# SimBench: Open source time series of power load, storage and generation for the simulation of electrical distribution grids

Christian Spalthoff, Fraunhofer IEE, Kassel, Germany, christian.spalthoff@iee.fraunhofer.de Džanan Sarajlić, TU Dortmund University, Dortmund, Germany Chris Kittl, TU Dortmund University, Dortmund, Germany Simon Drauz, Fraunhofer IEE, Kassel, Germany Tanja Kneiske, Fraunhofer IEE, Kassel, Germany Christian Rehtanz, TU Dortmund University, Dortmund, Germany Martin Braun, Fraunhofer IEE/Universität Kassel, Kassel, Germany

# Abstract

In operation and planning of electrical grids, it is essential to account for temporal fluctuation of produced and consumed electric energy. Time series based studies often use standardized load profiles for this, which, however, cannot accurately represent the individual peak patterns in load and demand and their random combinations. As part of the SimBench project, we developed a dataset of energy time series to be assigned as individual profiles to grid nodes in high voltage (HV), medium voltage (MV) and low voltage (LV) grids to calculate local power flows in a more realistic way. Load profiles are classified and assigned to categories on the basis of similarity to standard load profiles to represent a broad range of energy users and generation profiles were created using weather data and an agent-based simulation tool. The subset presented in this paper comprises different 77 one-year-profiles with a 15 minute resolution, containing commercial consumers, household consumers, storage, and production units based on real measurements from Germany with a focus on MV and LV levels.

# 1 Introduction

## 1.1 Motivation

Energy grids around the world currently face a fundamental transformation: Structures change from centralistic design and monodirectional power flows to a complex network of distributed prosumers with volatile and locally coinciding peaks in demand and generation. The planning and operation of stable and cost-efficient future power grids requires to accurately predict the proximity to operational boundaries of such systems also when actual measurement data is not available. Traditionally, grid computations are usually derived from a worst-case scenario based on predicted maximum coincident values of load and generation, which can lead to over-dimensioning of infrastructure. Time series based approaches can offer a higher degree of realism, but only when based on representative inputs.

The SimBench project aims to provide a comprehensive set of power grid models and time series to serve as a basis for grid calculations, especially where measured data is not available. Those calculations need profiles that exhibit the random and short-scale fluctuations of real measurements, but are representative for the units typically present in the respective power grid.

Further requirements are:

- individuality, i.e. a large bandwidth of different real types
- inclusion of both active and reactive power

- temporal extend of one full year, without repetition of generic weekdays
- temporal resolution of 15 minutes
- suitability for both LV and MV grid levels

While solar and wind power generation can be modelled relatively well based on weather data, demand profiles exhibit an inherent inhomogeneity based on several external factors. Especially commercial consumers vary strongly in opening hours or technology used, while a high variance in peak timing induced by variations in the occupants behaviour dominates household loads [1].

In time series based grid calculations, these external factors are hard to predict and therefore especially relevant. Thus, we chose to compose a dataset based on real measurements instead of synthetically modelled profiles.

## 1.2 Standard load profiles

Power utilities commonly use generic load profiles to group commercial customers with similar load shapes into categories. In Germany, standard load profiles (SLPs) are used for trading energy of consumers with annual consumption below 100 MWh [2]. Here, the most commonly used profile set is developed by the German Association of Energy and Water Industries (BDEW, formerly VDEW) [3]. It comprises eleven aggregated profiles, one for residential consumers (H0), three for agricultural (L0-L3), and seven for commercial consumers with different opening hours (G0-G6). They are differentiated into workdays, Saturdays and Sundays as well as three seasonal categories winter, summer, and transitional. The set includes two profiles for street lightning (B0) and band load (G7). Lately, with more available customer data and the roll-out of smart meters, the forecast accuracy of these profiles has been challenged [4], and in the future direct measurements will most likely gain more impact. However, the long practical experience in categorizing customer loads to SLPs and the readily available data for SLP types usually present in various grid types makes them a valuable basis for the development of synthetic grids where loads have to be extrapolated from incomplete data.

We choose to follow the classification into load categories provided by BDEW profiles, and selected a set of measured load time series based on similarity to these SLPs.

### 1.3 Classification of measured load profiles

Especially with commercial customers, a large amount of load measurement data is already collected. The private nature of this data, however, requires careful anonymization of individual customers before using or publishing. In our case, we had access to a large number of commercial "registering load measurement" (RLM) customers, but no information of the names or types of business associated with them. We decided to classify these profiles based on their similarity to known SLP categories, and select those with high similarity as "typical" profiles. These can be used as more realistic substitutes for SLP profiles in our dataset.

The classification, prediction and synthetic generation of consumer load profiles have been the topics of several scientific studies. Many of theses are based on clustering analyses of measured load profiles based on a variety of features (e.g. [5]). This approach is especially useful in the unbiased finding of actual centroids in a specific set of measured data, but has the disadvantage that the specific properties of the examined dataset might not be applicable for different or generic datasets.

In general, the large number of data points in a yearlong load profile and their periodic nature makes the selection of relevant features for analysis or comparison challenging. Discarding the time domain of the signal and only looking at parameters like average and standard deviation can reduce the complexity [6]. On the other hand, some approaches treat every single data point of a time series as independent and equivalent features [7], which increases the demand on computation power.

A common solution to reduce complexity is to group days of a year by similar weekdays or seasons, and calculating the average of these.

Since one of the most important factors for grid overload is simultaneity of peaks, we intend to classify profiles in a way that focuses strongly on the daily load profile shape, i. e. peak timing. This rules out methods that discard temporal information, like comparing total power consumption. For many categories of consumers, it is plausible to assume a high degree of scale invariance, so that normalized load profile shape is similar for small and large businesses, with only smoother and more predictable peaks in larger units. We base our analysis on this assumption and compared SLP profiles, which are typically used below 100 MWh annual consumption, with RLM customers who usually have higher annual consumption. The RLM dataset composition is likely to be dissimilar to the average smallscale customers for which we want to select profiles. Because of this, we use SLP-based classification instead of clustering for profile selection.

It is often the case that individual peaks in measured profiles exhibit similar sizes and shape, but vary in time. Typical examples for this are dairy farms, which use up to half of their energy consumption for milking equipment during short periods twice a day [8], which are often slightly shifted during the year. Methods calculating differences between hourly measurements without regarding the neighbouring data often tend to classify slightly shifted peak values very dissimilar.

To tackle this problem, a wide range of comparison methods have been presented in the literature for the shapebased clustering and classification of time series (for a review, see [9]). We implemented a two dimensional method that reduces data to a week average and uses a relaxed Euclidian geometric distance measurement.

# 2 Methods

## 2.1 Date adjustments

As we use data from a variety of sources and years, it is not possible to restrict profiles to measurements from a single year. Since especially the dates of weekdays and public holidays have a strong influence on load profiles, we use 2016 as reference year for the dataset and adjust data from different years by shifting days so that they correspond to the nearest weekday. We create data for the leap day or public holidays by duplicating the next similar weekday. In the weather based generation series, weekdays are not adjusted, but we reconstruct missing values for the leap day 2016 by averaging values formed from the previous and subsequent days.

## 2.2 Generation time series

The generation time series for photovoltaics (PV), wind energy and biomass generated for the SimBench dataset are created using the agent-based simulation tool for optimized grid expansion planning SIMONA developed by the TU Dortmund University [10], [11], further details in [12], [13]. This multi-agent simulation (MAS) tool is designed as a bottom-up model of an electrical energy system, which is freely configurable and takes into account operational degrees of freedom, interdependencies of individuals taking part in the energy system and innovative network resources. A comparison of measured data and time series created can be found in [14]. SIMONA includes models for PV, wind power and biomass plants. These plant models receive real weather data of Germany from the German Weather Service (DWD) in 2011 for Wind and 2012 for PV time series as input data. We use different geographical

locations for plants to include locally varying weather conditions in time series generation. The locations were selected taking into account the geographical location of the SimBench HV grid models. The locations selected for the wind time series correspond approximately to the locations of existing wind farms.

The general geographical location of the measuring points is as follows:

- North Sea and Baltic Sea
- North Sea coast, Baltic Sea coast and eastern Germany
- 3 locations around Hannover
- 3 locations around Lübeck

We also include three time series based on real measurements from 2017 of three hydropower plants.

For 2011, 2012 and 2017 generation data, we adjusted the time axis to 2016 as described above. Time series values are given as relative active power values, which can be applied by scaling to different plant sizes through multiplication by the installed power of the plants from the SimBench data set. For the PV time series it has to be taken into account that in some cases the rated power of the asset is not reached due to the volatility of the weather.

#### 2.3 Load profiles

Commercial load profiles are based on a dataset of anonymized commercial RLM profiles from the year 2016 provided by the German distribution grid operator Syna GmbH. Household profiles are based on the IZES 2010 dataset of smart meter measurements [15].

Single commercial load profiles with data inconsistencies like missing values are discarded (~10%). Out of the remaining 2539 profiles, 622 (~25%) show an annual consumption below 100 MWh, above which individual metering is required in Germany. Some of the profiles show a high amount of reactive power, sometimes exceeding active power. While different consumers might differ in their power factors ( $\cos \varphi$ ) due to differences in electrical appliances, we decide to treat profiles with power factor values lower than 0.8 as unusual enough to remove these (~14%) from the dataset. From this test dataset we select representative profiles by matching them to SLPs. Used for comparison are the BDEW profiles for the state of Hessen provided for 2016 on the website of the distribution grid operator EnergieNetz Mitte [16].

#### 2.4 Matching profiles to SLPs

Test profiles are reduced to a seven-day-week profile by averaging similar days (**Fig. 1**). Even though SLPs use a simpler form of five similar weekdays followed by a Saturday and Sunday, we decided against this form because this leads to a disproportional high influence of Saturdays and Sundays on matching results. In addition, we wanted to account for the fact that Fridays and Mondays often show a pronounced difference from other weekdays in many measurements. The resulting week profiles are normalized to the maximum used power.

In the resulting weekly profiles of 672 data points each, for every point in the test profile the Euclidean distance to the nearest point in the reference profile is determined, and vice versa. For scale, load and time factors are chosen so that 100% maximum load difference corresponds to 24h time difference.

$$d^{\text{TEST,REF}} = \sum_{m=1}^{672} \min_{n\{1,\dots,672\}} \sqrt{\left(t_m^{\text{TEST}} - t_n^{\text{REF}}\right)^2 + \left(e_m^{\text{TEST}} - e_n^{\text{TEST}}\right)^2}$$
$$d^{\text{REF,TEST}} = \sum_{m=1}^{672} \min_{n\{1,\dots,672\}} \sqrt{\left(t_m^{\text{REF}} - t_n^{\text{TEST}}\right)^2 + \left(e_m^{\text{REF}} - e_n^{\text{TEST}}\right)^2}$$

with d=distance measure, t=time value and e=load value. The sum of both values gives the overall distance measurement.

$$d = d^{\text{TEST,REF}} + d^{\text{REF,TEST}}$$

We compared all test profiles to all reference profiles in this manner. A test profile was chosen as potentially representative for a given SLP when:

*d* was lowest from all test profiles in the dataset, and *d* was lowest from all SLP profiles the test profile was compared with, the latter taking prevalence over the former.



**Fig.1**: Example comparison of a random week (Apr. 04 - Apr.10 2016) for the reference profile G1 (grey), RLM profile G1-A average over all week days (green) and RLM profile G1-A for this individual week (blue). Traces are normalised to yearly maximum amplitude.

In MV grids, the composition of directly connected loads is expected to differ from those connected to LV grids, but there is no universal threshold of consumption that warrants a direct connection to the MV grid. In our commercial load dataset, most consumers used about 150 MWh/a, with single datasets using up to 20 GWh/a. In comparison, the average German power consumption for non-industrial consumers was only between 1-3 MWh in 2010 [17].

We used 0-150 MWh/a, 400-800 MWh/a and above 3.5 GWh/a as classes for low, medium and high consumption, and matched and selected profiles from them individually to offer a broader range of profiles for MV (or HV) consumers.

### 2.5 Storage profiles

Based on input from distribution grid operators, we identified two relevant storage categories:

Residential PV-battery-systems represent a system with least possible grid interaction that runs on an internal operating schedule in order to maximise household self-consumption. Inputs are SimBench household loads (H0-A to H0-L) combined with PV profiles (PV1 to PV8, respectively) and battery storages. In our model, battery and PV size are proportional to total household power consumption. The resulting time series combine power production and demand into a prosumer series.

PV and wind profiles with added storage capacities at MV level could be a future option for grid support use: Production peaks shaved by storing power in battery systems to prevent grid overload. In Germany, grid operators are given a novel planning opportunity called 'peak shaving' that allows to consider an annual curtailment of maximum 3 % per generating unit during the planning process [18], [19]. We used this to create modified PV and wind generation profiles with a peak-shaving regime that stores the highest 3% of produced power and releases it in times of high demand.

### 2.6 Heat pump profiles

We model a controller based heating system with two thermal storages for domestic hot water and space heating supplied from a heat pump and a peak load heater. The algorithm sets the priority of the domestic hot water above the space heating storage, i.e. if both storages are empty, the domestic hot water demand is covered first. Both upper and lower limit can be set. As soon as one storage state of charge drops below the lower limit, the storage is refilled until the upper limit is reached. If none of the storages requires a refill, the heating system is turned off.

We model three different heat pump operation modes. The most common modes are bivalent-alternative, bivalentparallel and bivalent-semi-parallel [20]. Common for all modes is the bivalent temperature, i.e. the ambient temperature at which the peak load heater can be activated. In case of the alternative operation mode, the heat pump turns off at a certain temperature and only the peak load heater covers the heat demand, while in case of the parallel operation mode both systems run simultaneously. The semi-parallel mode combines both options.

Heat pumps also differ in their external heat source, with a current and predicted future use of both geo-thermal and air heat pumps [21], so we included both types.

Operation time of a heat pump is strongly dependent on the heat demand in a household and therefore to the time series of ambient temperature. We selected the central region used for PV and wind data (Kassel) as a location for ambient temperature profiles and heat demand profiles for a single family house.

We also include a daily time period when heat pumps are usually blocked by the grid operator based on profiles provided by EAM (11:00 to 12:45)[16].

### 2.7 Electric vehicle charging profiles

We created electric vehicle charging profiles based on a probabilistic combination of four data sets to calculate travel and charging times of single cars. Typical mobility behaviour parameters of German citizens are based on data from the "Mobility in Germany" survey [22]. We deleted implausible data like average velocities above 120 km/h. Car types are based on official sales statistics for Germany from 2013 to 2017 (Federal Motor Transport Authority, cited after [23]), and their individual data sheets provide data for battery types. Typical electrical car charging schedules based on measurements at the Fraunhofer IEE [24] were scaled accordingly assuming different applied power and capacities.

Exemplary profile for single charging processes are selected based on typical car distribution. Car types without available charging profiles (~35%) are randomly assigned to known types.

Battery capacities as well as power consumptions per 100 km travel distance are independently based on the typical existing distribution for electric cars, leading to a broad variety of combinations.

The initial battery state of charge (SOC) is set randomly to between 50 and 100 %

The number of potential trips and therefore charging occasions per weekday are determined from behaviour parameters and grouped into five trip categories: Workplace, home, into the city and out of the city.

Weekday and trip category determine the probability distribution of the time of arrival, velocity and distance. These factors are calculated for each charging occasion.

Since not all trips need to be followed by a charging process, we model the users charging decisions based on intervals between trips. If the car stands longer than three hours the probability to charge is 60 %. If a car stands less than three hours, the probability is reduced to 20%. If SOC falls below 20 %, the electrical car is always charged at the end of the trip regardless of time.

These car based time series are the basis for modelling the point of intercept between the electrical grid and the electrical vehicle at the charging stations. Forecasts for 2024 and 2034 predict that most charging stations have a power consumption of 3.7, 11.0, 22.0 and 50.0 kW, with smaller stations more common especially in residential environments. The profiles for the dataset contain household and workplace charging profiles for different charging station

powers, with charging types and car types taken from independent samples. Time series are calculated for the year 2016.

## **3** Results

### 3.1 Load profile matching

For all commercial profile classes (Type G1-G6), we could find matching week profiles from our RLM dataset. Lowest distances are found in profiles matched to SLP categories with relatively homogeneous loads like G3, G4 or L2, while profiles with accentuated peaks at certain day hours like L1 or G2 were less similar. G1, which represents a common business type with 8:00-18:00 working hours was also matched most often (**Fig. 2**). This profile is traditionally used for office buildings and stores, which make up the biggest faction in German non-industrial businesses. [17].



**Fig. 2**: Distribution of profiles (high, medium and low consumption commercial and household) after matching to SLPs. The profiles with lowest distances are selected for the datasets.

For medium and high consumption profiles, G1 was less often matched, and profiles with high baseline load appeared more often (data not shown). For high consumption customers, we included one profile (H0-H) that was matched to H0 due to a uniform load on all weekdays combined with a high base load. Several profiles in the dataset did not match standard SLPs but exhibited band load characteristics, either for the whole week or for working days only. One example of these profiles each was included as "BL-H" and "WB-H" in the profile set.

In the household dataset, the only commercial profile matched was G6, which represents weekend usage. In addition, the agricultural profiles L0-L2 were often matched. Since these profiles represent a high demand in the early morning, evening and weekend hours, their similarity to many household profiles is plausible.

When examined on a 15-minute-basis instead of weekly averages, even profiles with a low distance exhibit strong differences to their matched SLPs. Both short peaks and day-specific variations are much more pronounced (**Fig.** 1). This effect is more visible in the H and L type profiles, which show a higher overall variation in peak timing.

#### 3.2 Composition of the SimBench dataset

The dataset (**Fig. 3**) contains eight time series each for PV (PV-), elven for wind (WK-) and five for biomass (BM-) generation as well as three time series for hydroelectric generation (H-). Of the 15 storage series included, ten (PB-1 to WB-5 and WB-1 to WB-5) are based on peak-shaved generation series, and five series (PB-A to -L) are house-hold-based prosumer series. These time series and the five heat pump series (WP-1 to WP-5) are not divided into high, medium and low classes.

Of the five heat pump profiles, two (WP-1 and -2) are geothermal heat pumps and three (WP-3 to -5) are air heat pumps. WP-1 and -3 are operated parallel, WP-2 and -4 are operated in alternative mode and WP-5 is operated in semiparallel mode.

Five electrical car charging series each are provided for household (E#-HA to -HC) and workspace charging. (E#-GA and -GB) These profiles vary by max charging power available at the station (#), and model 3.7, 11.0, 22.0 and 50.0 kV chargers.

The consumer profiles consist of five household based time series (H0-A to H0-L) and 22 RLM based series. The latter are sorted by matching SLP type (G0- to L2-), with last letters –A to C indicating low consumption, -M medium consumption, and -H high consumption customers. In addition, a fulltime band load (BL-H) and weekday band load dominated (BW-H) profiles are included.



**Fig.3:** Profiles selected for the SimBench LV and MV dataset. Hatched boxes represent household type customers, while full boxes represent commercial type customers.

Timestamps for data are in CET (UTC+1) including German state holidays. All active power values are normalized to the maximum active power value. Since every grid node in SimBench grids comes with fixed scaling factors for active and reactive power, time series reactive power values are scaled up to meet the original load factor when multiplied with these factors.

With its input from variable sources, this dataset aims to provide the user with a high variety of time series, while keeping the number of individual profiles to download and manage at a practical level. The modular and interchangeable format of the data also allows add individually measured data or new profile classes in the future.

The SimBench grids and time series can be downloaded at www.simbench.de

# 4 Acknowledgements

We acknowledge the support of our work by the German Ministry for Economic Affairs and Energy and the Projekträger Jülich within the project "SimBench". The authors are responsible for the content of this publication.

We thank G. Naurath of Syna GmbH and J. Dickert of ENSO NETZ GmbH for providing data.

## 5 Literature

- L. Lutzenhiser, 'Social and behavioral aspects of energy use', *Annu. Rev. Energy Environ.*, vol. 18, no. 1, pp. 247–289, 1993.
- [2] Deutscher Bundestag, Verordnung über den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzzugangsverordnung-StromNZV). Bonn, 2005.
- [3] B. Schieferdecker, C. Funfgeld, H. Meier, and T. Adam, 'Repräsentative VDEW-lastprofile', VDEW-Mater. M-2899 Frankf., 1999.
- [4] D. Scholz and F. Müsgens, 'How to improve standard load profiles: Updating, regionalization and smart meter data', in 2017 14th International Conference on the European Energy Market (EEM), 2017, pp. 1–6.
- [5] O. Motlagh, A. Berry, and L. O'Neil, 'Clustering of residential electricity customers using load time series', *Appl. Energy*, vol. 237, pp. 11–24, Mar. 2019.
- [6] J. Henze, T. Kneiske, M. Braun, and B. Sick, 'Identifying Representative Load Time Series for Load Flow Calculations', in *International Workshop on Data Analytics for Renewable Energy Integration*, 2017, pp. 83–93.
- [7] G. Chicco, 'Overview and performance assessment of the clustering methods for electrical load pattern grouping', *Energy*, vol. 42, no. 1, pp. 68–80, Jun. 2012.
- [8] L. Murgia, M. Caria, and A. Pazzona, 'Energy use and management in dairy farms.', in *International Conference, Innovation Technology to Empower Safety, Health and Welfare in Agriculture and Agro-Food Systems, 15-17 September 2008, Ragusa, Italy,* 2008.
- [9] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, 'Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures', *Proc VLDB Endow*, vol. 1, no. 2, pp. 1542–1552, Aug. 2008.
- [10] J. Kays and C. Rehtanz, 'Planning process for distribution grids based on flexibly generated time series considering RES, DSM and storages', *IET Gener. Transm. Distrib.*, vol. 10, no. 14, pp. 3405– 3412, 2016.
- [11] J. Kays, A. Seack, and U. Häger, 'The potential of using generated time series in the distribution grid planning process', in *Proc. 23rd Int. Conf. Electricity Distribution, Lyon, France*, 2015.
- [12] A. Seack, *Time Series Based Distribution Grid Planning Considering Interaction of Network Participants with a Multi-agent System.* sierke Verlag, 2016.
- [13] J. Kays, Agent-based simulation environment for improving the planning of distribution grids. Sierke Verlag, 2014.
- [14] A. Shapovalov, C. Kittl, C. Rehtanz, L. Jendernalik, A. C. Schneider, and D. Giavarra, 'Application of time-resolved input data for smart grid simulation', *CIRED-Open Access Proc. J.*, vol. 2017, no. 1, pp. 2106–2109, 2017.

- [15] T. Tjaden, J. Bergner, J. Weniger, V. Quaschning, and Forschungsgruppe Solarspeichersysteme, 'Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis', Hochsch. Für Tech. Wirtsch. HTW Berl., 2015.
- [16] EAM GmbH & Co. KG, 'Standardlastprofilverfahren'. [Online]. Available: https://www.energienetzmitte.de/marktpartner/netzzugangnutzung/strom/standardlastprofilverfahren/. [Accessed: 17-Jan-2019].
- [17] B. Schlomann *et al.*, 'Energieverbrauch des Sektors Gewerbe, Handel, Dienstleistungen (GHD) in Deutschland f
  ür die Jahre 2011 bis 2013', 2015.
- [18] Deutscher Bundestag, 'Gesetz über die Elektrizitäts-und Gasversorgung (Energiewirtschaftsgesetz-EnWG)', *EnWG Vom*, vol. 7, 2005.
- [19] C. Wagner, Integration und Bewertung der Spitzenkappung als Planungsgrundsatz zur wirtschaftlichen Netzentwicklung in Mittelspannungsnetzen. Shaker Verlag, 2019.
- [20] K. Klein, K. Huchtemann, and D. Müller, 'Numerical study on hybrid heat pump systems in existing buildings', *Energy Build.*, vol. 69, pp. 193–201, 2014.
- [21] M. Braun *et al.*, 'Verteilnetzstudie Hessen'. [Online], 2018.
- [22] R. Follmer *et al.*, 'Mobilität in Deutschland 2008', *Methodenbericht Bonn Berl.*, 2010.
- [23] 'Neuzulassungen von Elektroautos nach Marke/Modellreihe bis 2018', *Statista*. [Online]. Available: https://de.statista.com/statistik/daten/studie/224041/umfrage/neuzulassungenvon-elektroautos-nach-marke-modellreihe/. [Accessed: 17-Jan-2019].
- [24] J. Prior, 'Testverfahren zur Bestimmung des elektrischen Verhaltens von Batteriesystemen in Elektrofahrzeugen', Universität Kassel, 2017.