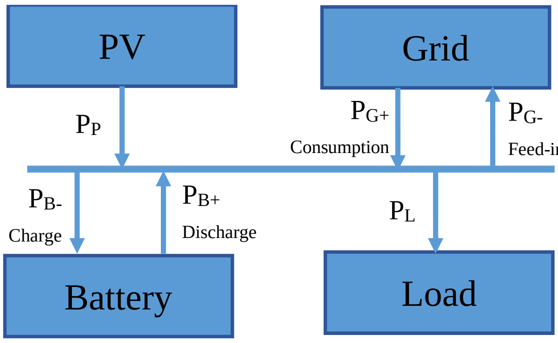

Fig. 1: Measurement data configuration

During this extensive period of measurement some issues regarding data quality arose. Aside from simply missing values, false values represented a major concern. Fortunately, most of the false values could be traced back to miscalculations in the raw data obtained from some inverters. These errors appear only in a limited time range. Datasets with more than 20 % of missing or false data have been removed from the evaluations, resulting in a dataset containing 115 households for further analysis as described below.

4. Methods

In this section the analysis of the total daily energy (TDE) profiles to derive at the characteristic values required for synthesis of new TDE profiles is described. The new TDE profiles are then used to construct high resolution load profiles.

4.1 Analysis of TDE profiles

From the dataset as described in chapter 3, $n_h = 115$ power consumption traces with a sample time of $\Delta T = 5$ minutes are available which will be denoted with $P(h, k)$ with the discrete time index $k = n\,\Delta T, n \in \mathbb{N}$ and h denoting the household index, $h = \{1,2, \dots n_h\}$. From this, the TDE is derived for every household by simply summing up $P(h, k)$ for single days. The TDE is denoted as $E_{TD}(h, d)$ where d is again a discrete time index and represents the day of the year, $d = 1, 2, n_d$. (For simplicity of exposition we are assuming only years with 365 days, i.e., without leap days.) The TDE signals are then normalized by subtracting the mean and dividing by the standard deviation to arrive at zero mean, unit variance signals denoted with $E_{TDN}(h, d)$. From those normalized signals, a seasonal variability signal s_V is derived by taking the mean value across all years and households, and applying a moving average filter of window length 40. This length was empirically found to result in a good trade-off between sufficient data smoothing and preserving the main characteristics in the data. The seasonal variability s_V is shown graphically in Fig. 2: the bold line represents the filtered signal while the thinner line is the mean value without any filtering.

The main algorithm for TDE characterization performs the following operations for every household h separately:

- Compute the individual deviation from a household from the seasonal variability, $e_1 = s_V - E_{TDN}$, and take the average $\bar{e}_1$ with respect to all available years of data.
- Remove the outliers from $\bar{e}_1$, based on the 25th (q_{25}) and 75th (q_{75}) quantile. A datapoint is classified as outlier if it is larger than $q_{75} + 1.5(q_{75} - q_{25})$ or smaller than $q_{25} - 1.5(q_{75} - q_{25})$. The signal without outliers is called $\bar{e}_{1OR}$.
- Filter the signal $\bar{e}_{1OR}$with a moving average filter of window length 14 to obtain $\bar{e}_{1ORF}$.
- Compute the second level deviation from the typical household seasonal deviation, $e_2 = \bar{e}_{1ORF} - \bar{e}_{1OR}$
- Fit a normal distribution $\mathcal{N}(\mu, \sigma)$ to the signal e_2 to obtain the two characteristic parameters μ (expectation) and σ (standard variation).

The moving average filter length was again found empirically to provide a reasonable data smoothing. Note that by performing the computations for all households, n_h different values $\mu(h)$ and $\sigma(h)$ are obtained. In a final step, two Burr distributions are fitted to those data. The probability density function (PDF) of the Burr distribution is given by (eq. 1) and depends on the three characteristic parameters α, c, κ.

$$f_B(x) = \frac{\frac{\kappa c}{\alpha}\left(\frac{x}{\alpha}\right)^{c-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{c}\right)^{\kappa+1}} \qquad \text{(eq. 1)}$$

As a result of the TDE profile analysis, two sets of those characteristic values were obtained, for the distribution $f_B(\mu)$ of $\mu(h)$ and $f_B(\sigma)$ of $\sigma(h)$. The main reason for choosing a Burr distribution was that it provides the closest fit to the data among all the tested PDFs available (The Mathworks Inc., 2021), see also Fig. 6.

4.2 Synthesis of new TDE profiles

The algorithm for synthesis of a new TDE profile follows the steps listed below

- Draw a pair of random values (μ^*, σ^*) from the distribution $f_B(\mu)$ and from $f_B(\sigma)$
- Construct a normal distribution $\mathcal{N}(\mu^*, \sigma^*)$ and draw a random sequence r of length 365 from it. This sequence is a substitute for e_2 with embedded statistical properties of the training dataset.
- Synthesize a new TDE sequence $E_{TDN}^S = s_V - \bar{e}_{1ORF}(h^*) - r$ by choosing a deviation $\bar{e}_{1ORF}$ from a specific (random) household $h = h^*$ as seed.
- Undo the normalization by multiplying with a standard deviation within the limits as observed in $E_{TD}(h^*, d)$ and adding a mean value within the limits as observed in $E_{TD}(h^*, d)$ to obtain the sequence E_{TD}^S.

When this algorithm is used multiple times, an entirely new dataset of individual TDE sequences is constructed, from which in turn a seasonal variability can be computed in the same was as described above for the initial measurement dataset. The seasonal variability of such a new dataset is shown in Fig. 2 where again the bold red line is the result of a moving average filter applied to the thin red line.

4.3 Synthesis of new high resolution load profiles

A synthesized TDE signal E_{TD}^S from the algorithm in section 4.2 describes the total daily energy consumption for a (virtually constructed) household, from which power signals with a sample time ΔT will be constructed next. To do so, for every single day $d = d^*$ in the sequence E_{TD}^S, a day $d = d'$ with a similar TDE is looked up in the original load dataset $E(h^*, d)$. The variable h^* denotes the seed as used in the algorithm in section 4.2. The check for similarity is done by searching for similar TDE values within a ± 5 % interval and randomly choosing one of the candidates within this interval. In case the similarity search is not successful, the closest match is chosen and the data scaled to obtain the desired TDE. Then, the consumption profile of that day d' is used as power consumption on the day d^* in the new sequence.

5. Results and Discussion

As described in the Methods section, a seasonal variability s_V is computed for the original measurement dataset and for a newly generated synthetic dataset. The two profiles are compared against each other in Fig. 2 where the bold lines are filtered versions of the tinner lines. The results shown in the figure were constructed with $n_h = 115$ households for both datasets. In general, a very good agreement between the two profiles is observed indicating that the synthesized load profiles resemble the statistical properties of the measurement dataset. It is of interest to note that the signals are slightly higher at the end of December compared to the beginning of January which reflects the general trend of increased electricity consumption of the society as a whole.

As far as the individual TDE profiles are concerned, Fig. 3 shows two examples of E_{TD} for two different houses ($h = 1$ for the top and $h = 35$ for the bottom panel) from the measurement dataset on the left. On the right side, two synthetic profiles E_{TD}^S based on the very same houses are shown. That is, in the 3rd step of the synthesis algorithm stated above, h^* is 1 in the top panel and 35 in the bottom panel. Even if both signals (top left and top right) are based on the same sequence $\bar{e}_{1ORF}(1)$ they are significantly different from each other as can be seen in the figure. The same statement can be made for the bottom two sub-figures based on $\bar{e}_{1ORF}(35)$ and in fact for all $n_h = 115$ dwellings considered.

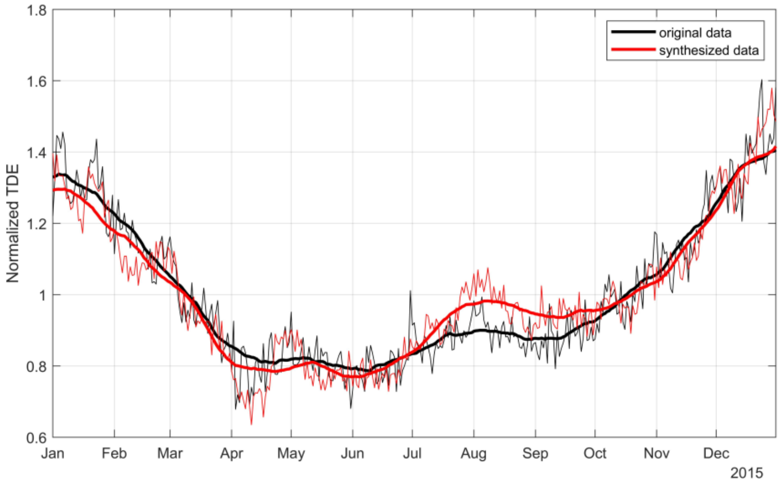

Fig. 2: Seasonal variability of the total daily energy for the measured data (black) and synthesized data (red)

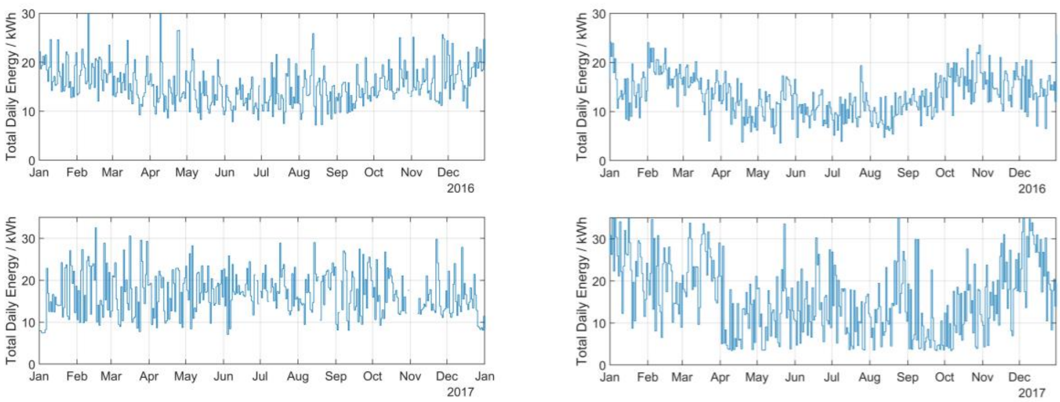

Fig. 3: TDE profiles of two dwellings (measured, left) and synthetic profiles (right)

In Fig. 4, two high resolution electrical load profiles are shown over a timespan of one year. While the top panel displays measured data, the bottom panel shows the result when applying the synthesis strategy described in section 4.3. Again, both signals rely on the same seed $\bar{e}_{1ORF}(h)$. A closer look at the signals, focusing on the month March in Fig. 5, reveals that the synthesized signal at the bottom follows a similar and realistic pattern of nighttime consumption and daytime peaks as the measured data.

To motivate the computations done in steps 3 and 4 of the analysis algorithm stated in section 4.1, consider first the graph shown in Fig. 6 on the left which displays the autocorrelation function of the second level deviations e_2 for the first 20 households. The general shape is close to a white noise random signal with a significant peak only at zero lag. This indicates that there is no more information to be extracted from the data and a probability density function can be fitted as done in step 5 of the algorithm. A detailed look at the autocorrelation functions reveals peaks at lags of 7, 14, and 21 which is in correspondence with the well-known fact of a weekly dependence of the load profiles. Those peaks are, however, below a threshold level of 0.2. The right graphic in Fig. 6 shows the histograms of the parameters μ and σ estimated in the final step of the analysis algorithm. The red lines are showing the PDF of the Burr distribution fitted to the histograms. The PDF $f_B(\mu)$ is given by the parameters $\alpha = 0.5, c = 280,\ \kappa = 1.36$ and the PDF $f_B(\sigma)$ by $\alpha = 0.33, c = 12.9,\ \kappa = 0.48$.

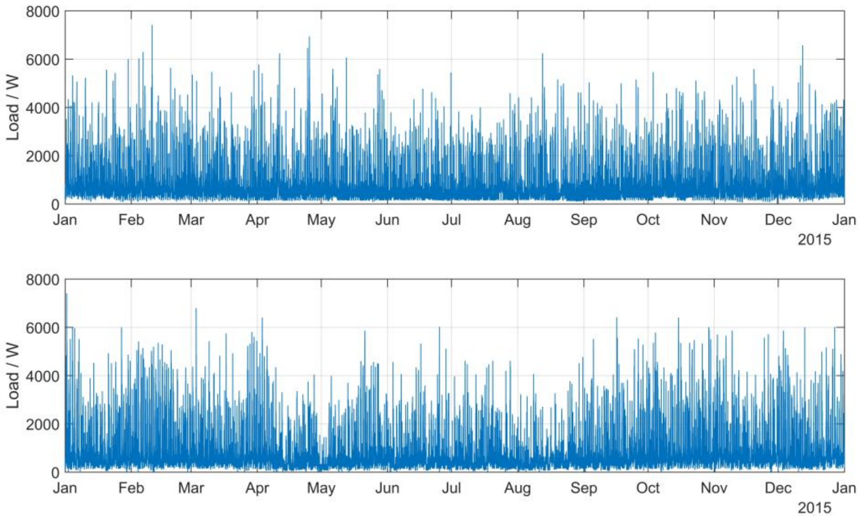

Fig. 4: Power signals $P(h = 1, k)$ (measured, top) and synthesized power signal (bottom)

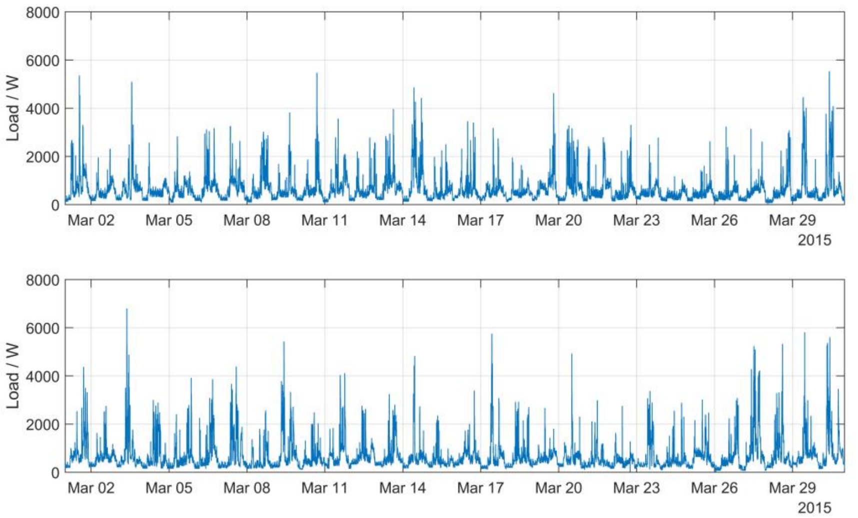

Fig. 5: Detail of Fig. 4 focusing on March: measurement (top), synthesized data (bottom)

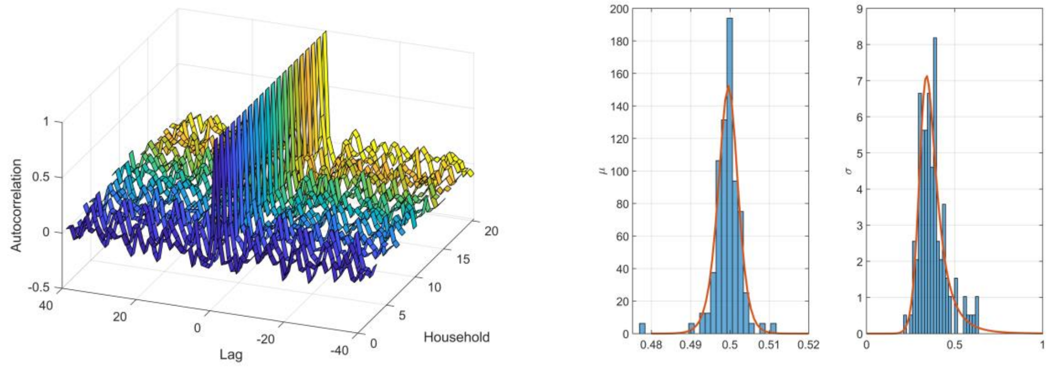

Fig. 6: Second level deviations e_2 (left) and fitted Burr distributions on $\mu(h)$ and $\sigma(h)$ (right)

6. Conclusion

A method to generate synthetic load profiles with a sample time of 5 minutes was presented. The method relies on a measurement dataset of 115 households over 5 years. It was shown that the synthesized datafiles as a whole possess a comparable seasonal variability as the original data while the individual traces are sufficiently diverse. Also on a smaller timescale, the synthesized data has similar patterns as the measurement data, for example day-night consumption changes.

The proposed method for generation of load profiles is well suited for defining individual load profiles of large residential areas in large-scale energy system simulation models. It is also well suited to generate load profiles over an extended time horizon of multiple years since the statistically learned seasonal variation inherently grows over time reflecting the general rise in energy consumption.

The generated load data are, however, not particularly suited for training load prediction algorithms. This is because as stated in section 4.3, on a short timeframe, the new and the original signal are identical, although it is a-priori unknown at which point in time this will be the case. A prediction algorithm could learn the patterns of this short timeframes in the data and correctly forecast them when they are recognized. Such a prediction would be impossible in a real-time environment using measurement data.

7. Acknowledgments

This project is financed by research subsidies granted by the government of Upper Austria. This project is funded by EU-European Regional Development Fund (ERDF) and Federal Province of Upper Austria Investments in Growth and Jobs (IGJ) under the research grant number WI-2020-701794/12-Cz. The work was also supported by the government of Upper Austria in the project Methodenentwicklung zur Energieflussoptimierung (Development of Methods for Optimization of Energy Flows)". Research Grant Wi-2017-450841/17.

8. References

Baetens, R., De Coninck, R., Van Roy, J., Verbruggen, B., Driesen, J., Helsen, L., Saelens, D., 2012. Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation. Applied Energy. Vol. 96, pp. 193-210. https://doi.org/10.1016/j.apenergy.2011.12.098

Good, N., Zhang, L., Navarro-Espinosa, A., Mancarella, P., 2015. High Resolution Modelling of Multi-Energy Domestic Demand Profiles. Applied Energy. Vol. 137, pp. 193-210. https://doi.org/10.1016/j.apenergy.2014.10.028

Gruber J.K., Jahromizadeh, S., Prodanović, M., Rakočević, V., 2014. Application-oriented modelling of domestic energy demand. Electrical Power and Energy Systems. Vol. 61, pp. 656-664. https://doi.org/10.1016/j.ijepes.2014.04.008

Kirchsteiger H., Steinmaurer G., 2020. Optimal Energy Management of Residential Solar PV with Battery

Storage: Effects of Fast Load and Generation Transients. *7th International Conference on Control, Decision and Information Technologies*, Prague, 29 June – 02 July. https://doi.org/10.1109/CoDIT49905.2020.9263853

Lombardi, F., Balderrama, S., Quoilin, S., Colombo, E., 2019. Generating High-Resolution Multi-Energy Load Profiles for Remote Areas with an Open-Source Stochastic Model. Energy. Vol. 177, pp. 433-444. https://doi.org/10.1016/j.energy.2019.04.097

The Mathworks Inc., Statistics and Machine Learning Toolbox, Natick, Massachusetts, United States, 2021. Available at: https://mathworks.com/products/statistics.html

Martínez Ceseña, E.A., Loukarakis, E., Good, N., Mancarella, P., 2020. Integrated Electricity-Heat-Gas Systems: Techno-Economic Modeling, Optimization, and Application to Multienergy Districts. Proceedings of the IEEE. Vol. 108, no.9, pp.1392-1410. https://doi.org/10.1109/JPROC.2020.2989382

Meier, H., Fünfgeld, C., Adam, T., Schieferdecker, B., 1999. Repräsentative VDEW-Lastprofile. VDEW Materialien M-32/99.

Paatero, J.V., Lund P.D., 2006. A model for generating household electricity load profiles. International Journal of Energy Research. Vol. 30, no. 5, pp. 273-290. https://doi.org/10.1002/er.1136

Pflugradt, N., Muntwyler, U., 2017. Synthesizing residential load profiles using behavior simulation. Energy Procedia. Vol. 122, pp. 655-660. https://doi.org/10.1016/j.egypro.2017.07.365

Richardson, I., Thomson, M., Infield, D., Clifford, C., 2010. Domestic electricity use: A high-resolution energy demand model. Energy and Buildings. Vol. 42, no. 10, pp. 1878-1887. https://doi.org/10.1016/j.enbuild.2010.05.023

Schinke A., Hirsch H., 2019. Impact of Electric Vehicle Charging and Photovoltaic Generation on Distribution System Voltage Volatility. *2019 IEEE Power & Energy Society General Meeting*, Atlanta, 04-08 Aug. https://doi.org/10.1109/PESGM40551.2019.8973901

Yamaguchi, Y., Yilmaz, S., Prakash, N., Firth, S.K., Shimoda, Y., 2019. A cross analysis of existing methods for modelling household appliance use. Journal of Building Performance Simulation. Vol. 12, no. 2, pp. 160-179. https://doi.org/10.1080/19401493.2018.1497087