

RESEARCH

Open Access



A machine learning approach to model the future distribution of e-mobility and its impact on the power grid

Paul Eitel* and Peter Stolle

From The 11th DACH+ Conference on Energy Informatics 2022
Freiburg, Germany. 15-16 September 2022

*Correspondence:
Paul.Eitel@fit.fichtner.de

Fichtner IT Consulting GmbH,
Sarweystraße 3, 70191 Stuttgart,
Germany

Abstract

It is to be expected that there will be a shift toward electromobility with regard to private passenger cars in the coming years. This will oblige the respective power grid providers to upgrade their networks in future years. So that grid operators can plan and operate their grids to meet future needs, they have to have as complete information as possible about the loads they will be required to handle. Depending on voltage level, geographic location, general grid load, and spread of e-mobility, the situation will vary. The assumption explored in this paper is that external factors influence the distribution of EV chargers. As a second task, the impact on the power grid is simulated by means of various scenarios on the basis of this identified distribution, with the focus on low voltage (LV) grids. Sociodemographic data is used as a geographic grid to determine potential distribution. For this, machine learning methods from the field of “Species Distribution Modeling” are applied for a prospective distribution concept. Using this distribution model, the results of simulation of power grid utilization reveal vulnerabilities scattered around the networks. It is shown that e-mobility will, in the future, present a challenge for power grid operators, for which solution concepts are needed.

Keywords: E-mobility, Distribution model, Species distribution modeling, Wallboxes, Power grid capacity utilization, LV grids, Machine learning, Data analysis

Introduction

As a survey by the German Association of Energy and Water Industries (Bdew: Private Ladeinfrastruktur foerdern 2022) reveals, 65% of respondents say they would prefer to be able to charge an EV from their own homes. Such surveys indicate that, in future, EV charging will primarily be done at the LV level of the distribution grid. Since simultaneous charging in regions with above-average numbers of EVs can lead to localized load peaks, grid operators are forced to act (Bdew: Intelligente Netze für Elektromobilität 2022). In step with the rise in the number of private EVs, installations of private chargers will likewise increase (Bdew: Private Ladeinfrastruktur foerdern 2022). Based

on this assumption, the remit of this paper is thus also limited to the impact on LV grids. Essentially, this paper examines the assumption that external factors, such as sociodemographics, influence the distribution of EV chargers and thus their impact on the power grid. Therefore, as the first objective of this paper, a potential geographic distribution of wallboxes can be derived. Endeavors are directed to identifying those factors that may be considered for modeling the potential future distribution of wallboxes. The rationale for such a model is the assumption that wallboxes are not distributed uniformly among all households, but that in the near future there could be regions with increased penetration and regions with low penetration (Arnhold et al. 2018). These regions have to be identified. The second objective of this paper is to run a simulation based on the identified distribution model, which will be used to investigate potential grid capacity utilization under various scenarios. For this purpose, a computational grid with concrete consumption data is available, on the basis of which potential load peaks are to be identified. The aim here is to build on other studies to achieve results that cover as wide an area as possible.

In summary, the two issues to be addressed in this paper can be formulated as follows:

- 1 **Which data can be taken to derive a spatial wallbox distribution and how can a potential distribution be modeled?**
- 2 **What is the impact of such distribution on an existing power grid?**

State of the art

Regarding charging behavior, different approaches identify peaks arising between 5 p.m. and 8 p.m. For the expected loads, especially in German publications the reference values of 3.7kW, 11kW and 22kW are used as charging power in the private sector, and this also decisively impacts charging duration. Echternacht et al. (2018), Gruosso (2016), Falco et al. (2019), Quirós-Tortós et al. (2015)

The aim of investigations into this topic is to identify potential overloading of transformers or power lines as well as violations of voltage tolerance limits (Weis et al. 2021). The aim is to show at what penetration of EVs it may be expected that such constraints will come into play (Echternacht et al. 2018; Weis et al. 2021; Held et al. 2019). In part, different grid configurations are also taken into account, depending on their geographical location, i.e. rural, suburban, or urban (Held et al. 2019). But for the most part, all studies have the following in common: they take specific actual grids or reference grids as a starting point for a capacity utilization analysis. In line with the assumption that not all grid participants draw power from the grid at full load at the same time, a coincidence factor is also sought for realistically simulating grid utilization due to e-mobility. Thus in Echternacht et al. (2018) and Held et al. (2019) it is assumed that not all EVs will be charged at full charging power at the same time. Calculation of such a factor depends on other aspects. On the one hand, with lower charging power, the charging time is longer and therefore more cars are likely to be charging at the same time (Echternacht et al. 2018). On the other, together with the battery's capacity, likewise influencing the coincidence factor is the state-of-charge (SOC) Held et al. (2019). In addition, the coincidence factor depends on the number of vehicles that are under consideration. The greater this

is, the lower is the coincidence factor (Echternacht et al. 2018; Held et al. 2019). The coincidence factor can thus be derived using the formula from (Roestel 2017).

In Echternacht et al. (2018) and Weis et al. (2021), the results indicate that e-mobility leads to an increase in power grid load but, in fact, even with high wallbox penetration and charging power up to 22 kW, the thermal limits of power transmission lines and transformers in the grids used as examples are not exceeded. This is mainly owing to the assumption of a low value for the coincidence factor. With a higher value, penetrations of as low as 20% to 30% could result in overloading (Echternacht et al. 2018; Weis et al. 2021; Held et al. 2019). In that study, however, a voltage drop that violates these limits is already to be expected at a penetration of between 50% and 80% (Held et al. 2019; Gruosso 2016). The fact that the results differ in part shows that the effect on the LV grid is mainly a function of the coincidence factor. Furthermore, the differing outcomes also depend on which power grids are taken as examples. But so that more general conclusions regarding critical penetration can be drawn, this methodology will have to be applied to a larger number of different grids (Echternacht et al. 2018).

Material and methods

Excursus: species distribution modeling

In developing a distribution model, Species Distribution Modeling (SDM) methodologies and practices have proven to be practicable for this paper. Specifically, the goal of SDM is to apply algorithms to infer the distribution of a particular species based on a set of geolocated occurrences of that species. The main problems are the limited number of observations, the bias of sampling and that in most cases only data on species presence can be drawn (Botella et al. 2018; Ward et al. 2009). Providing a workaround for this are so-called “pseudo-absence” data. In their simplest form, these are a randomly selected section of the background’s pixels and variables in its surroundings (Ward et al. 2009). SDM models can consequently be divided into the two subcategories of presence-only (PO) and presence-absence models (Ward et al. 2009). Thus, it is a binary classification problem and machine learning methods can be used in this case (Gastón and García-Viñas 2011).

Extent of the trial region

As shown in Figure 1, the trial region extends over large parts of Saarland.

As can be seen in Figure 1, the trial region belongs to the more rural part of Saarland. Overall, grid connection data from 36 (marked orange in Figure 1) of a total of 52 municipalities in Saarland are available for this trial.

Grid data

Grid connections: Around 172,000 grid connections to the respective premises are available as geographic points for the trial region. These connections can be assigned to a local power grid substation (LPGS) and thus to its associated LV grid.

EV chargers: Certainly most relevant for this paper are the data on EV chargers. Data such as power rating, year of manufacture and associated LPGS are available as geographical points. The point-by-point data can be classified as EV chargers or

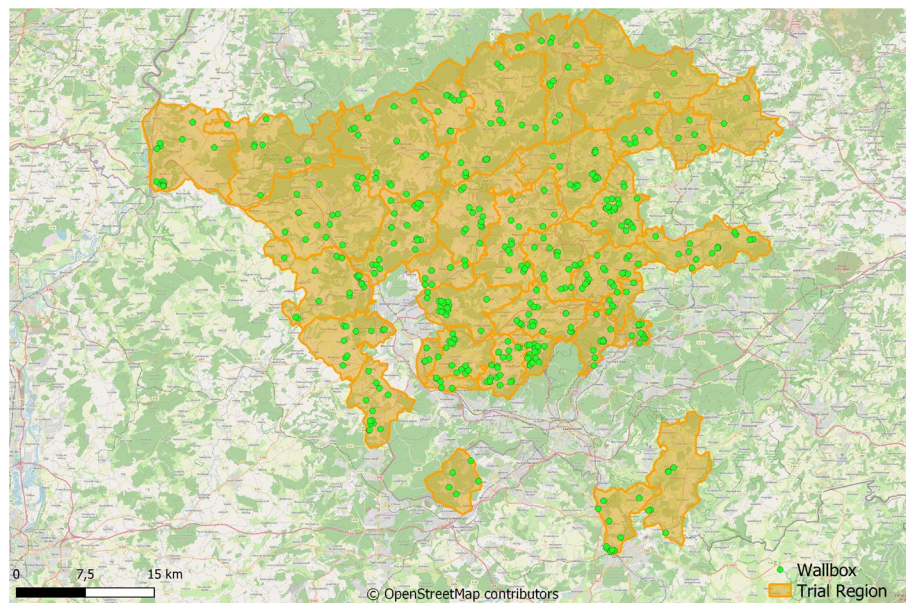


Fig. 1 Trial region and wallboxes

wallboxes, and a distinction can be made between “public” and “private”. Private wallboxes thus constitute the 437 private charging options registered with the grid operator, the distribution of which is shown in Fig. 1.

Substations: For grid calculations, the connections are spread over around 2500 local power grid substations. These may be regarded as geographic points with their basic rated power.

Consumption data: As consumption data of the individual connections, the time series of their load profiles are available for this paper. These load profiles for each grid connection result from the capacity utilization factor that has actually been metered throughout the year. For the annual value, an average power value is stored using standard load profiles for a time series with 15 minute intervals.

Environmental variables

For this paper, in addition to the power grid data, sociodemographic grid data for all of Saarland are available through the data package “DDS Data Grid”. With a 100×100 m geographic grid, these data provide a basis for homogenizing a wide variety of demographic data. The sociodemographic data are supplemented with additional geographic location surroundings variables generated for this paper during the process of feature engineering.

- **Basis:** Absolute figures for buildings, households and persons
- **Population:** Relative shares of gender and age brackets
- **Building:** Relative shares of building categories (by number of households per building) and relative shares of residential, mixed-use and commercial buildings
- **Purchasing power:** Relative shares of 6 purchasing power categories and single / multi-person households

- **Feature Engineering:** Intersection and categorization of existing features. Generation of information on PV installations, shares of "GREENS" party voters, EVs in the surrounding cells (100 m radius), Buildings and their geographic area, garages per building, "Points of Interest" (POI) per building and distance from city center.

Derivation of SDM

For the purposes of this paper, an architecture can be derived from the SDM methodologies. Consequently, in the following, the data are assigned to the respective terms. For this application, the sociodemographic grid cells presented in the previous section along with the feature engineering data serve as background data. The specific geographic position of the wallboxes results in the observation data, which serve as positive input variables for model training. In the case of wallboxes, it is not possible to draw any conclusions regarding the actual absence of e-mobility, as potential customers for wallboxes could be located in all cells. Thus, with regard to a distribution model, the generation of pseudo-absence data would be appropriate. Here, as shown by best SDM practice, the largest possible segment is chosen and about 5000 random cells are taken as pseudo-absence data for training and test data. This corresponds to about 20% of the total background data.

Results

Distribution model

Model training

In this subsidiary step, four machine learning algorithms are trained on the basis of the training data. These are: OCSVM (One Class Support Vector Machine), logistic regression, random forest, and neural network. For each algorithm, a set of well-established hyperparameters is defined and cross-validation is performed for their combinations. For the three binary algorithms, the imbalanced dataset is balanced with the unequally distributed pseudo-absence data, using class weighting. A strong regularization was chosen for all algorithms to counteract overfitting. Due to monotonicity in the data, the sigmoid activation function was chosen for the OCSVM and the neural network. With just one hidden layer, the neural network is not structured to be overly complex in this study. The number of hidden neurons is chosen to be about twice as large as the number of input neurons, which allows the model to learn to a greater depth of detail.

Model validation

To limit the spatial autocorrelation (see Griffith [1992](#)) of the data, spatial cross-validation is used for model validation. The subsets for this validation are thereby generated from the municipalities (see Fig. [1](#)). Thus, the algorithms are always trained for a subset of communities, their parameters optimized, and validated on a subset of "unseen" communities. Roberts et al. ([2017](#))

Results of trial runs and model selection

ROC curves

For the ROC curves (Fig. [2](#)), the test data are analyzed according to the cells classified as correct positive and false positive. Curves are obtained that provide information

about all combinations of the output score in relation to the two positive rates. An initial indication of the quality of the models is provided by these curves. Regarding the false positive rate, it must be noted at this point that the negative examples concern the pseudo-absence data. These are a randomly chosen large selection of background data and therefore also “positively contaminated” to a certain degree. The stronger the “contamination” of the data, the flatter the curves.

Model selection

Based on the test output as per AUC, the distribution of the output score and the interpretability of the models, logistic regression is shown to be the most suitable algorithm for this use case despite a slightly worse AUC compared to the neural network.

Evaluation and depiction of distribution model

Evaluation of coefficients

One of the questions addressed in this paper is what external factors influence the distribution of wallboxes. This will be answered in this section by performing a coefficient analysis. Firstly, Fig. 3 shows on the left side in which direction and to what degree the coefficient influences the model. The right-hand window shows the variability of this value with cross-validation. A high degree of variability implies correlations or multicollinearity in the data. In summary, it can be said that, among other things, numerous PV installations, high purchasing power and many “GREEN” voters per cell will favor the prevalence of e-mobility, whereas extremely large and small population densities, large building plots, low purchasing power and a limited age group will militate against its adoption.

Geographical depiction

In this section, the results of the logistic regression will be presented geographically for the reader. In Fig. 4, this distribution can now be represented in a high-resolution geographic map. The result thus appears as a 100x100 m geographic grid showing probabilities for the occurrence of wallboxes and depending on factors in the surroundings of the

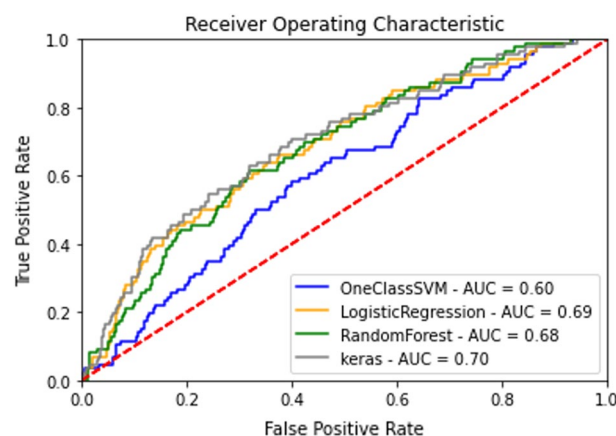


Fig. 2 ROC curves—model outputs from the test data

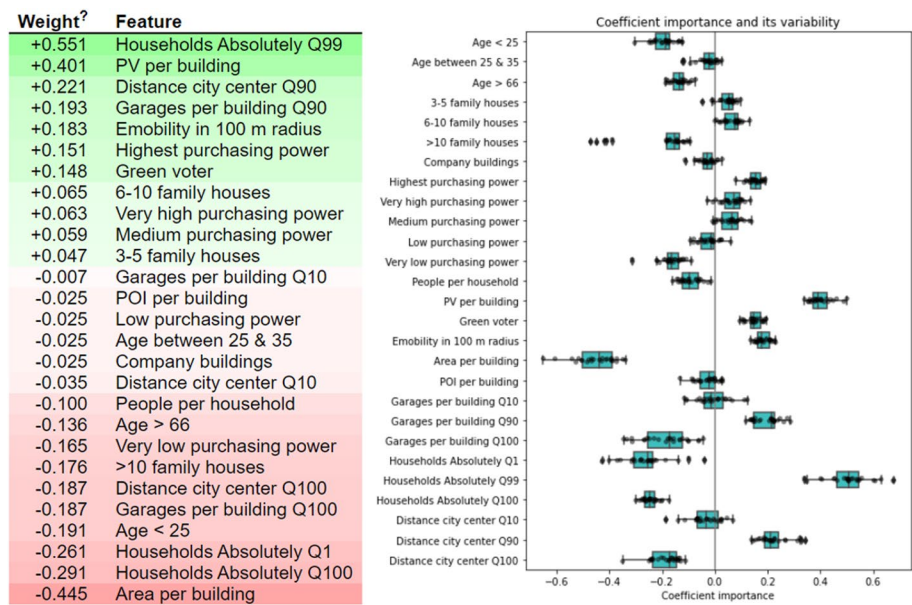


Fig. 3 Coefficient analysis of the logistic regression

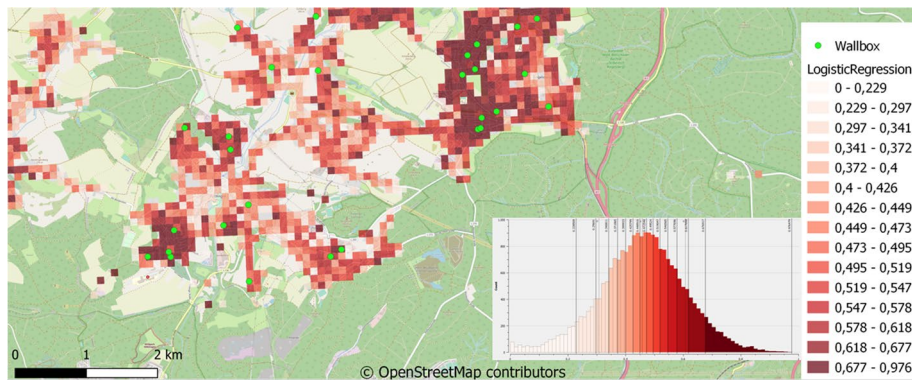


Fig. 4 Segment from the geographical depiction of the E-mob GRID

cells. As can be seen in Fig. 4, the wallboxes that were known at the time of the study are mostly located in cells where the model outputs a high probability.

Simulation of wallbox distribution

For blanket simulation of the impact of wallboxes on the power grid, the first step is to simulate an appropriate predicted distribution. For this, the outputs of the distribution model from the foregoing section are taken and, based on these, weighted random sampling is prepared. With the grid connection data and the geographic grid with its probability values, a probability can be assigned to each connection as a weighting factor. With these elements, a distribution simulation run can now be executed. To do this, each grid connection is extracted in turn from all connections depending on its assigned probability in successive simulation rounds. Each round defines a relative share of market penetration. A further distribution simulation parameter is the influence of the model.

In order to raise the influence of the model, the simulation rounds are performed more frequently for each penetration level, while retaining the most frequently selected connections in the simulation. In this way, the influence of the model can be increased and thus the degree of randomness can be lessened. Selecting simulation rounds of 1, 20 & 100 proves to be the most appropriate for multiple simulations with different numbers of rounds. In the following simulations the influence of the model is expressed as textual parameter:

- **Low:** 1 simulation round
- **Moderate:** 20 simulation rounds
- **High:** 100 simulation rounds

Simulation of the impact on the power grid

With the distribution simulation data, information about the impact on an actual power grid can be obtained in this section. The focus hereby is on the calculated distribution of wallboxes. For this purpose, assumptions regarding the charging behavior and the load profiles of the connections are simplified and the connections of each LV grid are analyzed cumulatively in this paper.

Simulation structure (impact on grid)

For the simulation run, the data are thus considered at the level of the grid connection. The simulated connections are available as wallboxes. The load profiles and thus the starting point for grid capacity utilization without simulated wallboxes are provided by the grid connection data together with their consumption data. For simulating the capacity utilization, these two load data are considered aggregated at the local power grid substation (LPGS) level and reconciled with the respective nominal power ratings. In this paper, for the sake of simplicity we choose points in time that are to be found in the literature. The grid capacity utilization is considered for a period during 2018 from 6:30 p.m. to 6:45 p.m., for one randomly selected working day in summer and one in winter. To test whether the currently installed infrastructures can cope with the simulated charging profiles, the following parameters are considered:

Penetration: 1% to 30% market penetration is investigated for the trial region

Influence of the model: The degree of randomness of the simulation

Coincidence factor: Formula from (Roestel 2017) with multiple values for each scenario

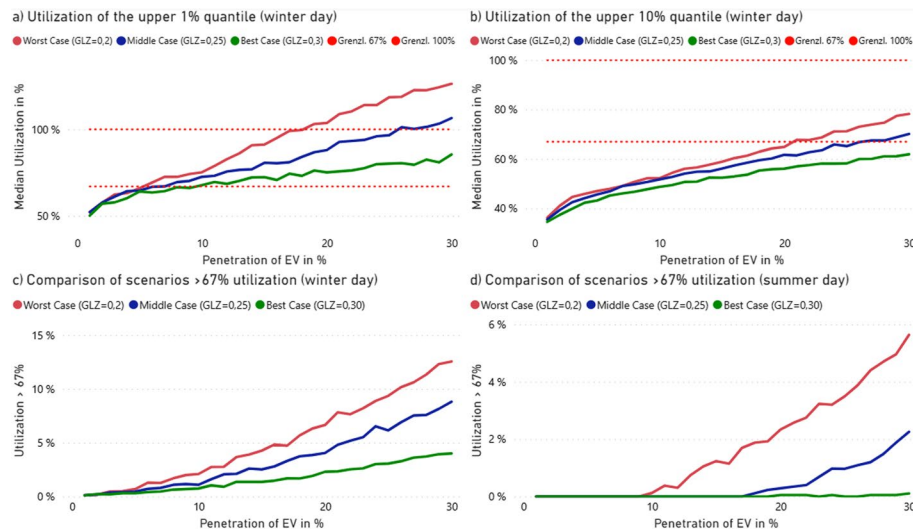
Average charging power: charging powers between 7 kW and 15 kW

Loading limits: Rated power (100%) of the transformers and 2/3 (67%) of this capacity

From the parameters presented, different scenarios are derived for the simulation runs in this study. The goal in developing the scenarios is to achieve a result that is as realistic and informative as possible. For these simulation runs, the three scenarios from Table 1

Table 1 Scenarios

Scenario	Best case	Mid case	Worst case
Influence of the model	Low	Moderate	High
Average charging power	7 kW	11 kW	15 kW
Coincidence factor	$g_{\infty} \in [0, 25; 0, 4]$	$g_{\infty} \in [0, 2; 0, 35]$	$g_{\infty} \in [0, 15; 0, 3]$

**Fig. 5** Comparison of scenarios

are examined with regard to penetration, time of year, and possible coincidence factors (g_{∞}).

Results of simulation runs

The above derived scenarios are examined here for their impact on the power grid. Each scenario is evaluated for different parameters. The focus is on wallbox market penetration, which serves for temporal ordering of the simulation runs. Also analyzed are the impact of the time of year and the coincidence factor. For a meaningful evaluation, among other things, the LV grids are considered at their peak capacity utilization. For this purpose, LV grids are split into their top 1 and 10% quantiles based on their relative capacity utilization and evaluated using the median.

Comparison of scenarios

From Fig. 5a, b, it is evident that the worst-case scenario stands out from the other scenarios, especially at the peak. Thereby, the differences between the scenarios are significantly larger for the 1% quantile than for the 10% quantiles. For the top 1% of networks, impacts are already evident at low wallbox penetration. However, the gap between the scenarios only becomes apparent at moderate to higher penetrations. For the upper 10% quantiles, the curves diverge much later.

The impact of time of year on each outcome is evident in Fig. 5c, d. Here, compared to the other scenarios, the worst-case scenario shows a significantly larger gap

between the scenarios on a summer day than on a winter day. Thereby, PV installations that are still feeding into the grid at this time of day during summer can no longer cope with the EV charging load. Overall, the critical impacts in all scenarios are concentrated in a small percentage of all LV grids in the trial region. In the worst-case scenario, with a maximum penetration of 30% and a realistic coincidence factor of 20%, up to 13% of the grids are impacted on a winter day (see Fig. 5c). In a best-case scenario with a coincidence factor of 30%, only up to 4% of all networks are impacted under the same conditions. Thus, in the near future, with a penetration of between 5% and 10%, isolated limit violations are only to be noted under the assumption of extreme conditions. Considering a somewhat more distant future with a penetration of 10% to 20%, the grid load will be significantly higher. Here, already 1% in the best case and in the worst case up to 7% of the LV grids show limit violations. With a much higher proliferation of wallbox installations and a market penetration of up to 30%, almost 13% of the grids show limit violations under the worst conditions (see Fig. 5c).

Geographic evaluation

The simulation results show that isolated limit violations can occur in the power grids. By undertaking a geographic evaluation, the affected networks can be analyzed in the geographic region. The goal of a geographic evaluation is to show specific clusters in the trial region where the power grid exhibits vulnerabilities and, in addition, where a high prevalence of wallboxes is predicted. Thus, as shown in Fig. 6, the substations can be visualized on a map with regard to their geographic location and their relative capacity utilization. A perfunctory examination shows spatially distributed point-by-point limit load violations in a worst-case scenario with the parameters shown, but also potential cluster zones. In these locations, there is thus a combination of a high density of simulated wallboxes and an infrastructure that is not designed to cope with this situation.

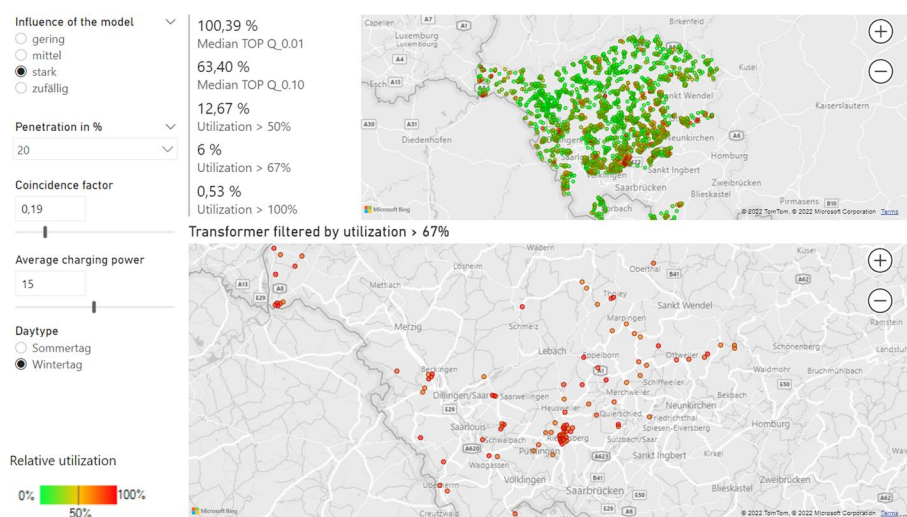


Fig. 6 Geographic evaluation of simulation results

Conclusion

The more rural nature of the trial region of this study cannot serve as a reference for the whole of Germany, since, among other things, it is not transferable to an urban environment. In this respect, further studies could show how EV charging behavior in a densely populated urban setting would impact the power grid. Moreover, in the simulation of this study, only the impacts on transformers are gauged. However, as previous research has shown, violations of the voltage tolerance band may arise even before the equipment is overloaded. Under the remit of this paper, neither these impacts nor violations of the thermal cable limits can be investigated. Further studies of these factors could reveal greater impacts on today's power grid even at a lower wallbox market penetration. Another optimization would be the assumption that regions also differ geographically regarding the battery state-of-charge (SOC). The SOC greatly influences the coincidence factor which, in turn, is a key driver of the grid load, as is evident in this paper. In conclusion, this paper has shown how potential vulnerabilities in the power grid can be identified, particularly if the geographic spread of e-mobility is not homogeneous. It is precisely for such cases that machine learning methods and models have proven to be useful tools for simulating the spread of e-mobility take-up and how to build on previous research, thus enabling predictions of power grid capacity utilization over an extensive region. These methods have shown that, in contrast to previous work, individual networks can become overloaded. These networks must be identified at an early stage in order to guarantee network stability.

Acknowledgements

Thank you to everyone who made this work possible. The FIT colleagues who made the contacts. VSE and PTV, which provided an important part of the data.

About this supplement

This article has been published as part of Energy Informatics Volume 5 Supplement 1, 2022: Proceedings of the 11th DACH+ Conference on Energy Informatics. The full contents of the supplement are available online at <https://energyinformatics.springeropen.com/articles/supplements/volume-5-supplement-1>.

Author contributions

PS conceived of the presented idea. PE developed the theory and implemented the idea. He was supported by PS in an advisory capacity. All authors read and approved the final manuscript.

Funding

This Research was funded by Fichtner IT Consulting GmbH.

Availability of data and materials

Sociodemographic data: © 2021 AZ Direct GmbH, Gütersloh; DDS Digital Data Services GmbH, Karlsruhe; OSM data: <https://wiki.openstreetmap.org/wiki/>, citycenter: <https://wiki.openstreetmap.org/wiki/DE:Key:place>, OSM POI: <https://wiki.openstreetmap.org/wiki/Key:amenity>, building and garage: <https://wiki.openstreetmap.org/wiki/Key:building>; Voting data: <https://www.wahlergebnis.saarland.de>; Power grid data: company property.

Declarations

Competing interests

The authors declare that they have no competing interests.

Published: 7 September 2022

References

- Arnhold O, Goldammer K, Pieniak N, Hübner K, Hartmann J (2018) Charging infrastructure planned right – a new opportunity for everyone. In: Liebl J (eds) *Netzintegration der Elektromobilität 2018*. Proceedings. Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-23393-8_10

- Bdew: Intelligente Netze für Elektromobilität. (2022) <https://doi.org/https://www.bdew.de/energie/elektromobilitaet-dossier/intelligente-netze-fuer-elektromobilitaet/>. Accessed 04 Jul 2022.
- Bdew: Private Ladeinfrastruktur fördern. (2022) <https://doi.org/https://www.bdew.de/energie/elektromobilitaet-dossier/private-ladeinfrastruktur-foerdern/>. Accessed 04 Jul 2022.
- Botella C, Joly A, Bonnet P, Monestiez P, Munoz F (2018) A deep learning approach to species distribution modelling. In: Joly A, Vrochidis S, Karatzas K, Karppinen A, Bonnet P (eds) Multimedia tools and applications for environmental & biodiversity informatics. Multimedia systems and applications. Springer, Cham, pp 169–199. https://doi.org/10.1007/978-3-319-76445-0_10
- Echternacht D, Haoati IE, Schermuly R, Meyer F (2018) Simulating the impact of e-mobility charging infrastructure on urban low-voltage networks. In: NEIS 2018; Conference on Sustainable Energy Supply and Energy Storage Systems, pp. 1–6
- Falco M, Arrigo F, Mazza A, Bompard E, Chicco G (2019) Agent-based modelling to evaluate the impact of plug-in electric vehicles on distribution systems. In: 2019 International Conference on Smart Energy Systems and Technologies (SEST), pp. 1–6
- Gastón A, García-Viñas J (2011) Modelling species distributions with penalised logistic regressions: a comparison with maximum entropy models. *Ecological Modelling* 222:2037–2041
- Griffith DA (1992) What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *Espace Géogr* 3(21):265–280
- Grusso G (2016) Analysis of impact of electrical vehicle charging on low voltage power grid. In: 2016 International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles International Transportation Electrification Conference (ESARS-ITEC)
- Held L, März A, Krohn D, Wirth J, Zimmerlin M, Suriyah MR, Leibfried T, Jochem P, Fichtner W (2019) The influence of electric vehicle charging on low voltage grids with characteristics typical for germany. *World Electric Vehicle Journal* 10(4)
- Quirós-Tortós J, Ochoa LF, Lees B (2015) A statistical analysis of ev charging behavior in the uk. In: 2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM), pp. 445–449
- Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Aroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8):913–929
- Roestel T (2017) Elektromobilität - bewertung der auswirkungen auf städtische verteilungsnetze. In: IEEE German Chapter Workshop, pp. 1–6
- Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the em algorithm. *Biometrics* 65(2):554–563
- Weis A, Biedenbach F, Mueller M (2021) Simulation and analysis of future electric mobility load effects in urban distribution grids. In: ETG Congress 2021, pp. 1–6

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
