

Feature-conserving gradual anonymization of load profiles and the impact on battery storage systems

Benedikt Tepe^{a,*}, David Haberschusz^{b,c,d}, Jan Figgner^{b,c,d}, Holger Hesse^e, Dirk Uwe Sauer^{b,c,d}, Andreas Jossen^a

^a Technical University of Munich, TUM School of Engineering and Design, Department of Energy and Process Engineering, Chair of Electrical Energy Storage Technology, Germany

^b Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University, Germany

^c Institute for Power Generation and Storage Systems (PGS), E.ON ERC, RWTH Aachen University, Germany

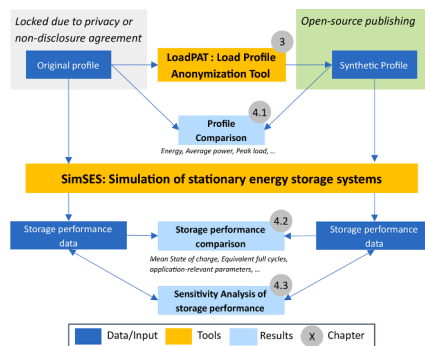
^d Juelich Aachen Research Alliance, JARA-Energy, Germany

^e Kempten University of Applied Sciences, Institute for Energy and Propulsion Technologies, Germany

HIGHLIGHTS

- Methodology for a feature-conserving gradual anonymization of load profiles.
- Application of the methodology to different load profiles and storage applications.
- Analysis of the effects on storage system operation using different KPIs.
- Demonstration of the open-source load profile anonymization tool LoadPAT.
- Methodology allows publishing load profiles similar to protected original profiles.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Open-source electric load profiles
Anonymization
Synthesizing
Household
EV charging station
Battery storage applications

ABSTRACT

Electric load profiles are highly relevant for battery storage research and industry as they determine system design and operation strategies. However, data obtained from electrical load measurements often cannot be shared or published due to privacy concerns. This paper presents a methodology to gradually anonymize load profiles while conforming to various degrees of anonymity. It segregates the original load profile into base and peak sequences and extracts features from each of the sequences. With the help of the features, a synthetic, anonymized load profile is created. Different levels of anonymization can be selected, which transform the original profile to the desired extent. A random permutation of the peak sequences or base sequences is used to achieve this transformation. Exemplary profiles from a household and an electric vehicle charging station are used to demonstrate the functionality of the anonymization. The indicators of the anonymized load profiles are compared with the original ones in both time and frequency domains, and the effects of load profile anonymization on the operation of battery storage systems in two scenarios are analyzed. While the anonymized load profiles retain the time-invariant indicators from the original profile, the permutation causes a loss of regularity

* Corresponding author at: Chair of Electrical Energy Storage Technology, Technical University of Munich (TUM), Arcisstr. 21, 80333 Munich, Germany.
E-mail address: benedikt.tepe@tum.de (B. Tepe).

in the load profiles. As a result, relevant indicators of battery storage systems subjected to these anonymized profiles deviate to a greater extent in time-dependent applications such as self-consumption increase. This is reflected in the overestimation of equivalent full cycles by up to 6% and underestimation of self-sufficiency by up to 9 percentage points. In time-independent applications such as peak shaving, however, the indicators can be well reproduced with deviations of up to 3% despite the lost regularity. In order to make the anonymization methodology usable for everyone, we present the open-source tool LoadPAT, in which users can anonymize their load profiles and choose their desired level of anonymization. This work is intended to further encourage the dissemination of open-source data.

Abbreviations

BSS	Battery storage system
BTM	Behind-the-meter
CDA	Conditional demand analysis
DFT	Discrete Fourier Transformation
EFC	Equivalent full cycles
EMS	Energy-management-strategy
EV	Electric vehicle
FTM	Front-of-the-meter
DOD	Depth of discharge
FCR	Frequency containment reserve
GUI	Graphical user interface
KPI	Key performance indicator
LoadPAT	Load profile anonymization tool
PS	Peak shaving
PV	Photovoltaic
RQ	Research question
SCI	Self-consumption-increase
SimSES	Simulation of stationary energy storage systems
SOC	State-of-charge

1. Introduction

We live in a time where individuals as well as companies and research institutions are constantly generating data. Between 2012 and 2020, the volume of digital data generated each year worldwide increased tenfold from 6.5 to 64.2 zettabytes [1]. Although not all of this data is stored, the data volume of storage capacity in 2021 also amounted to 7.9 zettabytes [2]. In the electricity sector, this development is being driven by the digitalization of grid monitoring and control and the increasing installation of smart meters [3,4]. At the same time, there is a trend towards openness of data, publications and code [5,6]. Research institutions, on the one hand, are interested in sharing collected and generated data and thus making it available to the community. Companies, on the other hand, often have privacy concerns about the possible sharing of data. This conflict gave rise to the present work. The goal is to develop a methodology for anonymizing electric load profiles. We present a load profile anonymization tool (LoadPAT), which can gradually anonymize load profiles, that may aid to a greater usage of sharing data through industry and facilitate an increase of resources available for applied research. In this work, we show the gradual anonymization in various levels that is possible using the tool and the impact on load profile key performance indicators (KPIs). In addition, simulations show the effects of the load profile anonymization on the behavior and load of battery storage systems (BSS) in different applications. With the help of LoadPAT, companies and research institutions can modify and thus anonymize load profiles according to their desired level before sharing with partners.

In the following, the existing literature on the topic of anonymization of load profiles is presented and the research gap is identified.

Subsequently, the research questions of this work are introduced before the scenario and the usage area of the tool are described.

1.1. Summary of existing literature

The topic of load profiles is relevant for research institutions and companies. The former need load profiles, for example, for simulations of the electricity grid or of BSSs. Companies such as distribution network operators make use of load profile data for estimating current and future consumption and for classifying customers [7]. Basically, in research on load profiles, a distinction can be made between load profile analyses including feature extraction and synthetic load profile generation. For the former, Table A1 in the Appendix shows existing literature, which will be described in more detail below.

1.1.1. Load profile analysis and feature extraction

In 2010, Price published methods for analyzing load profiles [8]. He defined five parameters for characterizing load shapes in the time domain: base load, peak load, rise time, high load duration and fall time. Li et al. built on Price's work in 2021 and published a load profile analysis in the time and frequency domains using Discrete Fourier Transformation (DFT) [9]. To do so, they used smart meter data from 188 buildings in Northern California with a time resolution of 15 min. The advantage of frequency domain analysis is that it captures the periodicity of the load profile as a baseline feature while allowing to reduce the amount of data to be stored [9]. DFT was also used by Campestrini et al. to evaluate SOC algorithms [10]. In this work, a number of driving profiles were analyzed to develop a representative synthetic profile in the frequency domain.

When clustering load profiles, a basic distinction can be made between direct clustering of time series profiles and indirect clustering via feature extraction [11]. In 2016, Haben et al. presented an analysis of smart meter data in which they identified the four key time periods overnight, morning, daytime, evening that should be considered for clustering residential load profiles [12]. In the same year, Al-Otaibi et al. published a feature extraction method to cluster daily load profiles based on these features [13]. In addition, clustering was performed after z-normalization, which considers the shape and disregards the magnitude of the profiles. The results showed that compared to using the whole daily load profiles (48 values), extracting features can reduce the dimensions significantly while the clustering is still successful. A similar procedure was used by Park et al. in 2019, who compared the direct clustering after z-normalization using the k-means algorithm with a Gaussian mixture model [14]. In this publication, the k-means algorithm produced better results with a shorter runtime. The same algorithm was used by Trotta 2020 to cluster one-hour annual load profiles of Danish households [15]. The result of this publication are four clusters describing the typical behavior of Danish households. Czétány et al. also used the k-means algorithm to cluster Hungarian households, as k-means was advantageous over a fuzzy k-means and an agglomerative hierarchical clustering [16]. In 2022, Elahe et al. published a new feature extraction technique for load profiles that can be used to identify households with plug-in hybrid vehicles [17]. For this purpose, they used different classifiers that utilize a set of extracted features from the load profile.

Table 1
Summary of literature on synthetic load profile generation and anonymization of load profiles.

Source	Date	Focus	Method	Sampling rate	Results
Aigner et al. [18]	1984	Creation with top-down approach	Conditional demand analysis	15 min	Breakdown of total household load to individual parts using the conditional demand analysis.
Bartels et al. [19]	1992	Creation with top-down approach	Conditional demand analysis	15 min	Publication of DELMOD which uses conditional demand analysis to create household load profiles for different days and seasons.
Pensa et al. [20]	2008	Anonymization of time series	k-anonymity	Arbitrary	Application of k-anonymity to records of sequences, which can be used for all types of records and profiles.
Widén et al. [21]	2010	Creation with bottom-up approach	Stochastic modeling	1 min	Activities of household members are simulated to create high resolution series. A validation with real data showed that the generated household load profiles are realistic.
Richardson et al. [22]	2010	Creation with bottom-up approach	Stochastic modeling	1 min	Creation of one-minute resolution household load profiles based on specific activities and patterns of active occupancy. Validation with field data in East Midlands, UK.
Efthymiou et al. [23]	2010	Anonymization of smart meter data	Escrow mechanism	1–5 min	Development of a method for anonymizing high frequent smart meter data using a third party escrow mechanism.
Ogasawara et al. [24]	2010	Normalization approach for time series	Normalization	Arbitrary	Presentation of ways to normalize all types of load profiles including dividing by maximum value, by using minimum and maximum value, and with adaptive neural networks.
Shouh et al. [25]	2013	Anonymization of time series	(k,P) - Anonymity	Arbitrary	Extension of k-anonymity by P-anonymity to (k, P)-anonymity. This can be used to preserve patterns effectively in addition to the standard k-anonymization.
Jambagi et al. [26]	2015	Creation with bottom-up and top-down approach	Activity based modeling	1 s	Development of a residential electricity demand model combining time use surveys with standard load profiles. A validation with measured data shows that the properties of the synthetic profiles are correct.
Müller et al. [27]	2020	Creation with bottom-up approach	Activity based modeling	1 min	Bottom-up approach model for electrical and thermal household load modelling regarding mobility behavior.
Han et al. [28]	2022	Creation with top-down approach	Decomposition and recombination	1 h	Generation of hourly household electrical load profiles using a statistical method of decomposition and recombination.

1.1.2. Synthetic load profile generation and anonymization

In addition to load profile analysis, clustering, and feature extraction, synthetic load profile generation and anonymization are most relevant to this work. Table 1 gives an overview of literature on synthetic load profile generation and anonymization. While some of the papers presented describe the generation and anonymization of general datasets and sequences, others deal explicitly with household load profiles. Basically, load profiles can be created using bottom-up and top-down methods [29]. The former use behavior and data from individual consumers to generate a load profile. The advantage of bottom-up methods is that user behavior can be aggregated, and thus general statements can be made. The disadvantage, however, is that user data must be available in sufficient quantity. For example, Widén et al. and Richardson et al. used household member activities to generate high-resolution load time series [21,22]. Müller et al. extended the idea of electric household behavior to thermal and mobility behavior [27]. Another bottom-up approach was presented by Li et al. in 2021, which can be used to generate synthetic load profiles for households, businesses, and industries based on the geographic location [30]. They use publicly available data and make three types of transformations: Temporal shifts, temporal permutations, and adding noise. Through these transformations, the authors perform a kind of anonymization without explicitly naming it anonymization. Top-down approaches do not use data from individual electrical appliances or persons, but aggregated data, such as national data, and break them down to individual households [29]. The advantage of these approaches is that no individual user data is required. The disadvantage is that it is not possible to draw conclusions about individual user behavior. As early as 1984, Aigner et al. published a study that made it possible to break down total household load to individual parts [18]. For this purpose, the authors used the conditional demand analysis (CDA) which uses various regression equations to account for factors such as desired temperature and size of the residence when creating load profiles [29]. The advantage of the methodology at that time was the significantly lower effort compared to direct measurements at the end user [18]. In 1992, Bartels et al. presented DELMOD, a model that also uses CDA to generate load profiles for different types of days for a given scenario [19]. This model

uses typical load profiles for different days and seasons and weather data that change the load profile. DELMOD has the advantage that detailed scenarios can be investigated and predictions can be made. In 2015, Jambagi et al. combined Richardson's approach (time use surveys) with a top-down approach of a standard load profile, making the aggregate results more realistic [26]. To generate hourly electrical load time series, Han et al. used statistical methods in 2022 [28]. In their top-down approach, they used three components of public data: seasonality, distribution of residuals and the trend. A different approach to the generation of synthetic load profiles was chosen by Pinceti et al [31]. They used generative adversarial networks to learn from hourly resolved real weekly load profiles and to generate synthetic load profiles. The disadvantage of this approach is that the training of the network takes a relatively long time. In contrast, the advantage is that the network can be used easily and quickly after training.

Regarding anonymization of load profiles, Efthymiou et al. have published a paper in which they present a method for anonymizing smart meter data [23]. For this purpose, they use an escrow service that aggregates data from different households. This method does not focus on the modification and adaptation of time series-based load profiles, but rather on the process of transmitting high-frequency data via a data aggregator. Focus on the anonymization of temporal sequences put, for example, Pensa et al., who use k-anonymization for this purpose. [20]. The k-anonymization can be applied to data sets and describes that the information of an individual cannot be distinguished from at least k-1 other individuals [32]. Pensa et al. applied this type of anonymization to datasets of sequences achieved k-anonymization of the data [20]. Shouh et al. extended the k-anonymization approach to another level, P-anonymization, which represents patterns within grouped time-series datasets [25]. This (k, P)-anonymization was shown in experiments to be resistant to linkage attacks while preserving pattern data. The concept of k-anonymization was extended by Machanavajjhala et al. to include l-diversity, which requires that for each sensitive attribute in data sets, at least l different attributes must occur [33]. A combination of k-anonymization and l-diversity consequently enables an anonymization that is more secure against attacks.

The presented variants of the anonymization of data sets have in

common that several load profiles must exist. However, the focus of the present work is on the anonymization of individual time series-based load profiles, which means that the presented variants are not applicable. For the modification of individual load profiles, Savov et al. published a paper in 2017 in which they analyzed the degree to which load profiles can be discretized [34]. The result was that the greater the discretization of the data, the greater the error in power loss evaluations in distribution grids. Ogasawara et al. presented ways to normalize load profiles in 2010 [24]. The normalization can be done by dividing by the maximum value, by using minimum and maximum of a profile and more complex by adaptive normalization with adaptive neural networks. The latter are a kind of sliding window techniques and have the advantage of creating individual data sequences from which statistical properties are determined for normalization. In their bottom-up approach, Li et al. used the aforementioned temporal permutation to change individual load profiles [30]. In their hourly load profiles, they permuted pairs of values every 50–100 h to add randomness.

Few open-source tools already exist for anonymization of data and creation of load profiles. The ARX tool, for example, is an anonymization tool for structured data sets [35]. With this tool, data can be adapted according to k-anonymization and l-diversity, among other methods. It also has a graphical user interface (GUI) that allows users to anonymize their data sets. However, multiple datasets are required for anonymization in this tool. Thus, it cannot be used for anonymizing individual load profiles. To generate synthetic load profiles, the aforementioned Pinceti et al. have published the LoadGAN tool [36]. This tool uses the methodology of generative adversarial networks and users can directly generate a desired number of load profiles in a desired resolution and length. The LPAT tool by Schaefer et al. splits load profiles into sub-load profiles, which are then used to dimension storage for a hybrid energy storage system [37]. The division into sub-profiles is done by DFT, low-pass filter and inverse DFT. As a result, a load profile is decomposed into several load profiles that have different frequencies and thus can be covered by different storage technologies. However, the goal here is not anonymization, but the determination of the storage requirements. Therefore, no permutation or normalization takes place.

1.1.3. Battery storage systems

Load profiles have a fundamental influence on the design and operation of BSS [38–40]. A collection of load profiles or a generation of anonymized load profiles is especially relevant for battery research, since simulations on storage applications depend strongly on the load profile of the household or company [38]. In general, stationary BSS are operated in various applications. These applications can be divided, for example, into in-front-of-the-meter (FTM) and behind-the-meter (BTM) applications [41]. The former are related to markets and the electricity grid, such as arbitrage trading and frequency containment reserve (FCR). The latter describe applications behind the meter on the consumer side, which can be peak shaving (PS) or self-consumption increase (SCI) of energy generated by photovoltaic (PV) systems. In addition to these singular applications, the applications can also be combined in multi-use scenarios, which can increase the profitability, but bring regulatory barriers with it [41]. Since the focus of this paper is on the use of the load profiles in the SCI application and the PS application these are considered in more detail below. At SCI, BSSs are installed to increase the consumption of self-generated energy and feed less energy into the grid. This can be done by private households or businesses. In Germany, for example, 430,000 home storage systems have been installed by the end of 2021 [42]. The household load profile has an impact on the design of home storage systems [40]. If a large part of the energy consumed is covered by PV during the day, the BSS can be designed smaller than if a large part is consumed in the evening or at night. BSS in PS application are used to cover peaks in the load profile [39]. This is mainly relevant for industrial customers, who must pay a fee per kW for the peak power in a year over a 15-minute period. Accordingly, the load profile is also relevant for the storage design in this

application.

Several KPIs are relevant for BSSs, which are being discussed in the following [38]. The parameters extracted at this point will be used in Section 4.2 to estimate the storage performance with the anonymized load profiles versus the original load profiles. The appendix Section 7.2 shows the equations of the different KPIs with a short description. In general, the degradation of BSSs is highly relevant in every application as it is decisive for the profitability. The degradation can be divided into cyclic and calendric ageing. On the one hand, cyclic degradation depends on the number of equivalent full cycles (EFCs), the depth of discharge (DOD), and the C-rate [43,44]. More EFCs, deeper DODs and larger C-rates lead to increased cyclic aging [44]. Calendar degradation, on the other hand, depends primarily on temperature and the state of charge (SOC) [45,46]. Particularly high, as well as particularly low temperatures and SOC, usually lead to accelerated aging [44]. Two further parameters are of relevance especially for the SCI application: One is the self-consumption rate and the other is the degree of self-sufficiency [47,48]. The self-consumption rate describes the proportion of PV energy consumed locally and not fed into the grid. The degree of self-sufficiency specifies the independence from the electricity grid, thus the proportion of electricity consumption that can be covered by PV energy and battery discharge. The load profile of the household has a major impact on the two parameters [40]. For the PS application, another KPI is of particular importance: the fulfillment factor or performance criterion [49]. This factor describes at what proportion of the time a storage system was able to deliver the power requested by the energy management system. If a BSS operates in the PS application, failures and non-fulfillments of requested power can lead to enormous increases in power-related costs for the business. In addition, round-trip efficiency is described as a relevant parameter in the PS application [49].

The research gap we identified is manifold. On the one hand, many load profile analyses are limited to extracting features to cluster a set of profiles. On the other hand, bottom-up and top-down approaches exist to generate load profiles from, for example, user data. Various methods exist for anonymizing data, but these cannot be applied to individual load profiles because they require a larger data set. Existing open-source tools can either anonymize data (but not load profiles) or generate load profiles (but not anonymize existing ones). Often, the analyses, clustering methods and load profile generation methods are also limited to profiles with resolutions of 15 min [9,16], 30 min [13] or one hour [15,30,31]. In our point of view, what is missing is a methodology that can be used to gradually anonymize individual original load profiles with a high resolution of one to five minutes. In a publication by Beck et al, for example, 5 min was shown to be relevant for the sizing of the power of a storage system [50]. The presented methodology gives users the flexibility to decide to which degree the original load profile should be modified. Since modifying or anonymizing the load profile affects the design and operation of BSSs as described, we also investigate the impact of load profile anonymization on stationary BSSs in different applications. We will answer the following research questions (RQs) throughout the paper:

RQ1) How can load profiles be anonymized gradually and how could an open-source tool look like that allows anonymization and enables an easy and straightforward use in industry and research? (Section 3).

RQ2) In which parameters do the anonymized load profiles differ from the original profiles and in which are they similar? (Section 4.1).

RQ3) How much may an original load profile be modified to maintain parameters critical to a storage application? (Section 4.2).

RQ4) How sensitive are the results from RQ 3 to storage system design and adjustable parameters, such as the threshold between base and peak sequences? (Section 4.3).

1.2. Scope of this work

The goal of the present work is to develop a methodology through

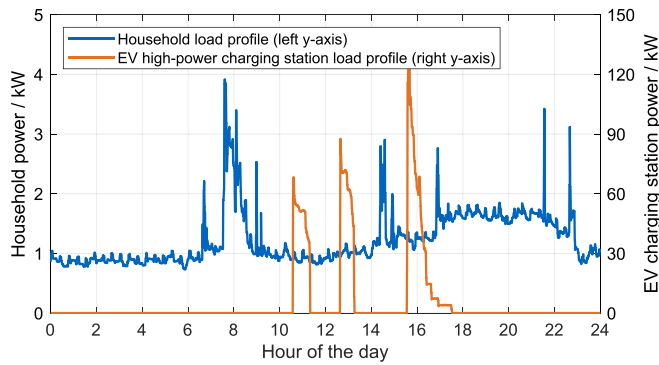


Fig. 1. Exemplary day of the household load profile (January 14th, 2021) and EV high-power charging station load profile (February 2nd, 2022).

Table 2

Database of load profiles of the different applications.

	Household load profile	Charging station load profile
Storage application	Home storage system	High Power EV Charger with buffer
Length of datasets	12 months	6 months
Time period	2021	Mid-January 2022 to Mid-July 2022
Resolution	1 min	1 min
Consumption	10.8 MWh	9.47 MWh
Peak Power	8.685 kW	248.1 kW
Industry/ Research Partner	ISEA RWTH	Industry Partner

which users from industry and research can modify load profiles that are protected by data privacy laws so that they can share them with partners. The extent to which the load profiles must be modified for this purpose is to be decided by the owner of the data. For this reason, our methodology is intended to be flexible and allow users to anonymize gradually. Our approach is to extract features from the original load profile and then to recreate it based on these features. For this purpose, we divide the load profile into base and peak sequences, whereby the threshold between the ranges can be freely adjusted. Subsequently, features such as the mean value and the length of the sequence are determined for each sequence. The anonymized synthetic load profile is then formed from these stored features. The degree of anonymization can then be varied via levels, which determine the variant of the permutations. For example, simple normalization is possible in level 1, while only peak sequences are permuted in level 3 and base and peak sequences are permuted in level 5. The flexibility for users derives from the choice of the threshold between base and peak sequences and the choice of the anonymization level. After generating the anonymized load profile, a storage system simulation model can be called and executed that simulates the original and the anonymized load profiles in different storage applications. In this way, we test the impact of anonymization on BSS operation. The BSS applications we analyze in this work are PS (where peak loads are covered by storage systems) and SCI of PV-generated energy for households using storage systems.

On the one hand, the developed methodology allows the anonymization of classical, continuous load profiles. On the other hand, it can also be used to anonymize load profiles that have idle sequences and peak sequences with high-power levels. In our analyses, such load profiles are from charging stations at which electric vehicles (EVs) charge or whose load is zero apart from standby consumption. Users can thus anonymize daily, weekly, or annual load profiles. In addition, the tool can be used to generate several similar synthetic load profiles from one original load profile. In this way, small data sets can be multiplied and used for data augmentation. Furthermore, storage operators can use the

tool to test a possible storage operation without having to give out or receive original load profiles. The contributions of this work are as follows:

- Methodology to gradually anonymize load profiles by permutation of base and peak sequences while retaining the parameters essential for battery storage use
- Analysis of effects of load profile anonymization on the parameters essential for battery storage utilization
- Demonstration of the open-source tool LoadPAT for providing research and industry the opportunity to share data and present results to public without conflicting with non-disclosure agreements

2. Database

For the present work, two types of load profiles were collected in exchange with an industrial partner and a research institution. These data can be classified in load profiles of a household and of an electric vehicle (EV) high-power charging station (see Fig. 1). The first is a typical load profile of a household, consisting of base load sequences and peak load sequences. The base load sequences result from appliances that always use electricity, such as refrigerators in private households. Peak load sequences are caused by more power demanding consumers used for a short time, such as kettles or electric stoves in private households. The EV high-power charging station load profile has a different pattern. It is composed of charging events and idle phases, in which the load is zero except for standby consumers. The charging stations for EVs provide high charging power during charging times but have very low power during resting periods without a connected EV.

The data used in this work and their parameters are shown in Table 2. The household load profile was measured by the ISEA of RWTH Aachen University within the WMEP home battery storage program [51]. It is used over a whole year with a resolution of 1 min. The EV high-power charging station load profile was provided by an industry partner. It also originates from a storage application as it was measured at a charging station with a buffer BSS. It has a length of 6 months with a resolution of 1 min. The annual energy consumption of the household is approximately equal to the half-year consumption of the charging station. However, the peak power of the household is only 8.7 kW, while the charging station load profile showed a peak power of 248.1 kW within the six months. The complete load profiles are shown in the appendix.

For this work, Python 3.8 was used with the Tkinter package to create LoadPAT and its GUIs [52]. Furthermore, MATLAB was used for data analysis and for the creation of the figures.

3. Methodology of the gradual anonymization

In this chapter, we describe the methodology of the feature-conserving load profile anonymization. First, the differences between the levels of anonymization are explained (Section 3.1). Subsequently, the features used for anonymization are described (Section 3.2). Afterwards, the computational process of the anonymization is described in Section 3.3. In Section 3.4, the interface to the BSS tool Simulation of stationary energy storage systems (SimSES) is described, and the storage parameters of the applications are defined. Finally, we describe how Monte Carlo simulations can be used to determine the influence of the level of anonymization and to measure the respective deviations from the original profile (Section 3.5). Beyond this presentation of the methodology, there is a presentation of the open-source tool including screenshots in the appendix (Section 7.4).

3.1. Levels of anonymization

The gradual anonymization methodology allows differentiation into different levels. Before the exact synthesis of load profiles is presented in

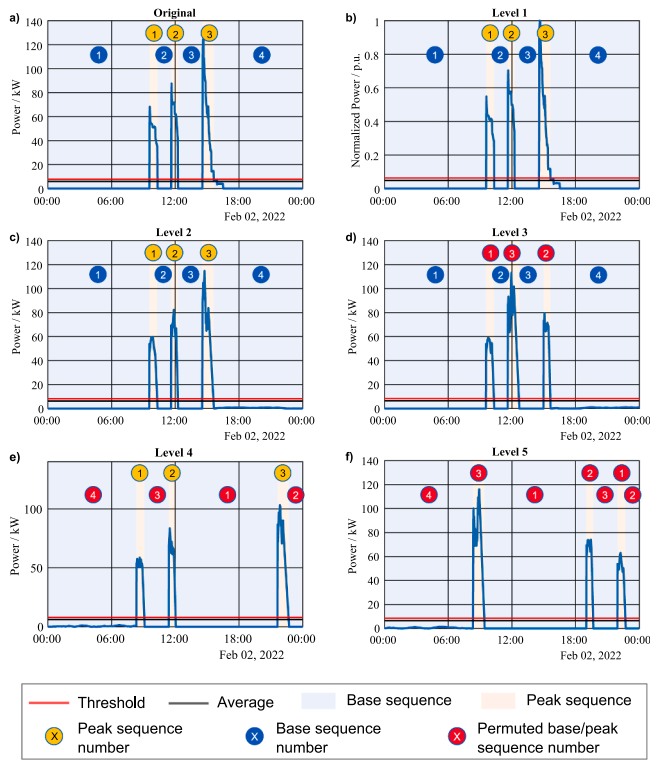


Fig. 2. Examples of daily anonymized load profiles depending on the level of anonymization.

Table 3

Difference between the levels of anonymization.

Number of Level	Description
Level 1	Copying of profile and normalization
Level 2	Feature extraction of sequences and profile creation based on features
Level 3	Level 2 + permutation of peak sequences
Level 4	Level 2 + permutation of base sequences
Level 5	Level 2 + permutation of peak sequences and permutation of base sequences

Table 4

Features extracted for the anonymization. Every feature is calculated for every base and peak sequence.

Peak Features	Base Features
Length/ timesteps of peak sequence without ramp-up and ramp-down (n)	Length/ timesteps of base sequence (n)
Maximum (P_{max})	Maximum (P_{max})
Minimum (P_{min})	Minimum (P_{min})
Mean of delta between two consecutive values (μ_{Δ})	Mean of delta between two consecutive values (μ_{Δ})
Standard deviation of delta between two values (σ_{Δ})	Standard deviation of delta between two values (σ_{Δ})
Probability of change of sign of the delta between two consecutive values ($P_{change\ of\ sign}$)	Probability of change of sign of the delta between two consecutive values ($P_{change\ of\ sign}$)
Mean of peak without ramp-up and ramp-down (μ)	Mean (μ)
Ramp-up length (n_{up})	
Ramp-down length (n_{down})	

the following sections, an example is used to explain the levels of anonymization. To illustrate the various levels of anonymization, Fig. 2 and Table 3 show the differences. The top left diagram (a) shows an

exemplary original load profile. This profile is an EV charging station load profile over the course of a day. At this point, a daily profile is chosen to explain the procedure of the anonymization. In the later course, the permutations are carried out over the half-year (EV charging station) or full-year profile (household). The original daily profile has three peak sequences and four base sequences.

In the anonymization **level 1**, the original profile is merely copied and can be normalized if desired (Fig. 2 b). This is to enable an exclusive normalization of load profiles. From **level 2** onwards, features are determined for every peak and base sequence from the original profile and the profile is then reconstructed based on these features. In this level, the order of base sequences and peak sequences is maintained. Fig. 2 c shows an example of an anonymized load profile of level 2. This profile is remarkably similar to the original profile at first glance. However, at second glance, differences within the peaks become clear. These are due to the recreation of the profile using the features presented later. Thus, this level of anonymization can be applied by users of the method who want to modify or hide variations within base or peak sequences. The occurrence of peaks and base sequences in this level takes place at the same time as in the original profile. In **level 3**, sequences are exchanged for the first time (Fig. 2 d). The base or rest sequences are created identically to the original profile, while the peak sequences are randomly permuted (red colored circles). In the example, the maximum peak is now the second peak, whereas in the original profile it was the third one. Peak sequence 2 and 3 have swapped accordingly. It can also be seen that the peak shape differs. The random creation of the shape of the sequences will be explained in Section 3.3. The third level of anonymization can be applied when users want to hide the exact times at which conspicuous peaks occur. For example, times when industries use certain machines could be randomly shifted in the load profile. In **level 4**, instead of the sequence of peaks, the order of base sequences is randomly varied. In Fig. 2 e the maximum peak is again the third peak, for example. The very long base period (number 1) of over 9 h between midnight and 9:30 in the original profile is no longer present at the same place in this level. Instead, the profile starts with base sequence number four of about 8 h, which is at the end of the day in the original profile. Base sequence number one is now after the second peak. Anonymization in the fourth level changes the load profile more than in the levels before. The times at which characteristic peaks occur change due to the permutation of the base sequences, which means that the use of machines or larger household appliances can no longer be assigned to typical times in the load profile. **Level 5** finally combines the variations of level 3 and level 4, randomly varying the order of base sequences and the order of peak sequences. As a result, it is now no longer possible to read the order of the peaks from the load profile. An example of this level is shown in Fig. 2 f. Users who select this level of anonymization change the load profile the most. Typical characteristics of the load profile still occur, but at completely randomized times. We will discuss the effects of the levels of anonymization in Section 0. The use of a daily load profile in this section serves to illustrate the levels. Within the results, the full profiles are used, resulting in a larger number of base and peak sequences. The permutations are always performed over the entire load profile present. For the EV charging profile this means over half a year and for the household load profile over one year. At this point, it should be noted that the division into the two types of sequences, base and peak, is only one possible way of analyzing the load profile that is commonly used. Alternatively, we could also divide the load profile into four types of sequences, as Haben et. al did with household load profiles [12].

3.2. Features extracted and used for the anonymization

To generate the anonymized load profiles according to levels 2 to 5, features of the base sequences and the peak sequences are calculated. These features are shown in Table 4. For each base sequence, the length of the sequence, the mean value, the standard deviation, and the

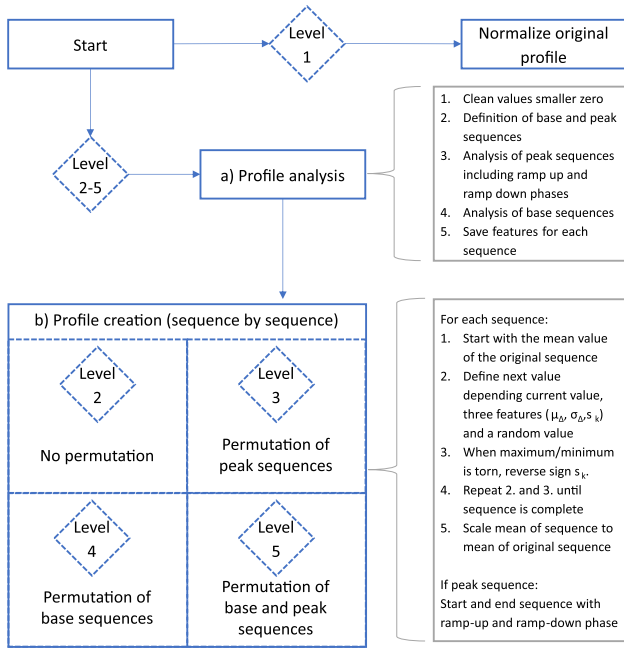


Fig. 3. Flow chart of the process of the anonymization.

maximum are saved. In addition, the mean delta between two values and the standard deviation of this delta are calculated. Moreover, the probability of a change in the sign of this delta is determined. For each peak sequence, basically the same values are determined. However, a distinction is made between ramp-up phase, peak without ramping and ramp-down phase. The mean value and the standard deviation are calculated for the peak without ramping. In addition, the length and the slope are determined for the ramp-up and ramp-down phases.

3.3. Computational process of the anonymization

This section describes the process of anonymization including the calculation of the individual values. Fig. 3 shows a flow chart of the process for clarification. If anonymization is to be performed in level 1 with normalization, the profile is normalized to its maximum value. If, in contrast, the desired level is between 2 and 5, the original load profile is analyzed (a) and a synthetic load profile is created (b).

In the profile analysis (a), the values of the original load profile are cleaned up by setting values smaller than zero to zero. Next, peak sequences are determined by checking at which points in time the load profile has values above the defined threshold. Conversely, base sequences are also determined in this way. If peak sequences are only one time step short, they are filtered out and count as part of the current base sequence. Afterwards, all peak sequences are analyzed one after the other. During this process, the features mentioned above are determined. The average value of the respective peak sequence is used as the threshold value for the end time of the ramp up. Likewise, for the start of the ramp down phase the time is used in which the load falls below the average value for the last time before the end of the sequence. With the help of the lengths, the start and end values of the ramp up and ramp down sequences, their slopes are then determined. After the peak sequences, the base sequences are analyzed, and the features shown in Table 4 are calculated. Finally, all values are saved for the subsequent creation of the synthetic load profile.

In the profile creation (b), the procedure depends on the selected level. According to the selected level, the order of the peak and base sequences is kept (level 2), the order of the peak sequences is randomly mixed (level 3), the order of the base sequences is randomly mixed (level 4) or both orders are randomly mixed (level 5). The synthetic profile is

then created sequence by sequence.

The procedure for generating the sequences is identical for peaks and base sequences. The general idea is to use the mean and standard deviation of the delta of the original sequence together with the probability that the delta changes its sign to generate synthetic values successively as displayed in equation (1). The average value μ of the sequence is selected as the start value of the synthetic profile P_1 . Starting from this, the next value is determined using the mean μ_Δ and standard deviation σ_Δ of the delta of the original sequence. For this purpose, the standard deviation is multiplied by the absolute of a random value from the standard normal distribution X . As shown in equation (2), the sign s_k used depends on the sign used in the last time step and the probability of a change of sign of the delta of the original profile $p_{\text{change of sign}}$. For example, if the probability were 100 %, the sign would be multiplied by -1 each time and thus the direction of the delta would be reversed in each timestep. Since a low probability of a change of sign could result in very low or very high power values, the values are limited by the maximum P_{max} and minimum P_{min} of the sequence displayed in equation (3). If one of the limits is torn, the sign s_k is reversed for the respective time step so that the range is maintained. If the other limit is exceeded due to the change of sign, the limit that was torn first is selected as the next value. This way, the synthetic sequence contains few more sign changes than the original sequence, but this is accepted by the advantage of the more realistic power range. Restriction to the range between maximum and minimum power may cause the mean value of the synthetic sequence to differ significantly from the mean value of the original. This is the case when the maximum is far above, or the minimum is far below the mean value. An outlier upwards then leads to higher values, for example. To reduce this discrepancy between the original and synthetic mean values, the sequence is scaled to the mean value of the original sequence.

$$P_k = P_{k-1} + s_k \cdot (\mu_\Delta + \sigma_\Delta \cdot |\mathcal{X}_k|) \quad (1)$$

$$s_k = s_{k-1} \cdot s_0 \quad (2)$$

$$P_{\text{min}} \leq P_k \leq P_{\text{max}} \quad (3)$$

With :

$$k = 2 \dots n$$

$$P_1 = \mu$$

$$s_1 = 1$$

$$s_0 = \begin{cases} -1, & p_{\text{change of sign}} \\ +1, & 1 - p_{\text{change of sign}} \end{cases}$$

$$\mathcal{X}_k \sim \mathcal{N}(0, 1)$$

n : Length of original sequence

In addition to the individual base and peak sequences, the synthetic profile also includes ramp-up and ramp-down phases before and after peak sequences. To take the ramp-up into account, the ramp-up is synthesized based on the last power value of the previous sequence and the mean value of the following sequence. From these two values and the known length of the ramp-up and ramp-down from the original profile, the slope is determined, and the values are appended to the profile. The slope of the original profile is not used here because the last value of the previous sequence fluctuates over each simulation and using the original slope and length may cause an overshoot or jump in the profile. In the ways described, the base sequences, ramp-up phases, peak sequences, and ramp-down phases are iteratively appended to each other to generate the synthetic profile. The order of the sequences depends on the selected level of anonymization, as described above.

Table 5
Storage parameters of the different applications used in the SimSES simulations.

	Home Storage System	High Power EV Charger with buffer
Profile description	Household load profile	Required load by charging EVs
Energy-management strategy (EMS)	SCI	Peak-Shaving
Battery capacity	8.8 kWh	140 kWh
Max. power of storage system	7 kW	250 kW
Photovoltaic (PV) – nominal power	9.3 kWp, Location: Munich	–
Peak-shaving-limit	–	32 kW

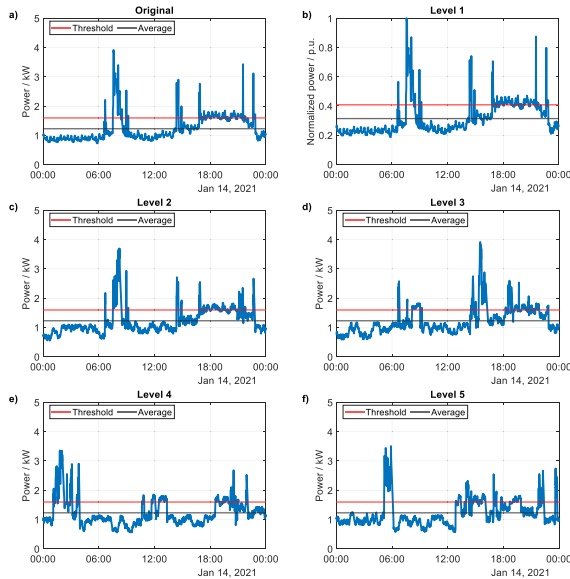


Fig. 4. Exemplary anonymized load profiles of a one-day household load profile.

3.4. SimSES interface and storage parameters

The storage system simulation tool SimSES is a python-based open-source tool that enables time-series-based simulation of storage systems [53]. In SimSES, a variety of parameters can be defined and varied. For example, users can select various energy-management-strategies (EMS) like PS or FCR or implement their own EMS [38]. Furthermore, various lithium-ion battery models with associated degradation models can be used. For a more detailed description of SimSES, please refer to [53].

In this work, SimSES is used to simulate BSSs in different applications using the original and the anonymized load profile. The applications are SCI for the residential load profile and PS for the EV charging station load profile. The standard sizing used in this work for the BSS in each application is shown in Table 5. In the SCI application the BSS is designed with 8.8 kWh and 7 kW, and the PV system has a power of 9.3 kWp, since this was the design of the original system surveyed by the ISEA of RWTH Aachen University. As the focus of this work is on the load profiles, the PV profile available in SimSES is used as the default PV profile. The PV generation profile was measured at the Professorship Power Transmission Systems of the Technical University of Munich in 2014. This profile corresponds to the location Munich and is scaled to the defined peak power of 9.3 kWp. The PV profile is open-source available as part of SimSES [53]. This peak power leads to a generated energy of the PV system of 8,373 kWh over the entire year. If the storage system is simulated with the described parameters and the original load profile over one year, 1,913 kWh are charged into the battery and 1,868

kWh are discharged from the battery.

In the PS application, the BSS is scaled to 140 kWh and 250 kW. The PS limit defined for the grid consumption is set to 32 kW in consultation with the industry partner whose load profile is being used. The designs are used for the simulations to show how the anonymization tool works and to quantify the impact on the BSS load. In Section 4.3, the influence of the design on the results is evaluated in more detail.

The battery model used in both storage applications simulations is a lithium-ion nickel-manganese-cobalt cell from Sanyo, whose parameters have been published by RWTH Aachen [45]. The DC/DC converter is modeled as lossless and the AC/DC converter is modeled following a publication by Notton et al [54].

3.5. Monte Carlo simulations

The generation of anonymized load profiles is based on random permutation of base and peak sequences in the different levels. This dependence on randomness leads to the fact that individual simulations can lead to special, non-representative results. Moreover, in the simulations, non-bijective mappings of load profile KPIs of the BSSs are created. Consequently, no inverse function can be formed to allow a correlation between storage behavior and load profile characteristics. Accordingly, 100 anonymized load profiles are generated for every level in Monte Carlo simulations. The results are ensembles in the solution space which are then evaluated statistically. In addition to the median of the typical KPIs, the scatter will also be shown in the results. In this way, we exclude the possibility that individual, very well-fitting results are displayed.

Overall, this chapter and the tool description in the appendix show how an open-source anonymization tool can be designed to allow gradual anonymization in different levels for easy and straightforward use in industry and research. LoadPAT is available open source, it has GUIs that make anonymization easy for users and it has setting options like level of anonymization and threshold between base and peak sequences.

4. Results

This chapter presents the results of the work. In Section 4.1, exemplary results of the anonymization are presented, and the load profiles are directly compared using various indicators. Subsequently, in Section 4.2 the simulation results of SimSES of the original and the anonymized load profiles are compared. For this purpose, relevant storage KPIs are compared for the different storage applications. Finally, Section 4.3 evaluates the influence of the threshold parameter between base and peak sequences, as well as the influence of BSS and PV design in sensitivity analyses.

4.1. Comparison of the original and the anonymized load profiles

In this section, exemplary results of LoadPAT simulation are presented. Subsequently, the results of the Monte Carlo simulations are compared with respect to the similarity of the load profiles. An example of the anonymization of the EV charging station load profile is already shown in Fig. 2, where a one-day load profile was anonymized to demonstrate the functionality of the tool. Fig. 4 shows a one-day original household load profile together with exemplary profiles of the different levels of anonymization. The diagrams show that, as with the EV charge point load profile, the profile changes more as the level of anonymization increases. Especially the permutation of the base and peak sequences is visible.

For the following results, the two systems are designed according to Table 5 and the load profiles from Table 2 are anonymized. Consequently, the annual load profile is used for the household and the six-month load profile for the EV charging station. Permutations therefore take place over the entire profile length and not just over one day. For

each level, 100 anonymizations are performed in Monte Carlo simulations and the storage is simulated with SimSES. The evaluations in this section refer to the direct comparison of the load profiles without SimSES.

Fig. 5 shows a comparison of four KPIs between the original and the anonymized profiles for the one-year household load profile. The KPIs are the mean (a), the standard deviation (b), the maximum value (c) and the energy (d). For each KPI, the results of the anonymization of the levels 2 to 5 are shown in boxplots. The boxes describe the inter quartile range from 25th to 75th percentile and the red line is the median. The black dashed vertical lines cover all values that are at most 1.5 times away from the maximum or minimum of the interquartile range. The red crosses show outliers that go beyond that. The horizontal dotted line represents the value of the original profile in each case.

Within the three KPIs of mean, standard deviation, and energy there is no major difference between the original load profile and the anonymized load profiles. Only the occurring maximum in the original load profile cannot be reproduced exactly (c). This KPI is underestimated by up to 20%. As described in Section 3.3 the values of the anonymized sequence fluctuate randomly around the average original sequence value and between the maximum and minimum of the original sequence. By generating the sequences in this way, the maximum of the original sequence is not reached in every anonymized sequence. This effect can also be seen in Fig. 4. In the original profile, the maximum value is 4 kW. In the anonymized load profiles of level 2, 4 and 5, however, the value of 4 kW is not reached. The outliers in Fig. 5 whose maximum is above the maximum of the original load profile exist because each sequence is scaled to the mean value of the original sequence after generation. This can cause the maximum of the original load profile to be exceeded in rare cases. In general, forcing the original maximum to be reached would be possible. For example, the synthetic profile could be scaled so that the maximum corresponds exactly to the maximum of the original load profile. However, this would severely overestimate the mean value and energy consumption. If instead only the maximum value of the synthetic profile is set to that of the original, an unrealistically short, large peak would result, which is also not representative of the original profile. For these reasons, an underestimation of the load peak is tolerated at this point. Furthermore, Fig. 5 shows that in all four KPIs, the differences between the levels of anonymization are not significant. This is because the permutation is not

considered within these four KPIs as they are all time independent.

Analogous evaluations are performed for the EV charging station load profile (appendix). The results are similar. Mean, standard deviation and energy are close to the original profile over all levels of anonymization. Furthermore, the maximum of the anonymized load profiles is again smaller than the original profile's maximum.

To determine the time-dependent similarity of the profiles to each other, a DFT of the original household load profile and of four exemplary profiles of the levels 2 to 5 was performed. This method has been used in the literature to cluster and classify load profiles based on periodic patterns [9,37]. The single-sided amplitude of the spectrum of the five profiles is shown in Fig. 6. The frequency range chosen is 0 mHz to 0.0275 mHz, since this corresponds to a period of 10 h to the length of the profile (12 months). The top graph represents the single-sided amplitude of the spectrum of the original profile. The gray shaded areas correspond to periods of about 12 h, 24 h, and 7 days. The original profile shows a pronounced amplitude especially for a period of 24 h and for a period of 12 h. This profile therefore seems to have a period or regularity over days and half days. The period over 24 h emerges because peaks occur with a regularity at approximately the same times of the day. The period over 12 h, whose amplitude is smaller, can be explained by morning and evening load peaks. The spectrum of the profile in level 2, generated using the features from the original load profile, also shows regularity over days and half days. From level 3, where the order of the peaks is randomly permuted, the amplitudes disappear almost completely. There is only a slight increase at the 24-hour period in the spectrum. In Level 4 and 5, in contrast, regularity is no longer recognizable. This is due to the random permutation of the order of the base sequences, whereby the occurrence of peak sequences is shifted in time.

Again, analogous to the industrial load profile, the frequency analysis was performed for the EV charge point profile. The results are shown in Appendix. The original EV charging station load profile shows a regularity over 24 h. In level 2 this regularity can be maintained, while the amplitude in level 3 is smaller but still existing. From level 4 upwards, the amplitude disappears completely.

Overall, the analysis of the load profiles shows that time independent KPIs are preserved despite anonymization. In contrast, the regularity of base and peak sequences is lost due to anonymization, especially from level 3 upwards.

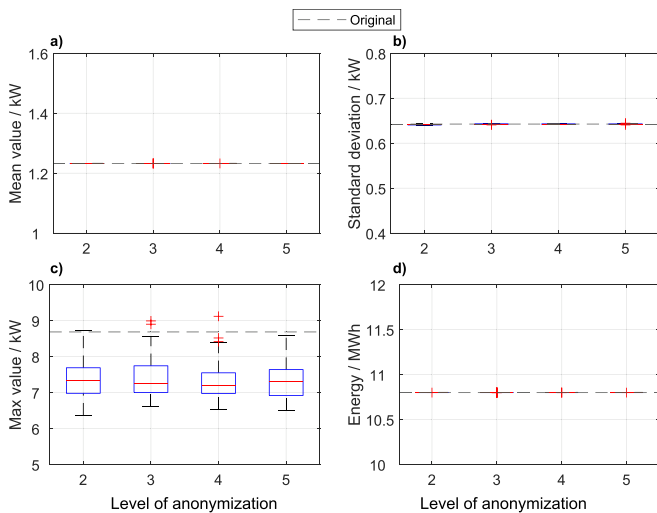


Fig. 5. Comparison of time independent KPIs of original and anonymized one-year household load profiles. 100 anonymized profiles were generated for each level of anonymization.

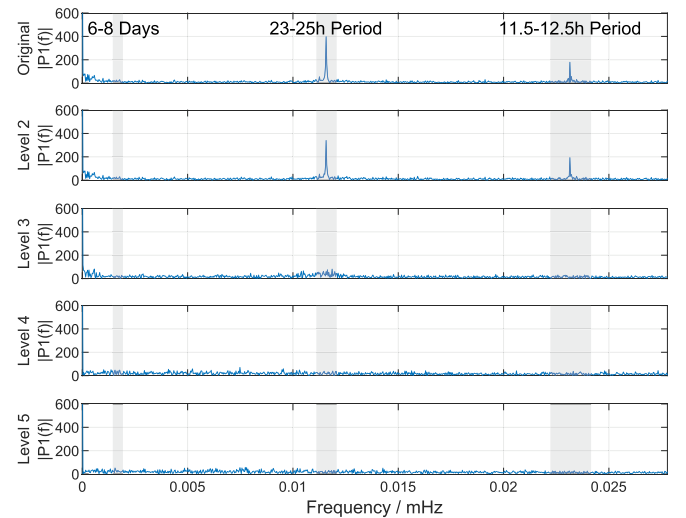


Fig. 6. Discrete Fourier transformation single-sided amplitude spectrum of one-year household load profile in the various levels. Resolution of data: 1 min.

4.2. Effects of the anonymization on relevant battery storage indicators

After the load profiles were directly compared in the previous section, the battery storage applications are now simulated in SimSES using the load profiles. For this purpose, SimSES is invoked as described in Section 3.4 and the operation is simulated once with the original load profile and 100 times with anonymized profiles in every level. Subsequently, various storage KPIs are compared, and it is checked to what extent the anonymization has influenced the behavior of the storage system.

Fig. 7 shows six storage KPIs for the annual household load profiles. The load profiles were used to simulate a BSS in SCI application as described in Section 3.4. The appendix shows the same KPIs with the absolute values of the six storage KPIs. The calculations of the KPIs are shown in the appendix Section 7.2. The deviation over the anonymization levels 2 to 5 compared to the result when using the original load profile. The KPIs shown are the mean SOC (a), the EFCs (b), the mean DOD (c), the mean C-rate (d), the self-consumption rate (e), and the self-sufficiency rate (f). Diagrams a), c), e) and f), on the one hand, show deviations in percentage points, since their original values are already in percent. Diagrams b) and d), on the other hand, show the deviation in percent since their original values are absolute. Above each diagram the absolute value of the KPI of the original load profile is given. Since the deviation from the original is considered at this point, the value of the original is drawn as a dashed line at zero in each case. As in the previous section, the values of the 100 anonymized load profiles are shown as boxplots for each level.

The figure shows that level 2 barely differs from the original in terms of the relevant storage KPIs. Mean SOC, EFCs, mean DOD and mean C-rate are close to the original load profile with small deviations within the 100 anonymizations. The self-consumption rate and the degree of self-sufficiency deviate from the original by only about 1% on average. From level 3 upwards, the deviations from the original are greater with up to 10 percentage points for the self-consumption rate. However, despite the permutations and the abandonment of regularity (see Section 4.1), greater anonymization beyond Level 3 does not lead to greater deviations in the KPIs relevant for BSS. Accordingly, anonymization in level 5 does not appear to be more critical than in level 3 with respect to the KPIs relevant for home storage systems. From level 3 upwards, the average SOC can still be reproduced well. In simulations of the degradation of the BSS with the anonymized load profiles, this would mean that the calendar degradation can be represented well. The mean C-rate deviates only slightly from the C-rate of the original profile, at about 5%.

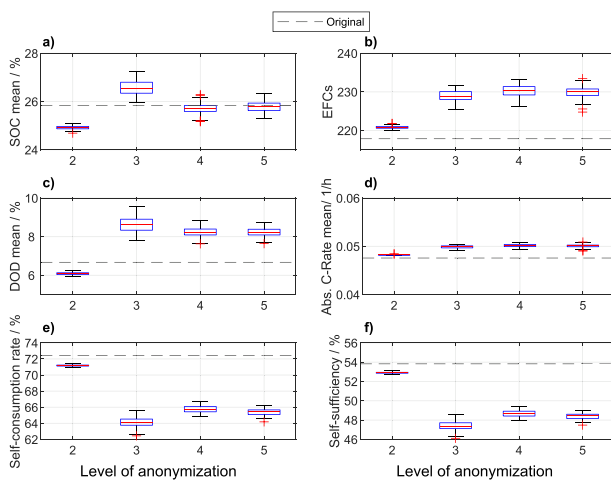


Fig. 7. Comparison of the deviation of BSS KPIs for the household load profile (1 year) in SCI application. a), c), e) and f) show the deviation in percentage points, b) and d) in percent. 100 anonymized profiles per level. PV: 9,3 kWp, E_{Bat} = 8,8 kWh, P_{Bat} = 7 kW.

The simulated BSSs make about 230 EFCs over the year from level 3 instead of 218 in the original. This corresponds to an overestimation of under 6%. Mean DOD is also overestimated by about 2 percentage points resulting in a value of 8 to 9% instead of the 6.7% in the original. The bottom two plots show that the self-consumption rate and self-sufficiency rate are underestimated from Level 3, with deviations of up to 9 percentage points.

In general, some of the results of the KPIs for the SCI case show a systematic overestimation (EFCs, DODs, C-rate) or underestimation (self-consumption rate and self-sufficiency rate) from level 3 to 5 compared to the original values. The original load profile has distinctive peaks especially during the day. The anonymized load profiles between level 3 and 5 have their distinctive peaks distributed over the 24 h. A fan chart over the course of the day with all values for the year once for the original load profile (a) and once for an exemplary load profile in level 3 (b) are shown in the appendix. This shows that the peaks from the daytime hours are distributed over the 24 h of the day due to the anonymization. The shift of the peak loads has an influence on the BSS KPIs. As a result, more energy is charged and discharged from the battery, so the BSS is fully discharged earlier at night and makes more EFCs overall. In addition, the BSS is discharged deeper during evening and night hours due to the load peaks which leads to an increase of mean DODs. The C-rate is slightly higher than the original since the BSS is charged more frequently during the day with high PV power because, for example, in level 3 higher day peaks are exchanged with lower peaks from the night and thus more power flows into the BSS. At level 4, the base sequences are permuted, which leads to even higher C-rates, since the higher base sequences from the daytime exchange with lower base sequences from the nighttime. Next, the self-consumption rate is lower, since less PV energy can be consumed immediately and, once the BSS is fully charged, a larger proportion of the PV energy is fed into the grid. In the evening and at night, more energy must be supplied from the grid, which decreases the degree of self-sufficiency. Since the effects mentioned occur in levels 3, 4 and 5, but do not increase in levels 4 and 5, the KPIs remain relatively constant at higher levels.

Fig. 8 shows BSS KPIs for the EV charging station load profile similar to Fig. 7 but over the period of six months instead of one year. Instead of self-consumption rate and degree of self-sufficiency, round-trip efficiency and fulfillment factor are displayed, as these are relevant for the PS application (see Section 1.1). The appendix again shows a diagram with the absolute values. The Y-axis scaling of the subfigures in Fig. 8 differs from the scaling in Fig. 7 to show the spread of the boxplots. Fig. 8

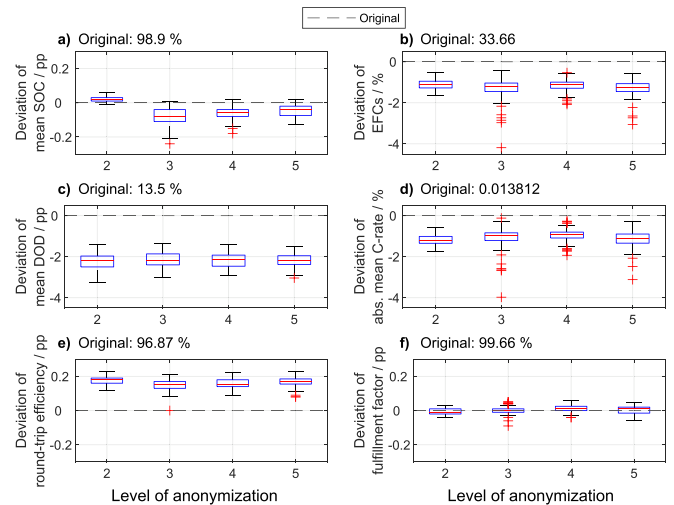


Fig. 8. Comparison of the deviation of BSS KPIs for the EV charging station load profile (6 months) in PS application. a), c), e) and f) show the deviation in percentage points, b) and d) in percent. 100 anonymized profiles. E_{Bat} = 140 kWh, P_{Bat} = 250 kW, Peak-Shaving with Threshold 32 kW.

shows that in all KPIs, the anonymized load profiles show only minor deviations from the value of the original load profile of up to 3% for the C-rate and 3 percentage points for the mean DOD. For example, the mean SOC deviates by only up to 0.3 percentage points. More importantly, the fulfillment factor corresponds to the original across all levels. In contrast to the SCI application, the deviation is constant over the four levels of anonymization. This is because the PS application does not depend on a time-varying PV generation. From level 3 onwards, the permutations cause a shift in the peaks, but this has no influence on the relevant BSS KPIs. This can also be seen from the fact that level 2, in which there are no permutations, deviates just as strongly in most KPIs as levels 3 to 5. The largest noticeable deviation across all six KPIs is mean DOD, which is underestimated by up to three percentage points, resulting in a mean of around 11.3% instead of 13.5%. To explain this underestimation, fan charts are included in the appendix showing the original load profile (a) and a level 3 anonymized load profile (b) over the course of the day. Similar to the fan charts of the household in, the diagrams show that there is a homogenization of the load peaks over the 24 h. This homogenization decreases the mean DOD of the BSS, since the BSS is discharged less frequently by several consecutive load peaks. Furthermore, the maximum of the peaks is smaller as described in Section 4.1, which leads to a decrease of the DOD. For the same reason, the average C-rate is smaller compared to using the original load profile.

Overall, our analyses reveal three results: First, load profiles for use in simulations of time-independent applications such as PS can also be anonymized up to level 5 and thus be strongly modified without changing KPIs relevant for BSS excessively. Smaller characteristic peculiarities in the sequences of the original load profile are consequently not relevant and can be replaced by mean values and random fluctuations. Second, load profiles in time-dependent applications such as SCI can easily be anonymized in level 2 without changing BSS KPIs drastically (approximately 1% in all KPIs). Vice versa, the anonymized profiles can well be used to design a BSS with suitable sizing and realistic estimation of battery degradation. From level 3, the deviations are larger by up to 9 percentage points, and users would have to decide for themselves whether the deviation is still within a reasonable range in order to realistically represent the original load profile through anonymization. Third, if it is decided to anonymize the load profile beyond level 2 in time-dependent applications, higher anonymization in level 4 or 5 does not lead to larger deviations in the relevant BSS KPIs compared to level 3. Accordingly, the load profiles can then be changed more strongly to achieve higher anonymization.

4.3. Sensitivity analysis of threshold and system design

In this section, we perform a sensitivity analysis of the SCI case with the household load profile to explain the impact of three sensitivity parameters on the results. The first sensitivity parameter is the threshold between base and peak sequences. This threshold describes the boundary between base and peak sequences as a multiple of the mean value of the load profile. It can be freely chosen by the users and is set to 130% in the base case, as this provides good separation between base and peak sequences. Basically, a shift in the threshold means that the division into base and peak sequences changes. If, on the one hand, the threshold is set very low, large power variations are all defined as peaks, which leads to larger fluctuations but lower maxima in the peak sequences of the anonymized load profile. If, on the other hand, it is set very high, the base sequences will have large fluctuations and the peak sequences become shorter but more pronounced.

Fig. 9 shows for the household load profile the six BSS KPIs for the original load profile and the medians of the four levels of anonymization over the varying threshold. The threshold is varied in a range from 90%

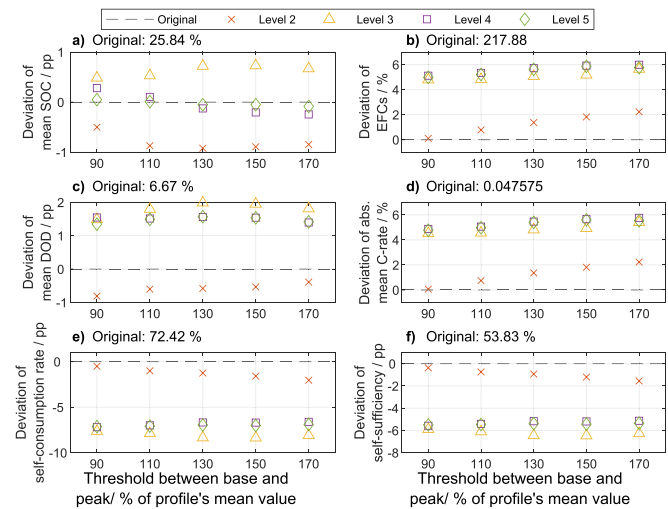


Fig. 9. Sensitivity analysis of the SCI use case with household load profiles: Threshold between base and peak sequences. Threshold defined as percentage of profile's mean value. The levels' values are the median values of 100 simulations. The subplots show the deviation of the results in the levels compared to the original in percentage (b, d) and percentage points (a, c, e, f).

to 170% in 20% steps. The diagram shows the deviations of the KPIs from the original value, which is set to 0 as in Fig. 7. The values of levels 2 to 5 correspond to the median value of 100 anonymizations. For a better understanding of the absolute values, a diagram showing the respective absolute KPIs is presented in Appendix. The number of EFCs increases over all levels slightly as the threshold rises (b). If the threshold is low, more shares of the profile are evaluated as peaks, so that the peak sequences have larger differences between their maximum and minimum. This leads to the fact that the maxima are reached less often and a larger part of the peak sequences can be covered by the PV energy. At a higher threshold the peak sequences become shorter, but more pronounced. These more pronounced peaks can no longer be covered directly by PV generation. Therefore, more peaks must be covered by the BSS, which increases the number of EFCs. Furthermore, the self-consumption rate (e) and self-sufficiency (f) decrease slightly in level 2 but remain relatively constant in levels 3 to 5. The decrease in level 2 comes from the larger fluctuations within the base sequences and the higher maxima in the peak sequences that occur at higher thresholds. This means that a slightly smaller proportion of the PV energy can be consumed directly. In the higher levels of anonymization, the permutations have already changed the load profiles in such a way that the influence of the threshold is lower than in level 2. Overall, Fig. 9 shows that level 2 deviates the least from the original across all thresholds, as expected. Over the varying threshold, however, the deviations within each level change only slightly. The influence of the choice of threshold is therefore small.

As a second parameter, the BSS capacity of the home storage system is varied. Fig. 10 shows the variation between 4.8 kWh and 12.8 kWh with a step size of 2 kWh in the same format as for the threshold variation. Again shows the absolute values, while Fig. 10 shows the deviation. In contrast to the variation of the threshold, this variation of capacity also changes the value of the original profile. Therefore, the original value is shown as a single line and not as a horizontal line. Since Fig. 10 shows the deviation from the original, the respective original value is again shown as 0.

Basically, shows that, for example, the EFCs decrease with increasing capacity, while the self-consumption rate and self-sufficiency increase. Here, the influence of the storage design on the six parameters exists as

expected. Analogous to the analysis of threshold sensitivity, level 2 again shows the smallest deviations from the original profile (Fig. 10). Furthermore, the deviation of levels 3 to 5 in terms of mean SOC (around +1 to −1.5 percentage points), mean DOD (around 2 percentage points), self-consumption rate (−9 to −5 percentage points) and degree of self-sufficiency (−7 to −4 percentage points) is approximately constant over the increase in BSS capacity. An increase in storage capacity, on the other hand, leads to greater deviations regarding EFCs (2 to 10 %) and mean C-rates (1 to 10%). These deviations are constant over the levels. As described in Section 4.2, the presented anonymization leads to more EFCs. If the BSS capacity is now increased, more energy can be discharged at night by peaks occurring only in the anonymized load profile. Thus, the deviations of EFCs increase with increasing BSS capacity. The same applies to the C-rate, which is higher on average because the BSS must more frequently cover the peak loads that cannot be covered by the PV system. Overall, this sensitivity analysis shows that anonymization is relatively robust with respect to the selected storage capacity.

Finally, Fig. 11 show the variation of the PV peak power between 5.3 kWp and 13.3 kWp in 2 kWp steps for the household load profile case. First, as with the BSS capacity variation, the influence of the size of the PV system on the KPIs can be seen: If the PV system is larger, the EFCs increase because more PV energy can be stored. Likewise, the mean C-rate increases, since the BSS is charged with a higher power due to the larger PV system. In addition, the self-consumption rate decreases and the self-sufficiency rate increases with increasing PV nominal power, as has already been confirmed using field data [51]. The results in Fig. 11 are similar to those in Fig. 10. The deviations across the enlarged PV system are relatively constant for mean SOC, mean DOD, self-consumption rate, and self-sufficiency. However, in contrast to the BSS evaluations, the deviations of the EFCs and mean C-rates are large for small nominal PV power of 5.3 kWp with up to 20% over the levels 3 to 5. If the PV system is small, the EFCs are therefore overestimated. The EFCs of the BSS in the original load profile are 137, while they are 163 when anonymized according to level 5 which is an overestimation of 19%. Basically, a smaller PV system can only insufficiently cover the load by the household (degree of self-sufficiency decreases). The permutation of the peak sequences and the possible shift to times without PV generation means that the BSS can be charged more during the day but is also discharged more at night. For this reason, EFCs increase especially when the nominal PV power is small.

Overall, this section answers the question of how sensitive the results

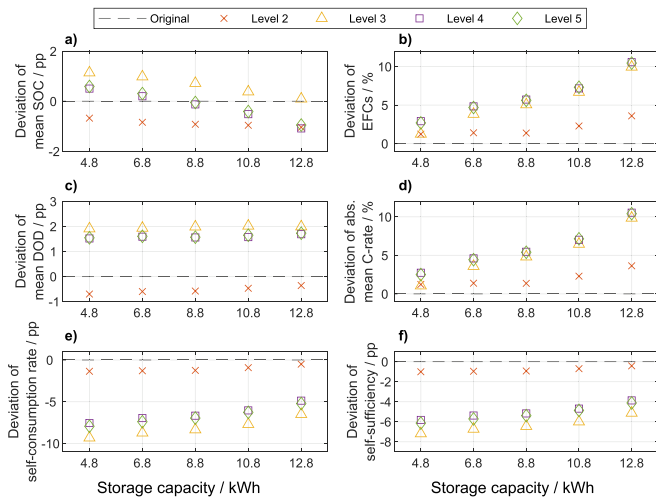


Fig. 10. Sensitivity analysis of the SCI use case with household load profiles: Storage capacity. The levels' values are the median values of 100 simulations. The subplots show the deviation of the results in the levels compared to the original in percentage (b, d) and percentage points (a, c, e, f).

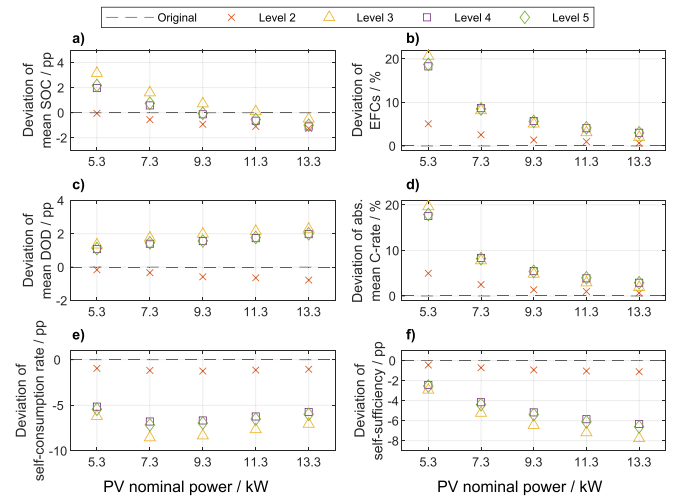


Fig. 11. Sensitivity analysis of the SCI use case with household load profiles: Nominal PV power. The levels' values are the median values of 100 simulations. The subplots show the deviation of the results in the levels compared to the original in percentage (b, d) and percentage points (a, c, e, f).

are to the threshold between base and peak sequences and the system design. The influence of the threshold is rated as low. The BSS KPIs change only slightly by varying the threshold and the deviations between original and the different levels of anonymization remain relatively constant. In principle, this also applies to the variation of the capacity of the BSS and the nominal power of the PV system. Only particularly large capacities of the BSS and small nominal PV power lead to stronger deviations between two KPIs of the anonymized load profiles and those of the original (EFCs and C-rate). A realistic dimensioning of the system design is therefore necessary.

5. Conclusion and outlook

This chapter summarizes the work in Section 5.1. Furthermore, in Section 5.2, we discuss the usability of the tool and the strengths and weaknesses of the methodology, and we give an outlook on how the presented work could be followed up.

5.1. Conclusion

In this work, a methodology of anonymization of load profiles is presented. For this purpose, the existing literature of load profile clustering and anonymization is presented first. Within this research we find that a methodology to gradually anonymize existing load profiles has not been published yet. Hence, we develop a methodology to fill this research gap. Our approach extracts features from the original load profile and separates the profile into base and peak sequences. A synthetic, anonymized load profile is then generated from the features of each sequence. The gradual anonymization is enabled by anonymizing in different levels. A simple normalization of the original load profile is possible in level 1. In Level 2, the features are extracted and used for profile generation along with random values of the standard normal distribution, which represent a type of noise. From level 3 on, peak or base sequences are permuted. In level 3, the peak sequences are permuted in a random way, which shifts peaks that are characteristic for the load profile in time. In level 4, the base sequences are permuted in a random manner. The order of the peaks is preserved, but they also shift in time due to the change in base sequences order. Finally, level 5 allows the combined permutation of base sequences and peak sequences. This results in more strongly modified load profiles. To make the presented methodology usable for the public, the open-source load profile anonymization tool LoadPAT was developed. LoadPAT is coded in Python

and can be used on any computer. This allows companies, research institutions and private individuals to keep the original data on their computers and to create and subsequently share the anonymized load profiles themselves. The methodology presented and LoadPAT answer our first research question, how load profiles can be anonymized gradually and how an open-source tool could look like that allows anonymization and enables an easy and straightforward use in industry and research (RQ 1).

Two load profiles were selected as use cases for testing the methodology within LoadPAT: A household load profile and an EV charging station load profile. For both use cases, 100 anonymized load profiles were created in Monte Carlo simulations across all possible levels. Subsequently, the anonymized load profiles were compared with the original load profile in each case. The results showed that the anonymized load profiles correspond to the original load profile in time-independent KPIs, such as the mean value or the standard deviation. However, a subsequent analysis of the amplitude of the DFT of the load profiles shows that the regularity of the load profiles is lost due to anonymization from level 3 upwards (RQ 2).

Afterwards, the original and anonymized load profiles are used as input profiles for the storage simulation tool SimSES and various storage applications are simulated. The household load profiles are used to simulate a SCI scenario. The EV charging station load profiles are used for a PS scenario. Our results show that anonymizing load profiles has only minor impact on KPIs relevant to BSS in time-independent storage applications such as PS. However, if time-dependent storage applications such as SCI (dependence on PV generation) are considered, anonymization from level 3 leads to larger variations in relevant KPIs. If, on the one hand, this variation is considered by users to be significant, anonymization should only be performed up to level 2. If, on the other hand, it is considered acceptable, anonymization can even be performed up to level 5, since the differences between the KPIs of level 3 and 5 are small (RQ 3).

Finally, we perform a sensitivity analysis in which we evaluate the influence of the threshold value between base and peak sequences and the influence of the system design (RQ4). This analysis indicates that although the choice of threshold leads to slightly different results, the deviations are relatively constant across the different levels of anonymization. The system design of storage capacity and PV system shows similar results. The KPIs relevant to BSS change with capacity and PV system size, as expected. The deviations across levels are also relatively constant here, except for particularly large capacities and particularly small PV systems.

5.2. Discussion and outlook

The presented methodology of LoadPAT allows the modification of load profiles so that characteristic times of peak and base phases are no longer identifiable. The methodology works for various types of load profiles such as the rather continuous household load profiles, but also the more event-based load profiles of EV charging stations. Moreover, the original load profile cannot be reconstructed from the anonymized load profiles, especially from level 3 onwards, due to the random permutations of a large number of peak or base sequences. In addition, the approach is easy to understand, does not require large data sets or processes with artificial intelligence involved. LoadPAT is easy to use because of the GUI including graphical representation of the load profiles. Users can select the threshold between base and peak sequences and the level of anonymization, and if needed, normalize the load profile to the maximum value. They can even simulate the created load profile in the tool in different storage applications. A weakness of the methodology is that characteristic peaks (e.g. typical machines of a company)

are shifted in time but can still be recognizable even at the highest level of anonymization. Here, the tool could be extended in the future to explicitly blur selected peaks. Furthermore, in the present version of the tool the resolution of the load profiles cannot be changed, and profiles cannot be shortened or extended. In addition, the permutations from level 3 are always performed over the entire length of the profile. A selectable period for the permutations, for example over weeks, could maintain seasonal fluctuations within the load profile. Another approach would be to extract a daily load profile from a monthly original load profile that is as representative as possible but anonymized. Similarly, a methodology could be developed to extract an anonymized, representative load profile from a set of load profiles. An approach that has already been published several times is the clustering of load profiles to summarize similar load profiles [14–16]. The LoadPAT methodology could be used to aggregate larger data sets of anonymized load profiles that would not be allowed to be shared without anonymization. Here, users could be allowed to upload their anonymized load profiles to a publicly available platform. This database could be made freely available to users from industry and research and thus contribute to the standardization of load profiles. In addition, the methodology could also be further developed and adapted so that real-time smart meter data can be anonymized with the goal of masking resident attendance times. Finally, the methodology could be extended to storage profiles that are not only positive but change sign through charging and discharging.

Data availability

The Load Profile Anonymization Tool (LoadPAT) presented in this paper can be downloaded as an open-source Python version from Gitlab [55]. The household load profile will be part of a future publication and the EV charging station load profile is subject to a non-disclosure agreement.

CRedit authorship contribution statement

Benedikt Tepe: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **David Haberschus:** Conceptualization, Methodology, Writing – original draft, Visualization. **Jan Figgener:** Conceptualization, Methodology, Writing – original draft, Visualization. **Holger Hesse:** Conceptualization, Writing – review & editing, Supervision. **Dirk Uwe Sauer:** Resources, Writing – review & editing, Supervision. **Andreas Jossen:** Resources, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was financially supported by the German Federal Ministry of Education and Research (BMBF) within the SimBAS project (Grant No. 03XP0338A), which is managed by Project Management Jülich. The household profiles were measured in the research projects “WMEP PV-Speicher” (funding number 0325666), and “WMEP PV-Speicher 2.0 (KfW 275)” (funding number 03ET6117)”, both funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). The responsibility for this publication rests with the authors.

Appendix

Existing literature on analysis and clustering of load profiles

Table A1

Equations of the storage KPIs

The KPIs used in this work for the behavior of the BSS are presented below. First, the mean SOC results from all SOC values of the storage profile (Equation (4)). We calculate the number of equivalent full cycles (EFCs) analogous to [38] after equation (5) using the energy charged into the storage system over the entire profile divided by the energy of the storage system. The mean DOD is also calculated analogously to [38] from the SOC at the start of the cycle minus the SOC at the end of the cycle (equation (6)). According to equation (7), the mean absolute C-rate is calculated using the absolute mean value of the current in ampere divided by the battery capacity in ampere-hours. The two KPIs relevant to self-consumption increase, the self-consumption rate and the degree of self-sufficiency, are derived from equation (8) and (9), respectively. The self-consumption rate, on the one hand, describes what proportion of the energy generated by the PV was consumed by the household or industry consumer. For example, if no energy is fed into the grid, the self-consumption rate is 100%; if half of the PV energy is fed into the grid, the self-consumption rate is 50%. The degree of self-sufficiency, on the other hand, describes the independence from the grid and therefore depends on the total consumption and the grid supply. If no energy is drawn from the grid, the degree of self-sufficiency is 100%, if half of the required energy is drawn from the grid, the degree of self-sufficiency is 50%. The round-trip efficiency is the energy discharged from the storage divided by the energy charged into the storage (equation (10)) [38].

The charged energy is corrected in the formula to account for the SOC offset between the beginning and end of the simulated period. The fulfillment factor, which is the percentage of time that the system was able to fulfill the requested service, is calculated using equation (11) [49]. Here, P^* stands for the realized power and P_{sys} for the power requested by the system.

$$SOC_{\text{mean}} = \frac{\sum_{t=1}^n SOC(t)}{n} \quad (4)$$

$$EFCs = \frac{E_{\text{total profile}}^{\text{pos}}}{E_{\text{BSS}}} \quad (5)$$

$$DOD_{\text{mean}} = \frac{\sum_{k=1}^m SOC_{\text{cycle } k, \text{start}} - SOC_{\text{cycle } k, \text{end}}}{m} \quad (6)$$

$$C - \text{rate}_{\text{abs, mean}} = \frac{|\bar{I}|}{C_{\text{Battery}}} \quad (7)$$

$$\text{Self consumption rate} = 1 - \frac{E_{\text{grid feed in}}}{E_{\text{PV Generation}}} \quad (8)$$

$$\text{Self sufficiency} = 1 - \frac{E_{\text{grid supply}}}{E_{\text{load total}}} \quad (9)$$

$$\eta_{\text{BSS}} = \frac{|E_{\text{total profile}}^{\text{neg}}|}{|E_{\text{total profile}}^{\text{pos}}| - [SOC_{\text{end}} - SOC_{\text{start}}] \bullet E_{\text{BSS}}} \quad (10)$$

$$\text{Fulfillment factor} = 1 - \frac{\int (|P^*(t) - P_{\text{sys}}(t)|) dt}{\int (|P_{\text{sys}}(t)|) dt} \quad (11)$$

With :

n : Length of load profile

Table A1

Summary of literature on load profile analysis and feature extraction.

Source	Date	Focus	Results
Price [8]	2010	Load analysis	Definition of five parameters to characterize load shapes in time-domain: Base load, peak load, rise time, high-load duration, fall time.
Haben et al. [12]	2016	Load analysis & clustering	Analysis of customer smart meter data including seven attributes that describe relative seasonal and intraweekly power and standard deviation. Those attributes of each profile are used in a finite mixture-based clustering.
Al-Otaibi et al. [13]	2016	Feature extraction for clustering	Calculation of specific maxima and minima in time range as features together with normalization and scaling leads to sufficient clustering results with much fewer features compared to 48 half-hour values of a daily load profile.
Wang et al. [11]	2019	Load analysis, forecasting & management	Review on smart meter data analytics: Load profiling can be done directly using the time-series and indirectly using suitable features extracted from the profile.
Park et al. [14]	2019	Clustering	Direct clustering of building load profiles extracting three fundamental profiles that 94% of 1,832,807 daily load profiles of 3,829 buildings can be assigned to
Trotta [15]	2020	Clustering (k-means)	Four clusters of Danish household load profiles are identified including seasonal fluctuations.
Li et al. [9]	2021	Load analysis	Combination of time-domain (based on the work of Price [2]) and frequency-domain load profile analysis of commercial office buildings. Time-domain analysis based on six key parameters of the load profiles.
Czétány et al. [16]	2021	Clustering	Three clusters of Hungarian household load profiles on daily and on yearly basis are identified. The K-means algorithm is favorable against fuzzy k-means and agglomerative hierarchical clustering.
Elahe et al. [17]	2022	Feature extraction	Identification of households with plug-in electric vehicles using a new feature extraction technique.

m : number of cycles

Illustration of the two original load profiles

The Figs. A1 and A2 describe the annual household load profile and the 6 months electric vehicle charging station load profile described in Section 2.

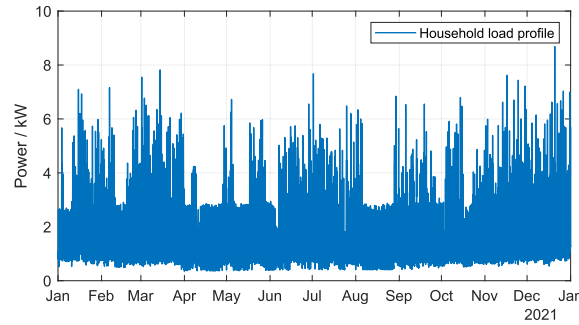


Fig. A1. Annual household load profile for 2021.

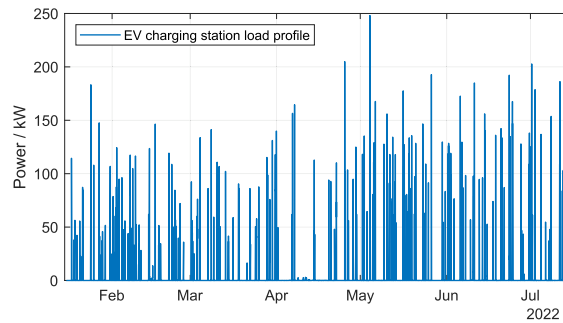


Fig. A2. 6 months electric vehicle charging station load profile from Mid-January 2022 until Mid-July 2022.

Description of LoadPAT

In this chapter we describe the open-source load profile anonymization tool LoadPAT. The basic idea of LoadPAT is to characterize a load profile and generate an anonymous, synthetic load profile based on various features. The similarity between the synthetic load profile and the original load profile can be distinguished by the user based on five anonymization levels. A screenshot of the GUI is shown in Fig. A3. On the left side, the levels of anonymization are explained. On the right side, the anonymization can be performed step by step. In step 1, the users browse to a CSV of the original load profile. Afterwards, the original profile is already plotted.

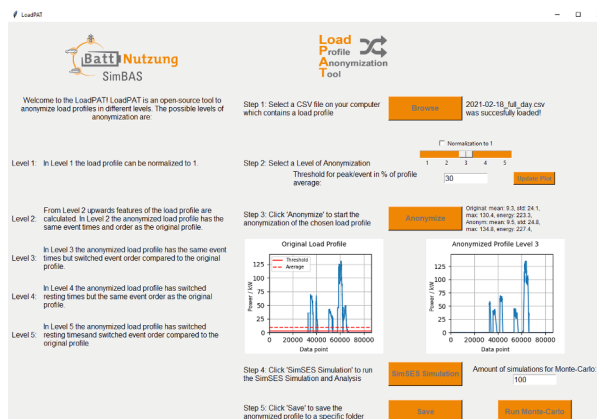


Fig. A3. Graphical User Interface of LoadPAT.

In step 2, users then select the level of anonymization. Here they can also click in a tick box on “Normalization to 1” to normalize the load profile to the maximum value. In addition, the threshold between base and peak sequences can be set in this step. This threshold is to be specified as a percentage value of the average value of the original profile. If, for example, 100% is selected, the threshold value corresponds to the mean value. In our simulations we choose as threshold 130%, as this provides good separation between base and peak sequences. To support the selection of the threshold value, the plot of the original profile shows once the average value of the profile dashed and once the threshold value. By clicking on “Update Plot”, users can update the original plot after changing the threshold value.

In step 3, the actual anonymization takes place. By clicking on the “Anonymize” button, the anonymization is performed. The methodology of the anonymization is explained in detail in Section 3.3. The anonymized profile then appears next to the original load profile, and the users can compare the profiles visually. Moreover, some key metrics like mean values, standard deviation and maximal values are displayed next to the “Anonymize” button. If users are not satisfied with the generated anonymous profile, they can generate a new profile by clicking the button again.

In step 4, users can simulate the original profile and the anonymized profile in a battery storage application via the storage simulation tool SimSES. Finally, in step 5, users can save their anonymized profile. To analyze the operation of LoadPAT, it is furthermore possible to run many anonymizations in Monte Carlo simulations followed by storage system simulations across all levels in succession.

In this work, SimSES is invoked by LoadPAT and a BSS is simulated in different applications using the original and the anonymized load profile. The GUI for using SimSES within LoadPAT, which appears when clicking the “SimSES Simulation” button in the main window of the GUI, is shown in Fig. A4. On the left side, users can set parameters. First, this is the storage capacity in kWh and the maximum power of the storage in kW, which is limited by the power electronics. In addition, the start SOC and the resolution of the time series simulation must be defined. Furthermore, a lithium-ion cell and the EMS can be selected via drop-down menus. Depending on the EMS, the size of the PV system (SCI) or the limit for PS must then be defined. Afterwards, the storage behavior can be simulated with the original load profile and with the anonymized load profile. By clicking the “Show Results” button after the simulations are completed, the results of the storage simulation are displayed on the right side. In the upper part, the key KPIs determined by SimSES for the original profile and the anonymized profile are displayed as a table together with the percentage deviation. Below this, histograms for the KPIs SOC, C-rate, DOD and temperatures are plotted, allowing users to compare these four KPIs graphically. Finally, users can save the results from SimSES by clicking on the corresponding button. If users are satisfied with the results, they can then save the anonymized profile in the main window of LoadPAT. As of this publication, users have access to three battery models, two NMC-based lithium-ion batteries (SanyoNMC and MolicelNMC) and one iron-phosphate lithium ion battery (SonyLFP), whose data have been published in various publications [43,45,46,56]. As EMS, users can choose between two strategies: A strategy of SCI of PV energy (ResidentialPvGreedy) and a PS strategy (SimplePeakShaving).

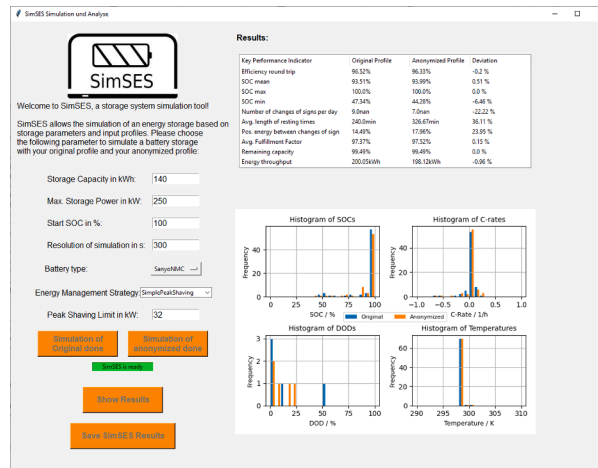


Fig. A4. Graphical User Interface of SimSES in LoadPAT.

Results of the comparison of the original and the anonymized load profile for the EV charging station load profile

Fig. A5 shows the comparison of load profiles for the EV charging station load profile. Fig. A6 shows the single-sided amplitude spectrum of the original profile and 5 exemplary anonymized load profiles for the EV charging station.

SimSES KPI comparison: Plots of absolute values

Figs. A7 and A8 show the comparison of the BSS KPIs for the household load profile anonymization and the EV charging station load profile anonymization in absolute values.

Fan chart of the original household load profile and an exemplary load profile anonymized in level 3

The Figs. A9 and A10 show fan charts of the original load profiles and one exemplary anonymized load profile from level 3 over the course of the day as mentioned in Section 4.2.

Results of the sensitivity analyses of Section 4.3

The Figs. A11–A13 show the absolute values of the BSS KPIs of the sensitivity analysis. In Section 4.3 Figure Fig. 9, Fig. 10 and Fig. 11 the deviations from the original values are displayed.

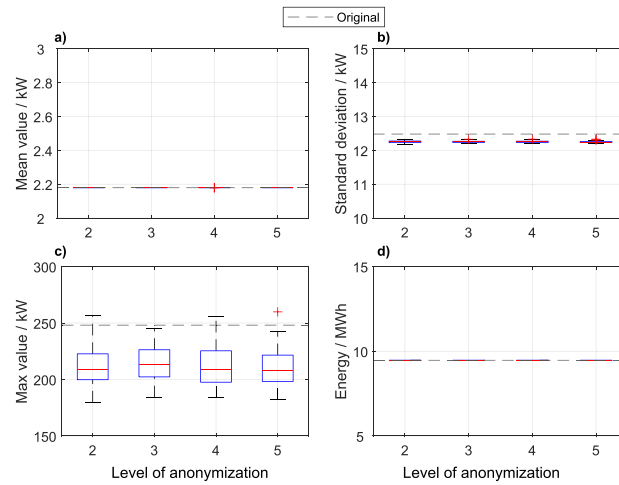


Fig. A5. Comparison of load profiles. EV charging station load profile (6 months). 100 anonymized profiles per original profile.

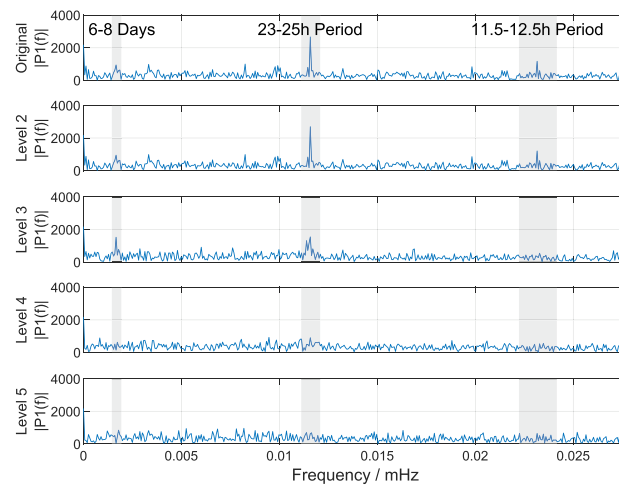


Fig. A6. Discrete Fourier Transformation single-sided amplitude spectrum of 6-months EV charging station profile in the different levels. Time resolution of profile: 1 min.

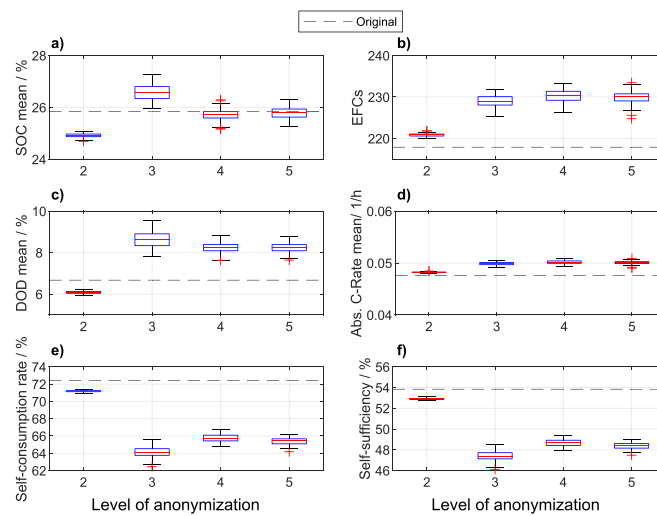


Fig. A7. Comparison of the BSS KPIs for the household load profile (1 year) in SCI application. 100 anonymized profiles per level. PV: 9,3 kWp, E_{Bat} = 8,8 kWh, P_{Bat} = 7 kW.

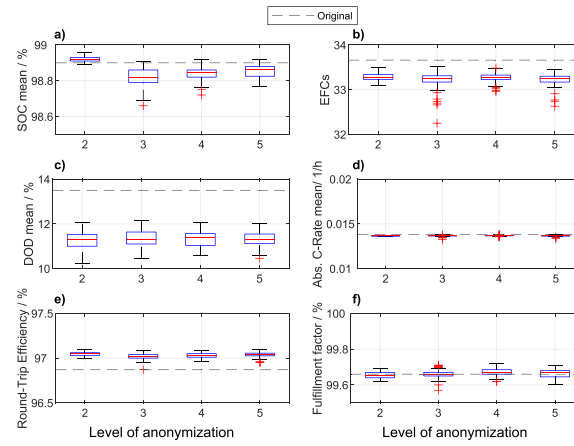


Fig. A8. EV charging station load profile (6 months). Comparison of SimSES characteristics. 100 anonymized profiles. $E_{\text{Bat}} = 140$ kWh, $P_{\text{Bat}} = 250$ kW, Peak-Shaving with Threshold 32 kW.

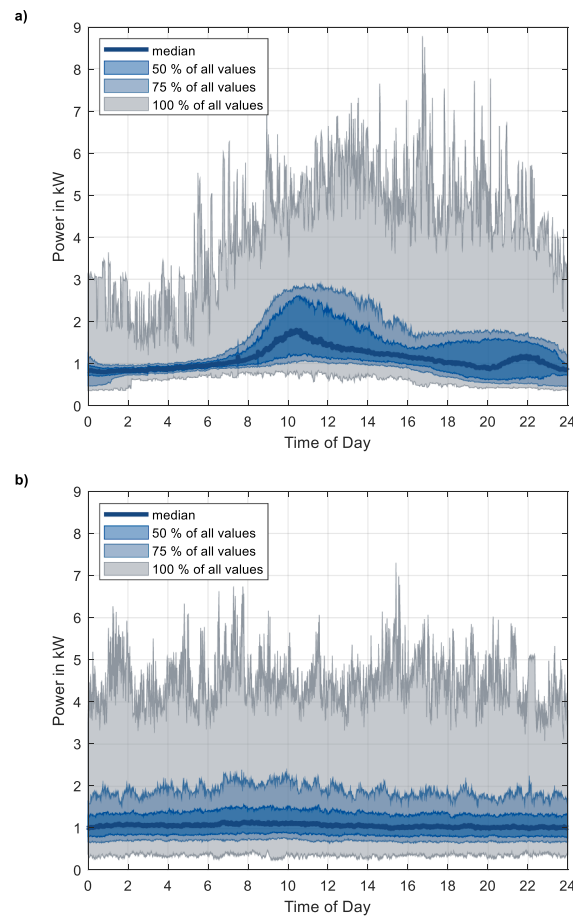


Fig. A9. Fan chart of the household load profiles over the course of the day. 365 values per minute (data for 1 year). The top diagram (a) shows the plot for the original load profile, the bottom diagram (b) the plot for an exemplary level 3 anonymized load profile.

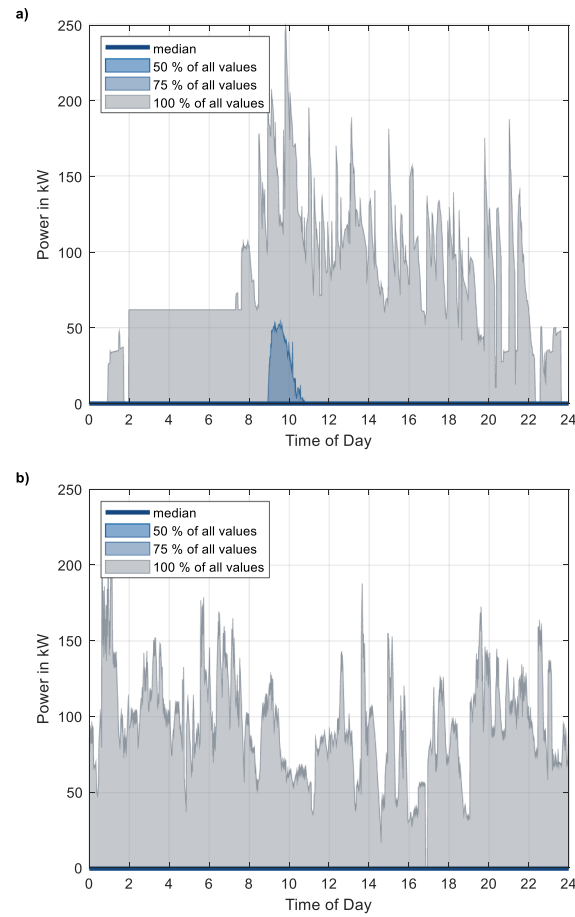


Fig. A10. Fan chart of the EV charging station load profiles over the course of the day. 181 values per minute (data for 6 months). The top diagram (a) shows the plot for the original load profile, the bottom diagram (b) the plot for an exemplary level 3 anonymized load profile.

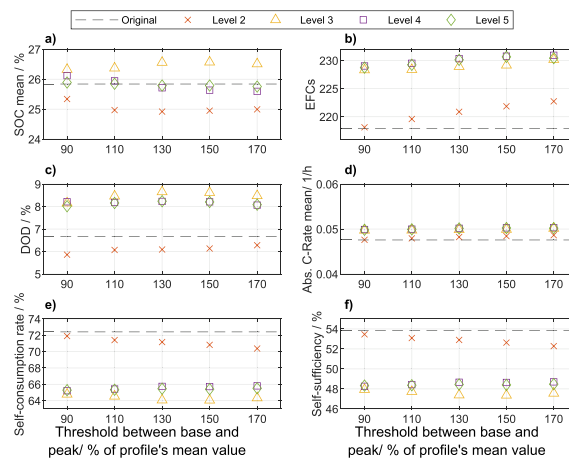


Fig. A11. Sensitivity analysis of the SCI use case with household load profiles: Threshold between base and peak sequences. Threshold defined as percentage of profile's mean value. The levels' values are the median values of 100 simulations. The subplots show the absolute values of the results in the levels.

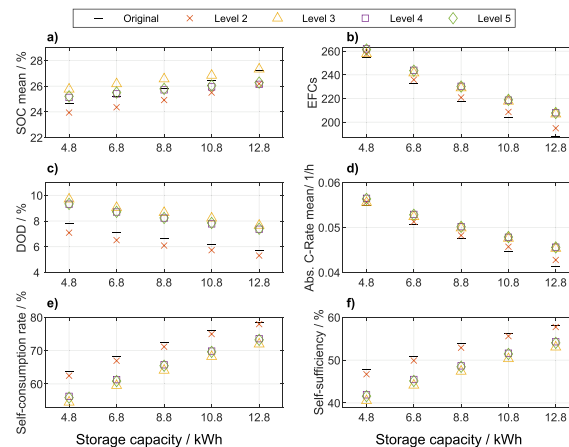


Fig. A12. Sensitivity analysis of the SCI use case with household load profiles: Storage capacity. The levels' values are the median values of 100 simulations. The subplots show the absolute values of the results in the levels.

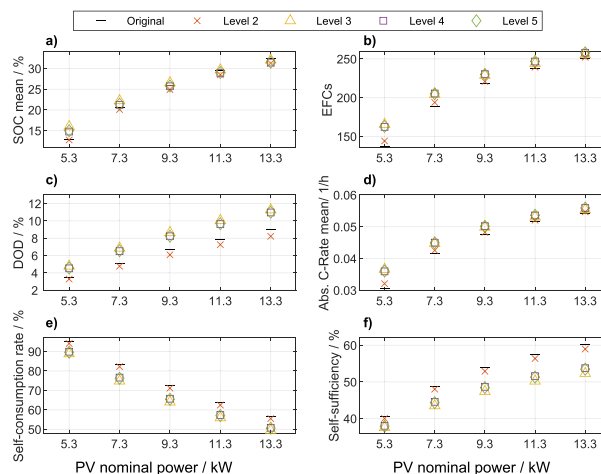


Fig. A13. Sensitivity analysis of the SCI use case with household load profiles: Nominal PV power. The levels' values are the median values of 100 simulations. The subplots show the absolute values of the results in the levels.

References

- [1] Reinsel D, Rydning J, Gantz JF. Worldwide global datasphere forecast, 2021-2025: The world keeps creating more data—now, what do we do with it all. IDC Corporate USA 2021.
- [2] Rydning J, Reinsel D. Worldwide Global StorageSphere Forecast, 2021-2025: To Save or Not to Save Data. IDC Corporate USA: That Is the Question; 2021.
- [3] Baidya S, Potdar V, Pratim Ray P, Nandi C. Reviewing the opportunities, challenges, and future directions for the digitalization of energy. *Energy Res Soc Sci* 2021;81:102243. <https://doi.org/10.1016/j.erss.2021.102243>.
- [4] Sovacool BK, Hook A, Sareen S, Geels FW. Global sustainability, innovation and governance dynamics of national smart electricity meter transitions. *Glob Environ Chang* 2021;68:102272. <https://doi.org/10.1016/j.gloenvcha.2021.102272>.
- [5] Robinson-Garcia N, Costas R, van Leeuwen TN. Open Access uptake by universities worldwide. *PeerJ* 2020. <https://doi.org/10.7717/peerj.9410>.
- [6] European Commission - Directorate-General for Research and Innovation. Trends for open access to publications. [October 27, 2022]; Available from: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/trends-open-access-publications_en.
- [7] Proedrou E. A comprehensive review of residential electricity load profile models. *IEEE Access* 2021;12:1114–33. <https://doi.org/10.1109/ACCESS.2021.3050074>.
- [8] Price P. Methods for Analyzing Electric Load Shape and its Variability. California Energy Commission: Lawrence Berkeley National Laboratory; 2010.
- [9] Li H, Wang Z, Hong T, Parker A, Neukomm M. Characterizing patterns and variability of building electric load profiles in time and frequency domains. *Appl Energy* 2021;291. <https://doi.org/10.1016/j.apenergy.2021.116721>.
- [10] Campestrini C, Horsche MF, Zilberman I, Heil T, Zimmermann T, Jossen A. Validation and benchmark methods for battery management system functionalities: state of charge estimation algorithms. *J Storage Mater* 2016;38–51. <https://doi.org/10.1016/j.est.2016.05.007>.
- [11] Wang Y, Chen Q, Hong T, Kang C. Review of smart meter data analytics: applications, methodologies, and challenges. *IEEE Trans Smart Grid* 2019;10(3): 3125–48. <https://doi.org/10.1109/TSG.2018.2818167>.
- [12] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Trans Smart Grid* 2016;7 (1):136–44. <https://doi.org/10.1109/TSG.2015.2409786>.
- [13] Al-Otaibi R, Jin N, Wilcox T, Flach P. Feature construction and calibration for clustering daily load curves from smart-meter data. *IEEE Trans Ind Inf* 2016;12(2): 645–54. <https://doi.org/10.1109/TII.2016.2528819>.
- [14] Park JY, Yang X, Miller C, Arjunan P, Nagy Z. Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. *Appl Energy* 2019;236:1280–95. <https://doi.org/10.1016/j.apenergy.2018.12.025>.
- [15] Trotta G. An empirical analysis of domestic electricity load profiles: Who consumes how much and when? *Appl Energy* 2020;275:115399. <https://doi.org/10.1016/j.apenergy.2020.115399>.
- [16] Czétányi L, Vámos V, Horváth M, Szalay Z, Mota-Babiloni A, Deme-Bélafi Z, et al. Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy Buildings* 2021;252:111376.
- [17] Elahe MF, Jin M, Zeng P. Knowledge-based systematic feature extraction for identifying households with plug-in electric vehicles. *IEEE Trans Smart Grid* 2022; 13(3):2259–68. <https://doi.org/10.1109/TSG.2022.3146556>.
- [18] Aigner DJ, Sorooshian C, Kerwin P. Conditional demand analysis for estimating residential end-use load profiles. *EJ* 1984;5(3). <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol5-No3-6>.
- [19] Bartels R, Fiebig DG, Garben M, Lumsdaine R. An end-use electricity load simulation model. *Util Policy* 1992;2(1):71–82. [https://doi.org/10.1016/0957-1787\(92\)90055-N](https://doi.org/10.1016/0957-1787(92)90055-N).
- [20] Pensa RG, Monreale A, Pinelli F, Pedreschi D. Pattern-preserving k-anonymization of sequences and its application to mobility data mining. In: *International*

- Workshop on Privacy in Location-Based Applications PiLBA'08. CEUR-WS.org; 2008. p. 44–60.
- [21] Widén J, Wäckelgård E. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Appl Energy* 2010;87(6):1880–92. <https://doi.org/10.1016/j.apenergy.2009.11.006>.
 - [22] Richardson I, Thomson M, Infield D, Clifford C. Domestic electricity use: a high-resolution energy demand model. *Energy Build* 2010;42(10):1878–87. <https://doi.org/10.1016/j.enbuild.2010.05.023>.
 - [23] Efthymiou C, Kalogridis G. Smart Grid Privacy via Anonymization of Smart Metering Data. In: In: 2010 First IEEE International Conference on Smart Grid Communications. IEEE; 2010. p. 238–43.
 - [24] Ogasawara E, Martinez LC, de Oliveira D, Zimbrão G, Pap GL, Mattoso M. Adaptive Normalization: A novel data normalization approach for non-stationary time series. In: The 2010 International Joint Conference on Neural Networks (IJCNN). IEEE; 2010. p. 1–8.
 - [25] Shou L, Shang X, Chen K, Chen G, Zhang C. Supporting pattern-preserving anonymization for time-series data. *IEEE Trans Knowl Data Eng* 2013;25(4): 877–92. <https://doi.org/10.1109/TKDE.2011.249>.
 - [26] Jambagi A, Kramer M, Cheng V. Residential electricity demand modelling: Activity based modelling for a model with high time and spatial resolution. In: 2015 3rd International Renewable and Sustainable Energy Conference (IRSEC). IEEE; 2015. p. 1–6.
 - [27] Müller M, Biedebach F, Reinhard J. Development of an integrated simulation model for load and mobility profiles of private households. *Energies* 2020;13(15): 3843. <https://doi.org/10.3390/en13153843>.
 - [28] Han M, Johari F, Huang P, Zhang X. Generating hourly electricity demand data for large-scale single-family buildings by a decomposition-recombination method. *Energy Built Environ* 2022. <https://doi.org/10.1016/j.enbenv.2022.02.011>.
 - [29] Grandjean A, Adnot J, Binet G. A review and an analysis of the residential electric load curve models. *Renew Sustain Energy Rev* 2012;16(9):6539–65. <https://doi.org/10.1016/j.rser.2012.08.013>.
 - [30] Li H, Yeo JH, Bornsheuer AL, Overbye TJ. The creation and validation of load time series for synthetic electric power systems. *IEEE Trans Power Syst* 2021;36(2): 961–9. <https://doi.org/10.1109/TPWRS.2020.3018936>.
 - [31] Pinceti A, Sankar L, Kosut O. Synthetic Time-Series Load Data via Conditional Generative Adversarial Networks. In: 2021 IEEE Power & Energy Society General Meeting (PESGM). IEEE; 2021. p. 1–5.
 - [32] Sweeney L. k-anonymity: A model for protecting privacy. *Int. J. Unc. Fuzz. Knowl. Based Syst.* 2002;10(05):557–70. <https://doi.org/10.1142/S0218488502001648>.
 - [33] Machanavajjhala A, Kifer D, Gehrke J, Venkatasubramanian M. L-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Discov Data* 2007;1(1):3. <https://doi.org/10.1145/1217299.1217302>.
 - [34] Savov K-K, Stoyanov P, Stanev R, Stoilov D. Analysis of errors in distribution networks power losses calculations with relation to the time discretization intervals. In: In: 2017 15th International Conference on Electrical Machines, Drives and Power Systems (ELMA). IEEE; 2017. p. 42–6.
 - [35] Prasser F, Kohlmayer F. Putting Statistical Disclosure Control into Practice: The ARX Data Anonymization Tool. In: Gkoulalas-Divanis A, Loukides G, editors. *Medical Data Privacy Handbook*. Cham: Springer International Publishing; 2015. p. 111–48.
 - [36] Pinceti A, Sankar L, Kosut O. Generation of Synthetic Multi-Resolution Time Series Load Data; 2021.
 - [37] Schaefer EW, Hoogsteen G, Hurink JL, van Leeuwen RP. Sizing of hybrid energy storage through analysis of load profile characteristics: a household case study. *J Storage Mater* 2022;52:104768. <https://doi.org/10.1016/j.est.2022.104768>.
 - [38] Kucevic D, Tepe B, Englberger S, Parlikar A, Mühlbauer M, Böhlen O, et al. Standard battery energy storage system profiles: Analysis of various applications for stationary energy storage systems using a holistic simulation framework. *J Storage Mater* 2020;28. <https://doi.org/10.1016/j.est.2019.101077>.
 - [39] Martins R, Hesse H, Jungbauer J, Vorbuchner T, Musilek P. Optimal component sizing for peak shaving in battery energy storage system for industrial applications. *Energies* 2018;11(8):2048. <https://doi.org/10.3390/en11082048>.
 - [40] Linszen J, Stenzel P, Fleer J. Techno-economic analysis of photovoltaic battery systems and the influence of different consumer load profiles. *Appl Energy* 2017; 185:2019–25. <https://doi.org/10.1016/j.apenergy.2015.11.088>.
 - [41] Englberger S, Jossen A, Hesse H. Unlocking the potential of battery storage with the dynamic stacking of multiple applications. *Cell Rep Phys Sci* 2020;1(11): 100238. <https://doi.org/10.1016/j.xcrp.2020.100238>.
 - [42] Figgner J, Hecht C, Haberschus D, Bors J, Spreuer KG, Kairies K-P et al. The development of battery storage systems in Germany: A market review (status 2022). arXiv; 2022.
 - [43] Naumann M, Spingler FB, Jossen A. Analysis and modeling of cycle aging of a commercial LiFePO₄/graphite cell. *J Power Sources* 2020;451:227666. <https://doi.org/10.1016/j.jpowsour.2019.227666>.
 - [44] Collath N, Tepe B, Englberger S, Jossen A, Hesse H. Aging aware operation of lithium-ion battery energy storage systems: a review. *J Storage Mater* 2022;55: 105634. <https://doi.org/10.1016/j.est.2022.105634>.
 - [45] Schmalstieg J, Käbitz S, Ecker M, Sauer DU. A holistic aging model for Li(NiMnCo) O₂ based 18650 lithium-ion batteries. *J Power Sources* 2014;257:325–34. <https://doi.org/10.1016/j.jpowsour.2014.02.012>.
 - [46] Naumann M, Schimpe M, Keil P, Hesse H, Jossen A. Analysis and modeling of calendar aging of a commercial LiFePO₄/graphite cell. *J Storage Mater* 2018;17: 153–69. <https://doi.org/10.1016/j.est.2018.01.019>.
 - [47] Moshövel J, Kairies K-P, Magnor D, Leuthold M, Bost M, Gähns S, et al. Analysis of the maximal possible grid relief from PV-peak-power impacts by using storage systems for increased self-consumption. *Appl Energy* 2015;137:567–75.
 - [48] Truong C, Naumann M, Karl R, Müller M, Jossen A, Hesse H. Economics of residential photovoltaic battery systems in Germany: the case of Tesla's powerwall. *Batteries* 2016;2(2):14. <https://doi.org/10.3390/batteries2020014>.
 - [49] Mühlbauer M, Böhlen O, Danzer MA. Analysis of power flow control strategies in heterogeneous battery energy storage systems. *J Storage Mater* 2020;30:101415. <https://doi.org/10.1016/j.est.2020.101415>.
 - [50] Beck T, Kondziella H, Huard G, Bruckner T. Assessing the influence of the temporal resolution of electrical load and PV generation profiles on self-consumption and sizing of PV-battery systems. *Appl Energy* 2016;173:331–42. <https://doi.org/10.1016/j.apenergy.2016.04.050>.
 - [51] Figgner J, Haberschus D, Kairies K-P, Wessels O, Tepe B, Sauer DU. Wissenschaftliches Mess- und Evaluierungsprogramm Solarstromspeicher 2.0: Jahresbericht 2018 2018. <https://doi.org/10.13140/RG.2.2.30057.19047>.
 - [52] Python Software Foundation. tkinter — Python interface to Tcl/Tk. [October 27, 2022]; Available from: <https://docs.python.org/3/library/tkinter.html>.
 - [53] Möller M, Kucevic D, Collath N, Parlikar A, Dotzauer P, Tepe B, et al. SimSES: A holistic simulation framework for modeling and analyzing stationary energy storage systems. *J Storage Mater* 2022;49. <https://doi.org/10.1016/j.est.2021.103743>.
 - [54] Notton G, Lazarov V, Stoyanov L. Optimal sizing of a grid-connected PV system for various PV module technologies and inclinations, inverter efficiency characteristics and locations. *Renew Energy* 2010;35(2):541–54. <https://doi.org/10.1016/j.renene.2009.07.013>. <https://gitlab.lrz.de/open-ees-ses/loadpat>.
 - [55] Tepe B. LoadPAT: Load Profile Anonymization Tool. TUM Chair of Electrical Energy Storage Technology; 2022.
 - [56] Schuster SF, Bach T, Fleder E, Müller J, Brand M, Sextl G, et al. Nonlinear aging characteristics of lithium-ion cells under different operational conditions. *J Storage Mater* 2015;1:44–53.