

GridCreator: An Open-Source Toolbox for Synthetic Low-Voltage Distribution Grids in Germany

Matthias Behr¹ , Gunter Grimm¹ , Rebecca Hofmann¹ , Mirko Schäfer¹ ,
Ramiz Qussous¹ , Markus Schumacher² , and Anke Weidlich¹ 

¹INATECH, University of Freiburg, Germany

²Institute of Physics, University of Freiburg, Germany

*Correspondence: Mirko Schäfer, mirko.schaefer@inatech.uni-freiburg.de

Abstract. GridCreator is an open-source Python tool used to generate approximations of real-world low-voltage distribution grids across Germany. The tool combines public data sources and open-source Python packages to enable the creation of entire low-voltage grids for user-defined areas with minimal effort. Technical grid data is enriched with socio-economic information from the 2022 census survey. Based on correlations between socio-economic census data and public statistics on the occurrence of generation and demand units, these technologies are distributed across the nodes of the grid area. Finally, the tool generates demand and generation time series, which approximate real-world patterns and provides an interface to network calculation software and energy system models.

Keywords: Low-Voltage Grids, Open-Source, Socio-Economic Data

1. Introduction

Existing low-voltage (LV) grids were originally designed to transport electricity unidirectionally from the transformer to end consumers. The increasing penetration of household flexibilities, such as heat pumps (HPs), electric vehicles (EVs) and battery storages, is altering load profiles and overall consumption patterns. In parallel, the growing number of rooftop solar photovoltaic (PV) systems feeding directly into LV grids is changing load flow characteristics. Both shifts in consumer behavior and weather-dependent generation lead to increasing prediction uncertainties and pose challenges for the planning and operation of LV distribution grids. Potential consequences include line and transformer overloading as well as power quality issues. Consequently, LV grids must be adapted to these new conditions [1], [2].

Existing studies on LV distribution grids often rely either on a limited set of synthetic networks [3] or on standardized IEEE reference grids [4]. However, in order to derive generalizable and practically relevant insights, it is crucial to examine potential solutions in the areas of market design, network tariffs, control mechanisms, and price incentives using realistic LV distribution grids. In this contribution we introduce GridCreator as a Python tool that enables the generation of low voltage grids for any region in Germany using only input coordinates as a bounding box (bbox). It creates realistic grids enriched with socioeconomic and technological data and supplements them with asset level generation and demand time series. The entire workflow runs automatically and provides a complete framework that outputs a fully assembled grid including all time series, making it possible to perform simulations that closely approximate real world conditions.

Table 1. Overview of reviewed tools.

Name	Focus	Area Covered	Open-Source	Documentation	LV Grid Generation	Infrastructure and Socio-economic data	Timeseries Generation	Optimal Capacity Expansion	Optimal Dispatch
PyPSA	Multi-energy system modelling	World	✓	✓				✓	✓
PyPSA distribution	Island multi-energy systems	World	✓	✓	✓		✓	✓	✓
City Energy Analyst	Designing low-carbon and energy-efficient cities	World	✓			✓			
SimBench	Representative LV and MV grids			✓			✓		
SING	Distribution grid generation	World	✓	✓	✓				
PyLovo	LV grid generation	Bavaria	✓	✓	✓				
osmTGmod	Transmission grid generation	Bavaria	✓	✓	✓				
open eGo - ding0	Distribution grid generation	Germany	✓	✓	✓				
RAMP	Demand timeseries generation		✓	✓			✓		
DemandRegio	Regionalization of electricity and gas demand	Germany	✓			✓	✓		
pyCity	Generation of time series		✓	✓			✓		
LoadProfileGenerator	Predefined households time series	Germany	✓	✓					
open eGo - eDisGo	Analysis and expansion of LV networks		✓	✓				✓	✓
PyStatis	Loading <i>census</i> data	Germany	✓	✓		✓			
osmrx	Loading OSM data	World	✓	✓		✓			

2. Related Work

As a first step, various existing tools in the field of distribution grid modeling were analyzed and evaluated with respect to their suitability for the described use case. Table 1 provides an overview of the features offered by each tool.

To the authors' knowledge, no existing tool provides all the functionality required to generate an LV network with realistically assigned generation and demand time series. Therefore, GridCreator is intended to enable the interaction of several individual tools along a modelling pipeline so that their respective strengths can be utilized collectively. The following section discusses each tool with regard to its properties and its specific role within GridCreator.

PyPSA [5] is a widely used framework for energy system modelling and analysis, which motivates to use NetCDF as a common output format for GridCreator. Although *PyPSA* can be used as a framework for grid modelling, it does not provide the functionality for deriving real-world LV grids based on road networks. Conceptionally, *SING* [6] and *ding0* [7] are able to fill this gap, but *SING* was found not to be usable because its repository is no longer maintained, and the applicability of *ding0* is severely constrained due to its reliance on a non-public database. Nevertheless, the developers of *ding0* provide ready-made LV grids for all of Germany, which are incorporated into GridCreator [8]. Another tool considered was *Pylovo* [9], but it is currently limited to generating grids for the state of Bavaria. *osmTGmod* [10] can create grids, but not at the low-voltage level; instead, it is primarily designed for the creation of large-scale transmission grids. *Simbench* [11] does not generate new grids but rather provides standardized, pre-built benchmark networks. *CityEnergyAnalyst* [12] is an open-source tool for modeling urban energy systems, but its complexity and city-wide focus made it unsuitable for generating individual load profiles within the intended workflow. Instead, *pyCity* [13] was selected, as it can produce both load and PV generation time series using only a few input parameters. *Ramp* [14] can create highly detailed household consumption profiles, including all electrical appliances, but this level of granularity was omitted in favor of a simpler and more lightweight approach. *LoadProfileGenerator* [15] provides only predefined German load profiles without supporting custom profile creation, and it

lacks commercial load profiles. Although tools such as *DemandRegio* [16] and *OpenSynth* [17] offer a broader selection of residential or commercial load profiles, *pyCity* was chosen to keep GridCreator user-friendly and to minimize external dependencies.

osmnx [18] as the data source for infrastructure information enables the direct download of OpenStreetMap (OSM) data for specified coordinates. The initial goal was to retrieve *census* data directly via an API, but *PyStatis* [19] provides access only to aggregated statistics rather than the required 100m×100m atlas data, so the data had to be collected manually.

PyPSA-Distribution [20] builds on the PyPSA framework and is designed to facilitate global distribution-grid studies, including grid planning, tariff development, and capacity expansion analyses. However, as the tool was still in an early development stage when GridCreator was initiated, it was not integrated into this work. *eDisGo* [21] is a PyPSA-based tool for distribution grid planning, including grid reinforcement, battery integration, and grid-supportive curtailment. Since GridCreator focuses solely on recreating existing networks, *eDisGo* is not suitable for this purpose, though it can be used for subsequent editing and analysis of the generated networks. Since GridCreator is an open-source tool accessible to all interested users, non-open-source tools were excluded from consideration.

3. Methodology

The GridCreator concept can be divided into five steps (see figure A1 in the appendix). After each step, the user has the option to modify the generated data, for example to deviate from predefined inputs or to represent specific scenario assumptions. The following sections describe the workflow step by step, using an example grid from Opfingen for illustration.

Step 1

Since no suitable open-source tool was available for generating LV networks directly from open data, GridCreator uses the pre-constructed *ding0* buses and lines, provided in a folder structure organized by medium-voltage (MV) supply areas [8]. As GridCreator currently focuses on LV networks, only LV-level components within the user-defined bbox are loaded. Figure 1 shows the Opfingen example, where the bbox was defined using EPSG4326 coordinates from 47.99434°,N to 48.00798°,N (latitude) and from 7.70691°,E to 7.72483°,E (longitude). If no transformer lies within the specified bbox, the bbox is automatically expanded until the nearest transformer is included. The network is then loaded by determining the shortest path through the network from each bus within the bbox to a transformer. All components located along these paths are added to the network, irrespective of whether they lie inside or outside the original bounding box. This ensures that a complete LV network, with all relevant and electrically connected components, is available for subsequent processing.

Step 2

In the second step, OpenStreetMap (OSM) data are assigned to the nearest bus based on coordinates, with nodes within the radius of multiple buses linked to the closest one to avoid duplicates. In preparation for the technology assignments in the third step, all nodes are mapped to the respective state and postal code using their coordinates. All relevant *census* features used to describe the occurrence of generation and demand units are assigned to the nodes using nationwide data aggregated to a 100m×100m raster. This ensures that each node receives the most accurate available socio-economic information for its location. Based on these data, the residential structure of the area can be characterized, enabling the subsequent assignment of PV systems, battery storages, EVs, and HPs.

The selection of *census* features is separately described in Section 4. Assignment is performed

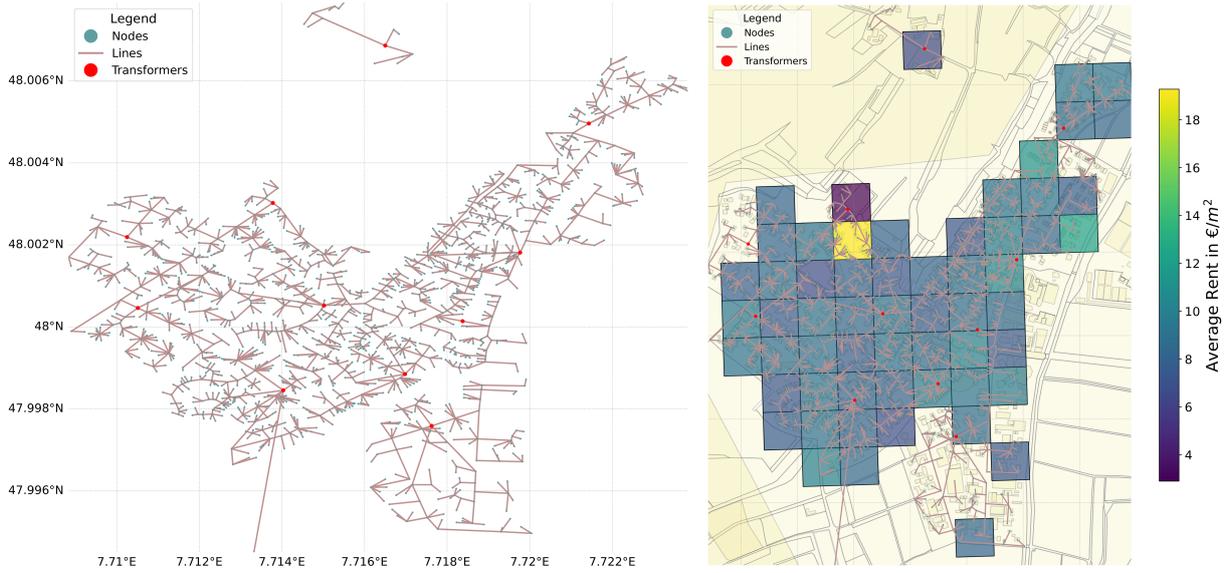


Figure 1. LV ding0 network for the coordinate range (EPSG4326) entered into GridCreator, extending from $47.99434^{\circ},N$ to $48.00798^{\circ},N$ (latitude) and from $7.70691^{\circ},E$ to $7.72483^{\circ},E$ (longitude). The left panel shows the raw ding0 grid after step 1, while the right panel visualizes the average rent per square meter as an example for socio-economic data and infrastructure shapes obtained from OSM as the background; both are assigned to the grid in step 2.

based on coordinates: the center of each raster cell is used to determine the shortest distance to each node, and all relevant data from the cell with the nearest center is then assigned to that node. Figure 1 illustrates the Opfingen network, showing *census* data for the average rent per square meter as a color map over an OSM-based background. It is important to note that some nodes do not lie within any *census* raster cell. This results from the confidentiality requirements of the Federal Statistics Act, which prohibits the publication of information that could reveal individual data [22]. In such cases, the corresponding *census* information is shifted to adjacent raster cells. Consequently, nodes located in these omitted cells do not contain direct information for the respective feature.

Step 3

The distribution of generation and consumption units (GCU) is based on downscaling of available statistical data. The *MaStR* [23] provides information on the number, peak power and orientation of PV systems and on the number and size of battery storages in each postal code area. The *Kraftfahrtbundesamt* [24] publishes data on the number of EVs registered in each registration district, and the number of HPs installations per federal state is available from [25]. Using this information together with *census* data, the likelihood of each bus hosting a GCU can be estimated. Probabilities are calculated by weighting the selected *census* features according to equation (1). The probability W for node j to accommodate a certain *GCU* is computed as

$$W_{GCU}^j = \frac{\sum_i f_i \cdot z_i^j}{\sum_i f_i \cdot z_i^{j_{tot}}}, \quad (1)$$

where z_i^j denotes the value of *census* feature i at node j , f_i is the weighting factor for *census* feature i , and $z_i^{j_{tot}}$ is the value of the *census* feature for the total area corresponding to the respective GCU. The resulting value represents the relative probability of the GCU occurring at node j . The factors f_i were calculated in a previous step, as separately discussed in Section 5. After calculating relative probabilities, GCUs are assigned to buses within each raster cell by

multiplying the probability with the total number of GCUs in the smallest enclosing area and then randomly distributing the GCUs among the buses until the total number is reached. For PV, battery storage and HP, only one GCU is assigned per bus, whereas EVs may occur multiple times per bus to account for households with several EVs. For PV systems, different orientations are distributed across buses proportionally to the orientation of PV systems in the corresponding postal code area, and the average peak power is assigned to all plants for simplicity. Depending on the ratio between installed PV systems and battery storage units in each postal code area, battery storages are randomly added to buses with a PV installation. The resulting distribution of GCUs is illustrated for the Opfingen network in figure 2.

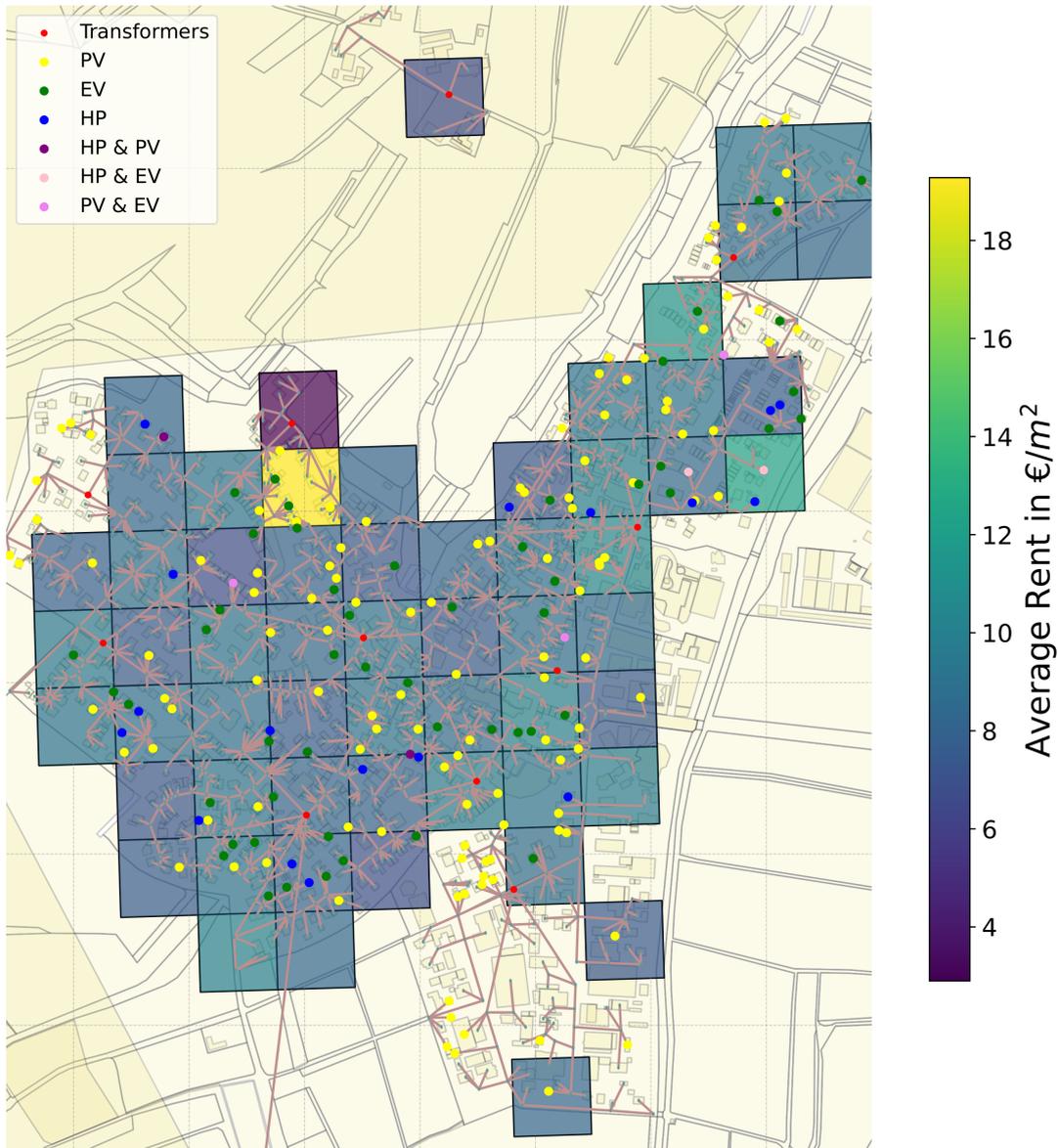


Figure 2. The low-voltage network of Opfingen was created using GridCreator. The average rent per square meter is displayed as an example census feature (color map). The OSM map serves as the background. The associated GCUs are color-coded on the bus points.

Step 4

Load time series are generated with *PyCity*, using 2013 Germany-wide weather data from the *Climate Data Store* [26] to obtain location-specific temperature and solar irradiation for

the bbox. The *census* data provides the number and size of households and population per $100\text{m} \times 100\text{m}$ raster. Apartments are distributed across buses, with multiple apartments per bus, and a time series for each apartment accounting for resident numbers is inserted as a load into the *PyPSA* network at the corresponding bus. For buses with assigned EVs, individual EV time series are generated based on household's occupancy patterns. PV generation time series are created in *PyCity* using weather data, peak power, orientation, and solar irradiance as input factors. The resulting time series are transferred to the *PyPSA* network as generation time series. For HPs, representative technical parameters are used, which can be adjusted by users. Using weather data, *PyCity* generates a consumption time series, which is added to the *PyPSA* network. Figure 3 illustrates the assigned time series for one week for a sample bus from the Opfingen network.

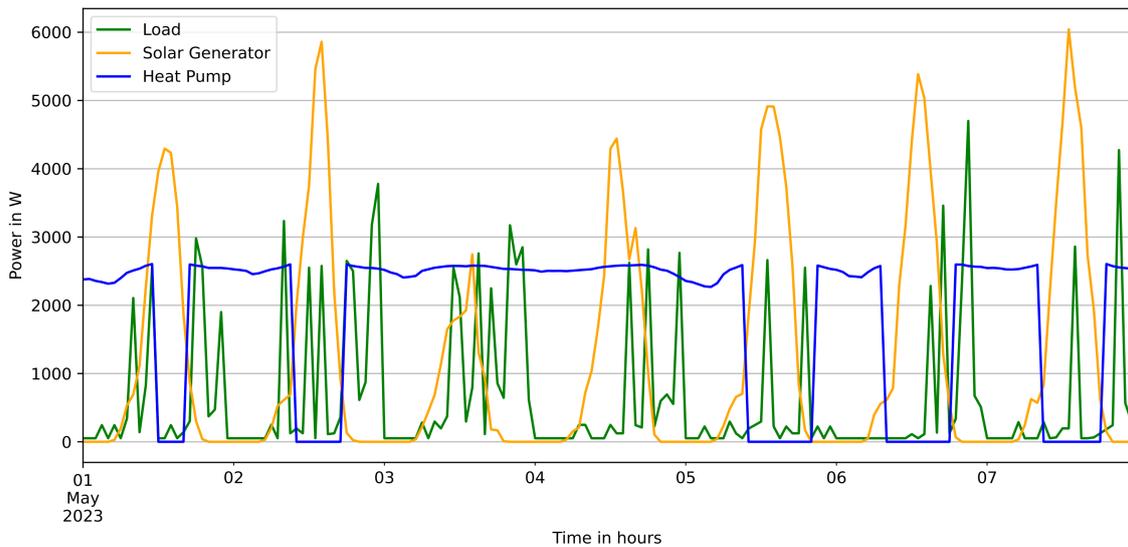


Figure 3. Visualization of the time series for one week, assigned to a bus in the Opfingen network. Shown are the household consumption time series, the generation of the PV system, and the HP consumption time series.

Step 5

The LV network is further processed using *PyPSA*. To ensure functionality, default parameters are defined for all components. Virtual generators are added on the MV side of the transformer to represent electricity flows to and out of the network. Additional components required for the modelling of EVs are added, including external charging points for cases of extended absence. Further details on the implementation can be found in the GridCreator repository.

4. Selection of *census* features

To reduce data volume and avoid over-identification, a subset of representative *census* features was selected, minimizing redundancy from highly correlated variables.

For the selection, *census* features were first aggregated to the spatial resolution of available GCU data. Correlations between population-normalized GCU occurrence and each feature were then computed, retaining only the top 20% most correlated features. Figure 4 illustrates the correlation between average rent per square meter and the number of PV systems normalized to population.

To further reduce features and avoid multicollinearity, the *Variance Inflation Factor (VIF)* was calculated for each dataset [27]. The features with the highest *VIF* were removed iteratively,

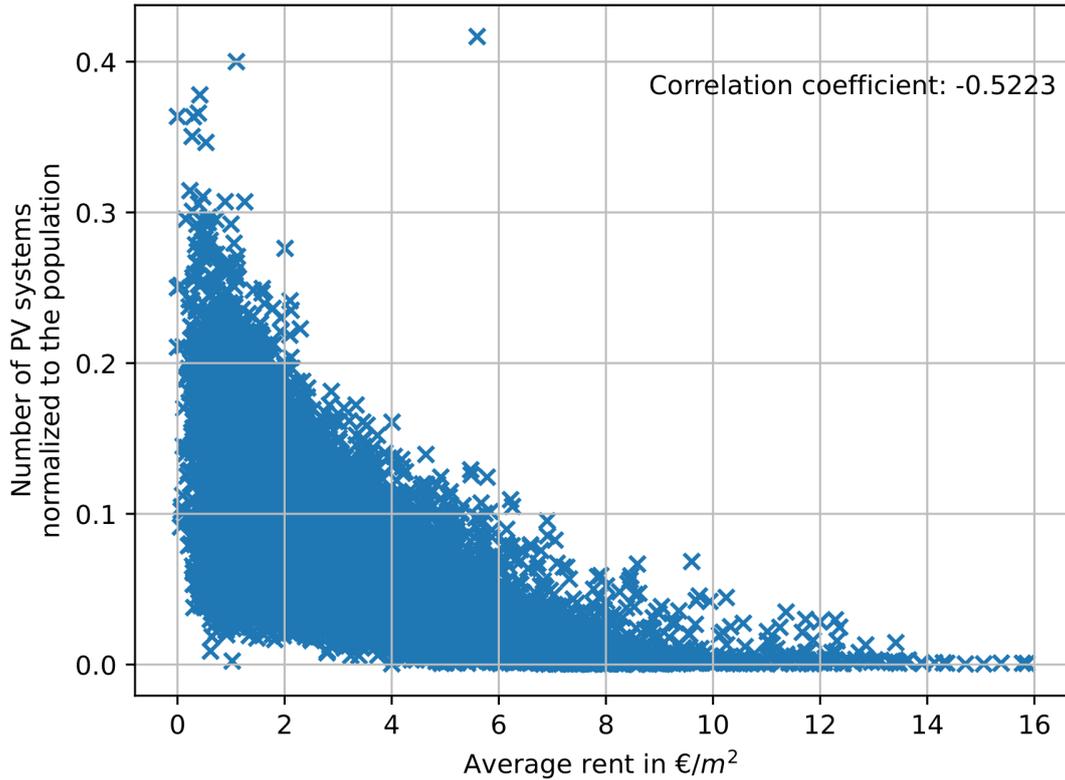


Figure 4. Correlation between the census characteristic describing the average rent per square meter and the occurrence of PV, normalized for population size. Both were aggregated at the postal code level. Two outliers with normalized PV counts of 2.3 and 1.6, each corresponding to an average rent of 0 €/m², were omitted from the figure.

recalculating *VIFs* each time, until all values were below 10, representing a compromise between dataset compactness and feature retention. The correlations of the remaining features with GCU occurrence and their *VIFs* are shown in table A2 in the appendix, providing a specific set of *census* features for assigning each GCU. It can be observed that the features selected to describe EV occurrence exhibit substantially weaker correlations than those for the other GCUs. Consequently, sample analyses sometimes produced EV counts that were significantly too high or too low. To address this, both the total number of EVs and the probability of an EV in each 100m×100m raster cell are derived from the population ratio instead.

5. Calculation of factors

As described in the methodology, GCU allocation is based on a weighted combination of selected *census* features. Weighting factors for each GCU category were calculated using equation (1), with the only difference being the spatial aggregation level: for PV and EVs, the higher level is the federal state and the lower level the regions with known absolute numbers; for HPs, the levels are Germany and the federal states.

Based on these hierarchies, an independent system of linear equations was formulated for each GCU category, with the weighting factors as the only unknowns. The probabilities were derived from the ratio of occurrence frequencies between the respective spatial levels: *res* represents the lower level (areas with available absolute GCU numbers, i.e. PV: postal code, EV: registration district, HP: federal state), and *RES* represents the higher-level aggregation (PV and EV: federal state, HP: Germany). Since the available factors do not describe each probability perfectly, a correction term *x* is added to each equation to capture the residual discrepancy. The equations

can then be written as

$$W_{GCU}^{res} - \frac{\sum_i f_i \cdot z_i^{res}}{\sum_i f_i \cdot z_i^{RES}} = x^{res} . \quad (2)$$

The optimal factors are determined when all x -values are minimize, using the *method of least squares* [28].

6. Application and Outlook

The primary objective of GridCreator is to enable users to easily generate synthetic yet realistic LV networks for individual locations. By manually specifying coordinates, LV networks of arbitrary size can be generated for any region in Germany. These networks are populated with generation and consumption time series, which are individually generated for specific coordinates using geo-socio-economic and technological data. Integrating these networks into the modeling framework *PyPSA* enables direct evaluation and manipulation of the networks. Moreover, GridCreator's modular input and output capabilities allow users to provide their own data, thereby generating more realistic results for a given region using local datasets. Future scenarios, such as simulating full electrification, can also be implemented.

Although GridCreator is ready for immediate use, several improvements are possible. Modeling commercial building load profiles more realistically and not relying on basic predefined standards would improve accuracy especially for dense urban areas with many small shops. Assigning technical parameters such as unit sizes or the coefficient of performance of HPs based on probability distributions would diversify the units present in a grid. A preliminary validation of the presented tool was carried out on a single real LV network. Additional empirical validation on a broader set of networks should be conducted.

List of abbreviations

Table 2. List of abbreviations used.

Abbreviation	Meaning
bbox	bounding box
EV	electrical vehicle
GCU	generation and consumption unit
HP	heat pump
LV	low-voltage
MV	middle-voltage
OSM	OpenStreetMap
PV	photovoltaic
VIF	variance inflation factor

Data availability statement

Input data and code is available on Zenodo (<https://doi.org/10.5281/zenodo.17884917>) and github (<https://github.com/INATECH-CIG/GridCreator>), both under a GPL-3.0 license. The authors are open for contributions and feedback.

Author contributions

Conceptualization: Matthias Behr, Gunter Grimm, Rebecca Hofmann, Ramiz Qussous, Mirko Schäfer, Anke Weidlich; **Methodology:** Matthias Behr, Gunter Grimm, Rebecca Hofmann, Ramiz

Qussous, Mirko Schäfer; **Software:** Matthias Behr, Gunter Grimm; **Formal analysis:** Matthias Behr, Gunter Grimm; **Writing – original draft:** Matthias Behr; **Writing – review & editing:** Matthias Behr, Gunter Grimm, Mirko Schäfer; **Visualization:** Matthias Behr; **Supervision:** Gunter Grimm, Rebecca Hofmann, Mirko Schäfer, Markus Schumacher; **Project administration:** Gunter Grimm; **Funding acquisition:** Anke Weidlich

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT-5.1, a browser-based AI developed by the company OpenAI, to translate, shorten, and improve text. DeepL, a browser-based AI developed by the company DeepL SE was used for translations. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Competing interests

The authors declare that they have no competing interests.

Funding

M.B., M.S., R.Q. and A.W. would like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501865131.

References

- [1] M. Ebad and W. M. Grady, "An approach for assessing high-penetration PV impact on distribution feeders", *Electric Power Systems Research*, vol. 133, pp. 347–354, 2016. DOI: [10.1016/j.epsr.2015.12.026](https://doi.org/10.1016/j.epsr.2015.12.026).
- [2] N. Damianakis, G. R. C. Mouli, and P. Bauer, "Grid impact of photovoltaics, electric vehicles and heat pumps on distribution grids — an overview", *Applied Energy*, vol. 380, p. 125 000, 2025. DOI: [10.1016/j.apenergy.2024.125000](https://doi.org/10.1016/j.apenergy.2024.125000).
- [3] D. Sarajlić and C. Rehtanz, "Analysis of the electrical properties of SimBench low voltage benchmark network models", in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2020, pp. 444–448. DOI: [10.1109/ISGT-Europe47291.2020.9248706](https://doi.org/10.1109/ISGT-Europe47291.2020.9248706).
- [4] S. Korjani, A. Facchini, M. Mureddu, G. Caldarelli, and A. Damiano, "Optimal positioning of storage systems in microgrids based on complex networks centrality measures", *Scientific Reports*, vol. 8, no. 1, p. 16 658, 2018, ISSN: 2045-2322. DOI: [10.1038/s41598-018-35128-6](https://doi.org/10.1038/s41598-018-35128-6).
- [5] T. Brown, J. Hörsch, and D. Schlachtberger, "PyPSA: Python for Power System Analysis", *Journal of Open Research Software*, vol. 6, no. 1, p. 4, 2018. DOI: [10.5334/jors.188](https://doi.org/10.5334/jors.188).
- [6] A. Latif and S. Farrar, "Developing synthetic distribution models using open-source data sets", [Online]. Available: <https://docs.nrel.gov/docs/fy23osti/86846.pdf>.
- [7] J. Amme, G. Pleßmann, J. Bühler, L. Hülk, E. Kötter, and P. Schwaegerl, "The ego grid model: An open-source and open-data based synthetic medium-voltage grid model for distribution power supply systems", *Journal of Physics: Conference Series*, vol. 977, p. 012 007, 2018, ISSN: 1742-6588. DOI: [10.1088/1742-6596/977/1/012007](https://doi.org/10.1088/1742-6596/977/1/012007).

- [8] J. Amme, P. Dubielzig, and B. Schachler, *Distribution grid data generated by ding0*. Zenodo. DOI: [10.5281/zenodo.10405129](https://doi.org/10.5281/zenodo.10405129).
- [9] B. Reveron Baecker, S. Candas, D. Tepe, and A. Mohapatra, "Generation of low-voltage synthetic grid data for energy system modeling with the pylovo tool", *Sustainable Energy, Grids and Networks*, vol. 41, p. 101 617, 2025, ISSN: 2352-4677. DOI: [10.1016/j.segan.2024.101617](https://doi.org/10.1016/j.segan.2024.101617).
- [10] M. Scharf and A. Nebel, "osmTGmod 0.1.1: Documentation, Version 0.1.0", 2016. [Online]. Available: https://github.com/wupperinst/osmTGmod/blob/master/osmTGmod_documentation_0.1.0.pdf.
- [11] S. Meinecke et al., "SimBench - Dokumentation: Dokumentationsversion DE-1.0.1: Elektrische Benchmarknetzmodelle", 2021. [Online]. Available: https://simbench.de/wp-content/uploads/2025/11/simbench_documentation_de_1.1.0.pdf.
- [12] J. Fonseca et al., *Architecture-building-systems/cityenergyanalyst: Cityenergyanalyst v4.0.0-beta.5*, 2025. DOI: [10.5281/ZENODO.598221](https://doi.org/10.5281/ZENODO.598221).
- [13] J. Schiefelbein, J. Rudnick, A. Scholl, P. Remmen, M. Fuchs, and D. Müller, "Automated urban energy system modeling and thermal building simulation based on openstreetmap data sets", *Building and Environment*, vol. 149, pp. 630–639, 2019, ISSN: 0360-1323. DOI: <https://doi.org/10.1016/j.buildenv.2018.12.025>.
- [14] F. Lombardi et al., "RAMP: stochastic simulation of user-driven energy demand time series", *Journal of Open Source Software*, vol. 9, no. 98, p. 6418, 2024. DOI: [10.21105/joss.06418](https://doi.org/10.21105/joss.06418).
- [15] N. Pflugradt, P. Stenzel, L. Kotzur, and D. Stolten, "LoadProfileGenerator: An agent-based behavior simulation for generating residential load profiles", *Journal of Open Source Software*, vol. 7, no. 71, p. 3574, 2022. DOI: [10.21105/joss.03574](https://doi.org/10.21105/joss.03574).
- [16] F. Gotzens et al., *DemandRegio - Harmonisierung und Entwicklung von Verfahren zur regionalen und zeitlichen Auflösung von Energienachfragen*, 2020. DOI: [10.34805/FFE-119-20](https://doi.org/10.34805/FFE-119-20).
- [17] S. Chai and G. Chadney, *Faraday: Synthetic smart meter generator for the smart grid*, 2024. [Online]. Available: <https://arxiv.org/pdf/2404.04314>.
- [18] G. Boeing, "Modeling and analyzing urban networks and amenities with OSMnx", *Geographical Analysis*, vol. 57, no. 4, pp. 567–577, 2025, ISSN: 0016-7363. DOI: [10.1111/gean.70009](https://doi.org/10.1111/gean.70009).
- [19] *Pystatis*, 2025. [Online]. Available: <https://correlaid.github.io/pystatis/>.
- [20] D. Fioriti, *PyPSA-Distribution documentation*, 2025. [Online]. Available: <https://pypsa-distribution.readthedocs.io/en/latest/index.html>.
- [21] open_ego-Team, *eDisGo 0.2.1 documentation*, 2025. [Online]. Available: <https://edisgo.readthedocs.io/en/dev/>.
- [22] T. Enderle and B. Kleber, "Geheimhaltung mit der Cell-Key-Methode im Zensus 2022", 2024. [Online]. Available: https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2024/06/geheimhaltung-cell-key-methode-zensus-062024.pdf?__blob=publicationFile&v=6.
- [23] *Marktstammdatenregister*, Bundesnetzagentur, 2025. [Online]. Available: <https://www.marktstammdatenregister.de/MaStR>.
- [24] *KraftfahrBundesamt - Statistikportal*, 2025. [Online]. Available: <https://das-kba-statistikportal.hub.arcgis.com/>.
- [25] Bundesverband Wärmepumpe (BWP) e.V. und BWP Marketing & Service GmbH, *Zahlen & Daten — BWP*, 2018. [Online]. Available: https://www.waermepumpe.de/fileadmin/user_upload/Mediengalerie/Zahlen_und_Daten/Absatzzahlen_Marktanteile/619_WPinDeutschland_2018.PNG.
- [26] *Climate data store: Era5 hourly data on single levels from 1940 to present*, 2025. [Online]. Available: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download>.

- [27] *statsmodels.stats.outliers_influence.variance_inflation_factor* - *statsmodels 0.15.0*, 2025. [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.stats.outliers_influence.variance_inflation_factor.html.
- [28] *least_squares* — *SciPy v1.16.2 Manual*, 2008. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html.

Appendix

A. Schematic representation of the GridCreator

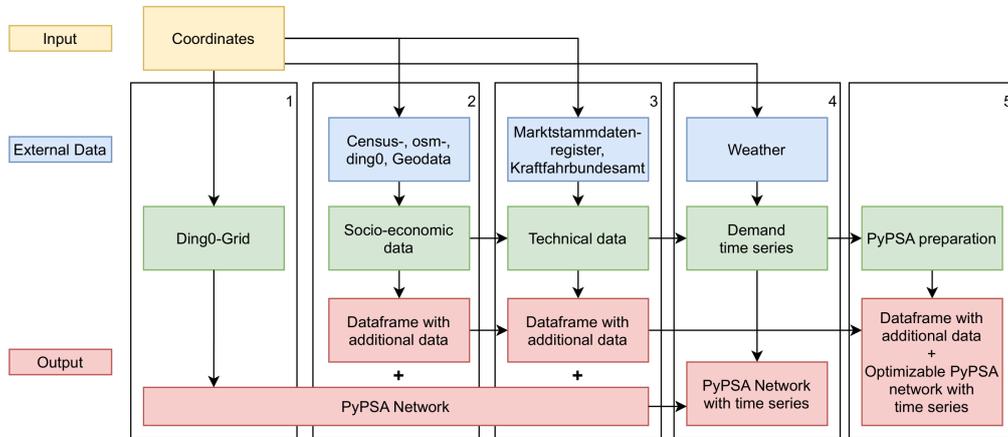


Figure A1. Schematic representation of the GridCreator workflow with its five steps.

B. Variance inflation factor values of the selected census features

Table A2. Selection of census features for each GCU with the associated correlation coefficients for the occurrence of the GCU and the respective VIF value in relation to the other data sets.

Feature		Correlation	VIF
Photovoltaic			
Description	Census Name		
Building heating with wood or wood pellets	Geb_Holz_Holzpellets	0.596674	3.775426
Apartments with a living area of 30 - 39 m ²	30bis39	-0.484745	5.061297
Average rent per square meter	durchschnMieteQM	-0.522255	5.212475
Unmarried couple with at least one child (<18 yr)	NichtehelLG_mind_1Kind_unter18	-0.495605	5.436315
Heating with gas	Gas	-0.565621	6.613232
Apartments in a detached two-family house	Freist_ZFH	0.526851	6.830665
Detached single-family house	Geb_FreiEFH	0.487688	7.076520
Number of apartments per building	7bis12_Wohnungen	-0.560625	7.815041
Apartments with a living area of 40 - 49 m ²	40bis49	-0.543683	7.903656
Electrical Vehicle			
Description	Census Name		
Persons with EU citizenship	EU27_Land	0.372078	3.868995
Residential buildings built between 1970 and 1979	a1970bis1979	0.373149	4.795826
Unmarried couple with at least one child (<18 yr)	NichtehelLG_mind_1Kind_unter18	-0.421189	5.413091
Residential building built before 1919	Vor1919	-0.463115	6.169144
Heating with coal	Kohle	-0.443703	7.615800
Building heating with coal	Geb_Kohle	-0.443774	8.210747
Heat Pump			
Description	Census Name		
Building heating with district heating	Geb_Fernwaerme	-0.580003	3.135556
Heating with wood or wood pellets	Holz_Holzpellets	0.704341	3.920469
Married couple with youngest child aged 18 or over	Ehep_Kinder_ab18	0.579440	7.131330