Computers & Security 146 (2024) 104042

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Fostering security research in the energy sector: A validation of open source intelligence for power grid model data

Anja Klauzer [a], Markus Maier [b], Lore Abart-Heriszt [c], Johanna Ullrich [b,*]

[a] Networks and Critical Infrastructures Security Research Group, SBA Research, 1040, Vienna, Austria
[b] Security & Privacy Research Group, Faculty of Computer Science, University of Vienna, 1090, Vienna, Austria
[c] Institute of Spatial Planning, Environmental Planning and Land Rearrangement, University of Natural Resources and Life Sciences, 1190, Vienna, Austria

ARTICLE INFO

Keywords:
Open source intelligence
Power grid security
Critical infrastructure
Data validation

ABSTRACT

Cyber attacks against power grids, interrupting utility service and causing blackouts are on the rise, and increasingly motivate researchers to investigate this topic. Thereby, models of real-world power grids are an indispensable prerequisite, but operators do not make them available, allegedly for reasons of protection. This security-by-obscurity strategy appears futile as grid artifacts (lines, plants, substations) are large and cannot be easily hidden. It seems promising to infer real-world model data from publicly available data, and indeed, multiple models were generated through Open Source Intelligence (OSINT). Questions on the models' quality remain, however, open but are of utter importance for research building on these models, especially as the results might have considerable impact on society and national security. This paper approaches this particular point and investigates whether OSINT leads to data on real-world power grids of sufficient quality; by the example of the European country of Austria, we investigate whether all parameters that are relevant for power flow analysis, a standard approach in power engineering, can be inferred from publicly available data (OpenStreetMap, national statistics, etc.), and validate this data against ground truths, including governmental land use plans, Google Street View and the power sector's information material. Our validation shows that the inferred data meets reality well — among others, the extra-high voltage level is 100% (lines) rsp. 98% (substations) complete. Beyond, the inferred data is up-to-date as the construction of lines or substations is always documented in OSM, in 76% of the cases even before finalization of the construction works. An analysis of 24 other European countries revealed that electric systems, substations, and power plants are documented in OSM to a similar extent as in Austria, motivating the application of our approach also to these countries. The contribution of our OSINT-based approach is twofold: First, it facilitates the development of models of real-world power grids, fostering research and discussion that is independent of the power grid operators, in the security domain and beyond. Second, our method represents an attack itself, challenging the energy sector's security-by-obscurity approach.

1. Introduction

For our technology-dependent society, reliable operation of the power grid is fundamental. With the advent of ubiquitous computing and networking, real-world cyber attacks – interrupting utility service and causing blackouts – are on the rise (Lee et al., 2016; Tidy, 2022), and increasingly motivate security researchers to investigate this topic (Dabrowski et al., 2017; Soltan et al., 2018; Huang et al., 2019; Ospina et al., 2021). Thereby, models of real-world power grids are an indispensable prerequisite to assess an attack's aftermath or to design novel protection measures. However, researchers are so far limited to either (I) coarse-grained models allowing only high-level conclusions, or (II) the few models representing actual power grids (but not necessarily for the country of interest) as the following examples show: Dabrowski et al. (2017) model the European power grid as a monolithic block. This approach allows to investigate attacks affecting Europe as a whole, but renders research on more local attacks which impact individual cities or regions only, infeasible. Soltan et al. (2018) even had to use a model of another country for their work on the US power grid. Initially, they used the coarse WSCC 9-bus grid model representing the Western American power grid – serving more than 70 million people – with only nine (!) power lines and changed to a more fine-grained model of Poland afterwards. Yet, simulations for Poland could only be performed for five points in time (in winter 1999/2000, winter 2003/04 and summer 2004) as parameters for generation and

* Correspondence to: University of Vienna, Kolingasse 14-16, 1090 Vienna, Austria.
E-mail address: johanna.ullrich@univie.ac.at (J. Ullrich).

https://doi.org/10.1016/j.cose.2024.104042
Received 25 January 2024; Received in revised form 22 March 2024; Accepted 5 August 2024
Available online 8 August 2024
0167-4048/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

consumption remain unknown for other moments in time; neither it is feasible to modify these parameters to reflect other scenarios (e.g., attacks against certain plant types or the replacement of fossil plants by renewables). Soltan et at. explicitly argue that *"there are no other real power grids at this scale and detail available for academic research"*, which supports our premise. Summarizing, fine-grained models of real-world power grids remain inaccessible for researchers.

Accurate models of real-world power grids do exist – otherwise, power grids could not be operated successfully – but their operators do not make them available, allegedly for reasons of protection. The exception are joint research projects of researchers and the energy sector, for which non-disclosure agreements must be signed, effectively hindering the publication of results and the discussion of their societal impact. From our perspective, *security-by-obscurity* as currently followed by the energy sector is futile: First and foremost, artifacts of the power grid (lines, plants, substations) are large and, also for technical reasons, cannot be easily hidden. Construction and maintenance require permits from state authorities, including public involvement and media coverage, and operators even open their facilities for visitors, e.g., during guided tours. As a public-sector undertaking, they are usually obliged to justify their work to parliament or other authorities. This raises the question whether this public information provides sufficient insight for modeling.

Open Source Intelligence (OSINT), i.e., the collection and analysis of public data, has turned out to be successful in recent years; for example, *Bellingcat* identified the people in charge of the MH17 downing, relying on materials found in social media and other platforms (van Huis, 2018/19). Similarly, it seems promising to infer real-world power grid models from publicly available data; and indeed, multiple models have been generated this way: *PyPSA-Eur* (Hoersch et al., 2018) is based on data provided by *ENTSO-E*, an association representing the European grid operators, *ELMOD* (Egerer et al., 2014; Egerer, 2016) on systematic overview maps of grid operators and national statistics, *SciGrid* (Matke et al., 2016), *GridKit* (Peles, 2021), and *osmTGmod* (Scharf and Nebel, 2016) infer power lines from the collaborative geographic database *OpenStreetMap (OSM)*, *eGo* combines OSM with national statistics, and a further approach relies on satellite images (Keliris et al., 2019). However, *PyPSA-Eur* and *ELMOD* turned out to be inaccurate, presumably due to the inaccuracies or coarseness of the used maps; and more importantly, validation of the remainder, OSM-based models are rare and, unfortunately, superficial, namely a comparison of *SciGrid*'s total line length (Medjroubi et al., 2017) and a check of 3.3% of *eGo*'s substation districts (Hülk et al., 2017). Yet, OSM-based approaches appear promising due to the accuracy of the underlying geographic database.

Questions on the model data's quality, including its completeness and adequacy, remain open, but are of utter importance for (security) research building on this data, especially as the results might have considerable impact on society and national security. The work at hand approaches this particular point and investigates whether OSINT leads to data on real-world power grids of sufficient quality. This paper contributes to security research in several ways:

- We develop a unified and automatable OSINT method to infer a power grid's parameters relevant for power flow analysis – a steady-state analysis of the power grid and de-facto standard approach in power engineering – from OSM, national statistics and other publicly accessible data.
- Following this approach, we infer these parameters for *Austria*, a medium-sized member state of the European Union, and validate them against ground truth from trusted sources, namely *Google Street View*, governmental land use plans and aggregated information materials provided by the energy sector.
- We assess how fast modifications of the power grid are documented in OSM by comparing the OSM's object history with national grid development plans to investigate whether OSM-based models can be kept up-to-date in an automatic way.
- We investigate the extent of power grid documentation in OSM for the other 24 countries that are connected to the Synchronous Grid of Continental Europe to see whether our approach appears worthwhile for application to other countries.

Our results show that the inferred data meets reality well and challenge the energy sector's security-by-obscurity approach. If we – researchers with limited resources – are able to infer such model data, adversaries – military, secret services, terrorists, etc. – can do so as well, thus becoming a threat for national security. Consequently, we recommend including these aspects into national security concepts. At the same time, our approach allows to assess the impact of cyber or terroristic attacks on real-world power grids, and fosters research and discussion in the security domain (and beyond) that is independent from the power sector. Thereby, our approach bears three advantages:

- It does not only allow to simulate all 15-min blocks of the year, but also facilitates easy adaption to investigate new scenarios (e.g., the outage of all gas turbines due to a cyber attack, the removal of a substation due to a terroristic attack, or the replacement of fossil plants by renewables due to energy transition).
- Power grid modifications in the real world are timely documented in OSM, and this way automatically update our model, guaranteeing up-to-dateness of OSM-based model data.
- Actual power grids in other European countries are documented in OSM in a similar extent as in Austria, leading to applicability of our approach to other countries.

Our data sets, which have been inferred following our OSINT approach, are provided to the public as open data.[1]

The remainder of the paper is organized as follows: Section 2 provides background information, followed by Section 3 describing related work. Section 4 gives an overview of our methodology. Then, we focus on processing and validation of data in detail; see Section 5 for power lines, Section 6 for substations, Section 7 for power generation, and Section 8 for consumption. Finally, Section 9 investigates OSM's up-to-dateness, and Section 10 our methodology's transferability to other countries. Our results are discussed in Section 11. Section 12 concludes.

2. Background

This section provides background on power grid terminology (2.1), open-source intelligence (OSINT) (2.2) and the geographic database OpenStreetMap (2.3).

2.1. Power grid terminology

The transmission grid serves to transport electric energy over large distances and operates at levels of extra-high voltage (Austria: 220 kV and 380 kV), while the distribution grid forwards the energy to the individual consumers, operating at lower voltage levels (Austria: 110 kV and below). The 110 kV level is referred to as "high voltage". The operator of the transmission grid is also referred to as Transmission System Operator (TSO), while those operating the distribution grids are the Distribution System Operators (DSO). In Austria, Austrian Power Grid (APG) serves as the TSO, and each of the nine regions is supplied by one or two distribution system operators (DSO). The DSOs typically supply within the boundaries of their regions; discrepancies by a few communities are feasible due to topographic specifics or for historic reasons.

Power lines forward electric energy, and typically contain multiple electric systems – potentially even operating at different voltage levels – which are mounted on the same power poles. From a power engineering

[1] https://github.com/sbaresearch/OSINT_grid_Austria.

Fig. 1. Power pole with four systems and three cables each.

perspective, these electric systems are matters of great interest. As the European power grid is three-phase, each system consists of three cables; this is also apparent as power poles always accommodate cables in multiples of three. Each cable consists of one to three wires; generally, the more wires, the more electrical energy can be forwarded via this system. Cables belonging to the same system have the same number of wires (Kirschner et al., 2007). Please see Fig. 1 for further illustration.

Substations serve as junction points for the lines and transform voltage levels. Thus, substations are typically attributed with two or more voltage parameters. Power plants produce energy by converting other forms of energy, e.g., gas or wind, into electric energy. The electric power is delivered to the next substation where it is transformed to higher voltages and forwarded over the power grid to the consumers. Consumers are either private households, industry, and trade, or governmental institutions, and show different patterns of consumption. For example, private households consume more on mornings/evenings and the weekends, while the others peak consumption is on work days.

2.2. Open source intelligence

Open Source Intelligence (OSINT) refers to the search, collection, processing, and analysis of information from public, accessible and legitimate sources to produce intelligence (Evangelista et al., 2021; Yadav et al., 2023). Initially, military and intelligence services conducted OSINT to gain information on their enemy, e.g., by reading the others' newspapers or listening to their radio broadcasts. With the advent of the Internet, publicly accessible sources have grown significantly, now encompassing social media, crowd-source databases, enterprise website, governmental records, and many more. Beyond that, it is now also feasible to search, collect, process, and analyze large quantities of data in an automatic way, a trend that is further amplified by AI/ML techniques. Thereby, the challenge of OSINT has shifted from finding information to identifying pertinent information among the plethora of different data sources available on the Internet.

The distinct process of OSINT is neither standardized nor defined beyond the reliance on public data, but rather depends on the particular objective of the actors, and typically also requires a basic understanding of the subject, referred to as fluid intelligence (in contrast to crystal intelligence) by Glassman and Kang (2012). Among others, current applications are fighting against cybercrime, cyber threat intelligence, language translation, or investigate journalism (van Huis, 2018/19; Pastor-Galindo et al., 2020).

2.3. OpenStreetMap

OpenStreetMap (OSM)[2] is a crowd-sourced geographic database: volunteers insert objects of the physical environment based on visual inspection, collected GPS data, aerial imaginary or other sources. In the database, individual objects are stored as a combination of geographic data structures (nodes, ways, relations), defining its geographic position in detail, and added tags (key–value pairs providing metadata). An ontology describes tags for various purposes and serves as a manual.[3] Objects related to electric power supply are tagged with the key *power*; its distinct value describes the type of the artifact, e.g., *line*, *plant*, or *substation*. Added tags describe the objects in more detail; according to the manual, they are either recommended or optional, but rarely enforced.

Changesets refer to multiple edits by a single user within a brief period of time,[4] and allow to track past changes in the geographic database, comparable to versioning software. The changesets are opened at the beginning of a modification and closed after its finalization, and are accessible via OSM's history function. An object's status at a specific point in time might also be visually presented, e.g., using overpass turbo.[5] QGIS.[6] is an open-source geographic information system, running locally on your computer, with a graphical user interface that is primarily used in the domain of geography and geoinformation It allows to import data from different geographic databases like OSM, and to process the data as needed. It provides a variety of standard tools such as the calculation of an object's area or length, cutting of lines, etc. Beyond that, QGIS has python support, referred to as PyQGIS,[7] that enables to script all functions that are available via its the graphical interfaces.

3. Related work

Previous work on power grid modeling is distinguished in OSM-based and other approaches. Table 1 provides an overview over related work with a particular focus on the data sources used for modeling. Table 2 shows the extent of validation that has been conducted in the past, and compares it with out approach.

OSM-based approaches: SciGrid (Matke et al., 2016), *GridKit* (Peles, 2021) and *osmTGmod* (Scharf and Nebel, 2016) infer power lines from OSM. They differ in the considered voltage levels – SciGrid and GridKit only consider 220 kV and 380 kV, osmTGmod also 110 kV – and the OSM attributes used for modeling. All three lack generation and consumption which are indispensable for in-depth analysis. A comparison of the three resulting models for Germany shows different numbers of lines and substations (Heitkoetter et al., 2019). Yet, it remains unclear how well these models represent reality as related publications reveal only a single figure: SciGrid's aggregated line length is 95% of the length that is specified by the energy sector (Medjroubi et al., 2017).

eGo (Mueller et al., 2018) uses the lines as inferred by *osmTGmod*, and models consumption by spatially distributing aggregated power consumption based on land use and population density (Hülk et al., 2017). While a validation of the model's topology is missing (see discussion on osmTGmod above), the authors can validate the assignment of consumption to distribution substations. While assignment based on distances works well, generation per substation district is underestimated and demand overestimated. However, one must bear in mind that the validated substation districts represent only 3.3% of the total (and are therefore not necessarily representative), especially as structural characteristics might differ in other parts of the country.

[2] https://www.openstreetmap.org/.

[3] https://wiki.openstreetmap.org/wiki/Map_features.

[4] https://wiki.openstreetmap.org/wiki/Changeset.

[5] http://overpass-turbo.eu/.

[6] https://www.qgis.org.

[7] https://docs.qgis/3.34/en/docs/.

Table 1
Data sources of related work. SciGrid, GridKit, osmTGmod, and eGo are like our work primarily based on OSM, while the others rely on overview maps from the energy sector.

	Country	Lines	Substations	Generation	Consumption
OSM-based approaches:					
SciGrid	Germany	OSM	OSM	–	–
GridKit	Germany	OSM	OSM	–	–
osmTGmod	Germany	OSM	OSM	–	–
eGo	Germany	OSM	OSM	Open power system data	National statistics
Our work	Austria	OSM	OSM	OSM	National statistics
Other approaches:					
PyPSA-Eur	Europe	ENTSO-E map	ENTSO-E map	Powerplantmatching database	ENTSO-E database, national statistics
ELMOD-DE	Germany	TSO maps	TSO maps	BNetzA database	ENTSO-E database, national statistic

Table 2
Validation of models. ○ no validation, ◑ partial validations, ● full validation, ✖ failed validation. Previous OSM-based models were only partially validated, e.g., aggregated line lengths or the assignment of consumption to substations.

	Lines	Substations	Generation	Consumption
SciGrid	◑	○	-	-
GridKit	○	○	-	-
osmTGmod	○	○	-	-
eGo	○	○	○	◑
Our work	●	●	●	●
PyPSA-Eur	✖	✖	✖	✖
ELMOD-DE	✖	✖	✖	✖

Other approaches: PyPSA-Eur (Hoersch et al., 2018) combines the ENTSO-E map, ENTSO-E's national time-series on power demand, and the *powerplantmatching* project to create a model of the European power grid. While the aggregated length of lines appears to be plausible, further validation reveals significant mismatch and the underlying reason is found in the ENTSO-E map's inaccuracy; artifacts are either depicted multiple tens of kilometers apart from their actual location, or remain detached from the grid at all. ELMOD (Egerer et al., 2014; Egerer, 2016) relies on coarse-grained overview maps of the grid operators, that are created for public relations and not intended for technical purposes, and the ENTSO-E time-series to model the German power grid. Beyond the grid's topology, it also includes a model on price formation that yields higher prices than reality. It remains, however, unclear whether these deviations are caused by inaccuracies in topology or economic modeling.

Summarizing, hitherto existing approaches either relied on too coarse-grained maps from the energy sector and turned out to be inaccurate during validation (PyPSA-Eur, ELMOD), or base on the more fine-granular OSM and were only partially validated (SciGrid, GridKit, osmTGmod, eGo), see Tables 1 and 2. While OSM-based models appear promising, it is unclear whether the resulting power grid models meet reality. Knowledge on their quality is however of utter importance for (security) research building on these models, especially as the results might have considerable national and societal impact. This paper overcomes this gap and investigates whether OSINT leads to data on real-world power grids of sufficient quality by the example of *Austria*, a medium-sized member state of the European Union.

- For the very first time, we fully validate all aspects of an OSM-based power grid model inferred by OSINT, see Table 2 for comparison with related work, and indeed the resulting model reflects reality well.
- We are the first to investigate how fast power grid modifications are documented in OSM. The results provide an insight on how fast an OSM-based model is able to adapt to power grid modifications in the real world.
- We conduct multiple analyses to determine whether our approach, applied to Austria, is also adaptable to other countries, allowing the generation of further models according to our methodology, i.e., it is the first generalizable approach allowing the efficient creation of multiple models.

We decided for the example of Austria for two reasons: First, previous work focused on Germany, and might have motivated individuals to improve the grid's representation in OSM, potentially biasing our results; whereas in Austria, such modeling has been yet unknown. Second, we have access to sufficient ground truth, enabling comprehensive validation of the inferred model data.

4. Methodology: An overview

As typical in the field of security, we first define a threat model (4.1) and infer research questions based on the threat model (4.2). Then, we describe the data sources, that are accessible to the adversary, in more detail (4.3), and our approach on how an adversary could exploit these sources to gain sufficient data for power grid modeling (4.4). Finally, we describe ground truth data used for validation (4.5), and elaborate on the independence of this ground truth data from the data sources serving as an input of our approach (4.6).

4.1. Threat model

The adversary[8] needs model data of the national power grid to plan a (physical or cyber) attack against the power grid. For modeling, the adversary looks to gain information about the power grid and its characteristics. In particular, the adversary needs details about (see also Fig. 2):

- *Electric systems* transfer electric energy from A to B. For analysis, the adversary needs a data set covering the high (110 kV) and extra-high (220 kV and 380 kV) voltage systems. Furthermore, they need the electrical parameters *resistance R*, and *reactance X* for each system.
- *Generation* is carried out in power plants. This means that the adversary needs to know the power plants, including their parameters *output*, typically provided in Megawatts, and *type*, e.g., hydro, solar, gas, nuclear. The latter is necessary to infer the characteristic behavior over time. For example, nuclear plants continuously supply output over the year, while hydro plants produce in dependence of the rivers' water level.
- *Consumption* uses transferred energy for various purposes in households, agriculture, industry or commerce. The adversary needs not only to know which amounts of electric energy is consumed, but also its geographic location and point in time. Thus,

[8] The adversary might also be researchers needing a power grid model for their work, e.g., to assess the feasibility/aftermath of an attack or to plan deployment of renewables.

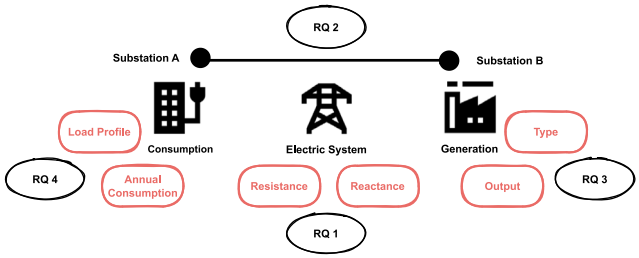


Fig. 2. Threat model and Research Questions (RQ). For success, the adversary needs to know electric systems, substations, generation and consumption parameters (as indicated in the red boxes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Research questions. RQ1 to RQ5 consider modeling by the example of Austria; RQ6 investigates the adaption of our approach to other countries.

Research Questions
RQ1: Is it feasible to infer the individual electric systems, including their electric parameters resistance R and reactance X, from OSM and other public sources?
RQ2: Is it feasible to infer all the substations at high and extra-high voltage level from OSM?
RQ3: Is it feasible to infer all power plants, including their production capacity and primary energy source?
RQ4: Is it feasible to spatially distribute the aggregated numbers on consumption using statistics on population and commercial activity? Beyond, is it feasible to connect this consumption with substations?
RQ5: How fast are actual modifications of the power grid documented in OSM?
RQ6: Are other European countries documented to a likewise extent in OSM as Austria, enabling adaption of our methodology to these countries?

the adversary seeks to infer consumption at the granularity of distribution substations in intervals of 15 min, or alternatively the annual consumption and load profiles to spread the aggregated consumption over time.

- *Substations* are the hubs of the power grid. Thus, the adversary needs to know how they relate to electric systems, generation and consumption. Electric systems connect two substations; whereas, generation, and consumption is attributed to a single substation connecting the respective power demand/provision with the power grid.

Depending on the intended purpose, models vary in their level of detail. Consequently, they vary in their requirements on model data. For this work, we assume that the adversary eventually wants to conduct power flow analysis that is comparable to the work of Soltan et al. (2018) using the Polish model; in their work, the authors successfully investigated manipulation-of-demand (MAD) attacks. Thus, we have defined the necessary parameters in alignment with this model.

4.2. Research questions

Accurate models of real world power grids do exist but their operators do not make them available, see our motivation in Section 1. Consequently, an adversary is only able to access public data to gain the needed information, i.e., the adversary has to conduct OSINT. Based on the threat model, we thus define the following research questions guiding through our methodology and the remainder of the paper.

- In OSM, power lines are documented as geographic objects. But for simulation, the adversary requires them at the level of electric systems leading to the following research question.
  RQ1: *Is it feasible to infer the individual electric systems, including their electric parameters resistance R and reactance X, from OSM and other public sources?* Details on methodology and results are found in Section 5.
- For full representation of the power grid, the adversary also requires knowledge on the substations connecting the electric systems with power generation and consumption. This leads to second research questions.
  RQ2: *Is it feasible to infer all the substations at high and extra-high voltage level from OSM?* Details on methodology and results are found in Section 6.
- Electric power generation occurs in power plants of different types and size. The first defines the primary energy source transformed into electric energy, and defines the power plant's production pattern over time, e.g., photovoltaics produce during sunshine while nuclear plants provide almost constant output. The latter defines the maximum output that can be fed into the power grid
  RQ3: *Is it feasible to infer all power plants, including their production capacity and primary energy source from OSM?* Details on methodology and results are found in Section 7.
- While generation is centralized in power plants, electric consumption in households, agriculture, industry or commerce is spatially spread all over the country. Yet, national statistics only provide aggregated numbers on consumption per region.
  RQ4: *Is it feasible to spatially distribute the aggregated numbers on consumption using statistics on population and commercial activity? Beyond, is it feasible to connect this consumption with substations inferred in the second step?* Details on methodology and results are found in Section 8.
- The power grid undergoes continuous, though slow modifications that need to be incorporated into an up-to-date model reflecting reality well, leading to another research questions.
  RQ5: *How fast are actual modifications of the power grid documented in OSM?* Details on methodology and results are found in Section 9.
- RQ1 to RQ5 consider power grid modeling by the example of the European state of Austria. As a final aspect, we consider whether our methodology is adoptable to other countries. As most of the input data comes from OSM, this leads to the final research question.
  RQ6: *Are other European countries documented to a likewise extent in OSM as Austria, enabling adaption of our methodology to these countries?* Details on methodology and results are found in Section 10.

For readability, the research questions are also provided in Table 3.

4.3. Data sources

Since power grid operators do not publish their model data, the adversary can only access public data to gain the required information. The following enumeration only includes publicly accessible data that has been used in this work.

- *OpenStreetMap* (OSM) provides free and open geographic data and information. Its data is crowd-sourced by volunteers and, among others, holds data related to the power grid.
- *Standard books about power engineering* provide insight into how electrical parameters depend on given system characteristics. They allow to infer technical parameters that are necessary for power flow analysis but unavailable in OSM. Such books, e.g., Kiessling et al. (2003), are typically used in courses for undergraduate and graduate students.
- *Standards* unify technical characteristics and allow to narrow down the range of electrical parameters' values in our scenario. EN50182/2001 (European Standards, 2013) defines a set of wire types for multiple European countries including their technical parameters.

Table 4
Overview on research questions, inferred power grid artifacts, data sources, ground truth, and section.

	Sources						Ground truth					Section
RQ artifacts	OSM	Standard books	Standards	National statistics	Expert reports	Load profiles	Google Street View	DSO/TSO data	Power plant map	Land use plans	Energy Mosaic	
1 Electric systems	✓	✓	✓	–	✓	–	✓	✓	–	✓	–	5
2 Substations	✓	–	–	–	–	–	✓	✓	–	–	–	6
3 Power plants	✓	–	–	–	–	–	✓	–	✓	–	–	7
4 Consumption	✓	–	–	✓	–	–	–	–	–	–	✓	8
5 Modifications	✓	–	–	–	–	–	–	✓	–	–	–	9
6 Adoption	✓	–	✓	–	–	✓	–	–	–	–	–	10

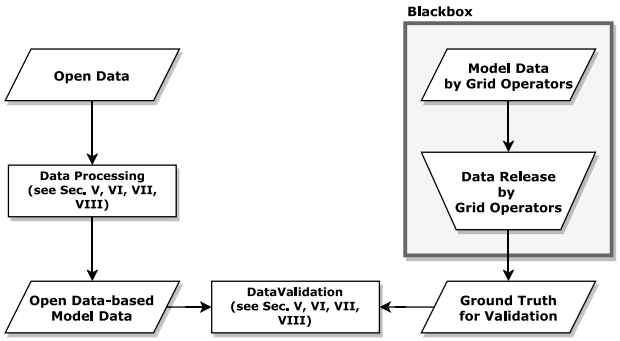


Fig. 3. Methodology. Model data is inferred from open data, and validated against ground truth, i.e., aggregated materials on the power grid from trusted sources.

- *National statistics* provide relevant figures on aggregated power consumption of individual regions, population per community, and economic parameters like gross added value (GVA) per sector/region.
- *Expert reports and official documents* are involved in governmental processes (e.g., environmental impact assessments), required for the construction of new lines and power plants, and typically published on the Internet. While the reports are tailored to the individual line or plant in planning, they allow to infer on characteristics of similar artifacts.
- *Load profiles* serve as a basis for billing among market participants, and must be provided by the national regulatory authority (Power Clearing & Settlement Austria, 2021). If the annual consumption is known, the load profiles allow to infer the share of consumption per 15-min-block of the year.

Table 4 provides an overview on the research questions, the inferred power grid artifacts and the data sources used for this purpose.

4.4. Approach

In this paper, we conduct the following steps, see also Fig. 3, to clarify whether model data, that is inferred from publicly accessible sources, comprehensively represents the actual power grid.

Data processing First, we act like an adversary and process publicly accessible data (open data) to gain model data for *Austria*. The details of data processing, and the results are described in the respective subsections on *Approach and Results* in Sections 5 (power lines), 6 (substations), 7 (power plants), and 8 (consumption).

Data validation An adversary would then arrange the obtained open-source based model data for a simulation tool, e.g., MATPOWER,[9] to conduct a power flow analysis. This way, they would be able to investigate the effects of a cyber or physically launched attack against the power grid, or optimize such attacks, e.g., by identifying and overloading the grid's weakest link. For example, Soltan et al. (2018) used such an analysis to assess the consequences of MAD attacks.

The mere convergence of a power flow analysis does not imply the representativeness of its input data. Validation of its results against ground truth remains, however, infeasible. Such insights into the power grid's operational state are not available to the public. In fact, this is the very motivation of our work. Consequently, we assume that a high-quality validated input data would be the best possibility to gather the power grid's operational state, and focus on validation of the obtained data instead. Therefore, we compare the obtained model data against ground truth, i.e., data that is provided by trusted sources, see Section 4.5. This way, we are able to assess whether the model data represents the real-world grid with sufficient quality. Validation and its results are described in the respective subsections on *Validation* in Sections 5 (power lines), 6 (substations), 7 (power plants), and 8 (consumption). In the following Section 4.6, we discuss the independence of open data, serving as an input to our approach, and ground truth used for validation.

4.5. Ground truths for validation

The ground truth for validation consists of the following data sets. The following enumeration only includes ground truth that has been used in this work.

- *Google Street View:* Through optical inspection of the power grid's artifacts on *Google Street View*,[10] we validated whether parameters found in OSM are consistent with the actual deployment.
- *Coarse-Grained Data from DSO/TSOs:* Providers do not publish their models but some of their aspects, typically in an aggregated or coarse-grained form. This way, we found a list of static network data provided by APG, the national TSO (Austrian Power Grid, 2021c), encompassing its 220 kV/380 kV electric systems, including length and other parameters. Furthermore, an overview map of these lines and the TSO's substations is available (Austrian Power Grid, 2021a). Finally, overview maps of distribution substations are available for three Austrian regions (Werner, 2012b; Netz Niederösterreich, 2021; Vorarlberger Übertragungsnetz GmbH, 2021), and aggregated numbers about the number of substations are found for other regions in presentation slides, network development plans, and on websites.
- *Overview Map on Power Plants:* This map shows all plants with more than 10 MW of output (OEsterreichs Energie, 2021). It includes information on the plants' power output, power source, and approximate location. It is maintained by the lobbying and advocacy group for the national energy sector, representing more than 140 actors in the field and can therefore be considered complete.
- *Land Use Plans:* Austrian communities are obliged to maintain maps defining the potential use of land (residential, agricultural, etc.) and make them accessible to the public. The maps must include electric lines and related artifacts and are accessible online. Unfortunately, each region maintains its own web map, and the granularity of the provided plans differ. While some regions include the lines and voltage levels, others only show the lines or do not record them at all.

[9] https://matpower.org.
[10] https://www.google.com/streetview/.

- *Energy Mosaic Austria:* This previous research project has determined energy consumption and CO2 emissions per community (Abart-Heriszt et al., 2019a,b) using more than 90 parameters. The provided results are the total energy consumption per community, including other forms of energy than electrical (e.g., thermal). While the mere electrical consumption is not publicly available, the authors provided us with this data.

Table 4 provides an overview on the research questions, the inferred power grid artifact, the data sources and ground truth used for this purpose.

4.6. Independence of open data & ground truth

One could argue that modeling would be easier if done directly based on ground truths as the latter are published by grid operators and other official entities. Further, it could be argued that the data found in OSM was inferred from ground truth, and the data sets are dependent on each other. We strongly argue against both claims.

Primarily, ground truth data is aggregated (see Fig. 3) and inferred from classified model data. Therefore, ground truth lacks relevant information and cannot be used for precise modeling. For example, geographic locations are provided at the granularity of multiple tens of kilometers and cause misalignment, particularly when merging individual data sets (e.g., an overview map on substations with another one on power plants). In this context, we also refer to the discussion on the models *PyPSA-Eur* and *ELMOD*, see Section 3; they are known to be inaccurate due to misalignment of geographic positions.

This argument holds similarly for OSM data: OSM consistently provides more information (e.g., precise location, number of wires and cables) than the aggregated ground truth. This means that users digitalizing an artifact in OSM need to gain more information by visual inspection, which is, either in reality, on satellite images or by *Google Street View*, *the* intended way to augment OSM. Beyond, OSM data is also available in regions where ground truth (e.g., land use plans) is unavailable. Nevertheless, it cannot be ruled out that OSM users have used ground truth for validation in a comparable manner as we did, i.e., that they checked the OSM data for completeness.

5. Topology: Power lines and electric systems

In OSM, power lines are documented as ways, i.e., linear features on the ground, enhanced by tags such as *voltage*, *wires* or *cables* providing further information about a line. From an electrical engineering perspective, however, electric systems, including their electric parameters resistance R and reactance X, are the matter of interest. Therefore, the power lines from OSM, typically containing multiple such systems, have to be split, and their length needs be inferred to eventually calculate R and X. In this section, we investigate whether it is feasible to infer all electric systems at high and extra-high voltage in Austria from public sources (RQ1). In particular, we describe how data on power lines is extracted from OSM, and further processed to gain electric systems and their electric parameters (5.1). Afterwards, we assess the quality of extracted extra-high voltage (5.2) and high voltage systems (5.3) by comparing them with ground truth.

5.1. Approach and results

For electric systems, we conducted the following three steps, see also Fig. 4:

1. *Line Extraction & Length Calculation:* From OSM, we imported all power lines with voltage level of 110 kV, 220 kV or 380 kV into QGIS (date of download 2020/02/17), and calculated their length using QGIS built-in *Field calculator*.

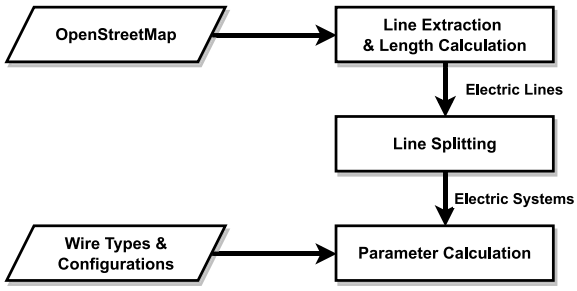


Fig. 4. Methodology. Power lines are extracted from OSM, split into systems, and enhanced by electric parameters.

Table 5
Common wire types (EN40182/2001) and configurations in Austria (inferred from planning applications, expert reports and construction plans).

Volt.	No. wires	Wire type	Cable distance	Comment
110 kV	*	238-AL1/82-ST1A	4.5 m	Old lines
110 kV	*	562-AL1/49-ST1A	4.5 m	New lines
220 kV	*	341-AL1/109-ST1A	8.6 m	–
380 kV	2	679-AL1/86-ST1A	8.0 m	–
380 kV	3	635-AL1/117-ST1A	8.0 m	–

2. *Line Splitting:* For splitting lines into individual electric systems, we exported the data from QGIS, and further processed it using python scripts. We inferred the number of electric systems per line from OSM's *cables* tag, showing the number of electrically power-carrying conductors. The individual characteristics (voltage level, wires, length) of each system are inferred from the line's characteristics. If only one value was provided, it is assumed for all electric systems. If multiple values were provided, e.g., systems of different voltage levels, we assigned the numbers according to the first digit of the numbers in the *ref* tag, a unique ID for systems and displayed at the power poles.
3. *Electric Parameter Calculation:* Finally, we calculated the systems' resistance R and reactance X,[11] according to the following equations (Kiessling et al., 2003):

$$R = \frac{R' \cdot l}{n} \tag{1}$$

$$X = f \cdot \mu_0 \cdot l \cdot [ln(\frac{D}{r_0}) + \frac{1}{4n}] \tag{2}$$

The number of parallel wires n is readily available in OSM's *wires* tag, and a system's length l has been calculated in our method's first step. μ_0 is the magnetic constant ($4\pi \cdot 10^{-7}$ H/m), also referred to as the permeability of the vacuum, and f the nominal system frequency of 50 Hz. Easing to infer the remainder parameters, grid operators tend to limit themselves to a restricted set of wire types for simplified maintenance, operation, and procurement. We found commonly used wire types for different voltages and number of wires from planning applications (Fichtinger and Kostner, 2013; Ämter der Steiermärkischen Landesregierung und der Burgenländischen Landesregierung, 2004; Bruny et al., 2016), expert reports (Oswald, 2007), construction plans (Weniger, 2019), and public tenders (Netz Niederösterreich GmbH, 2018), see Table 5 for an overview. With this information, we are able to infer R', the DC resistance per unit length, and r_0, the radius of the individual wire, from the standard *EN50182/2001* (European Standards, 2013). Also, typical values for D, the distance between a system's cables, are inferred from the same documents

[11] For aerial lines, capacitance is negligible as distances between cables exceed the individual cables' diameter.

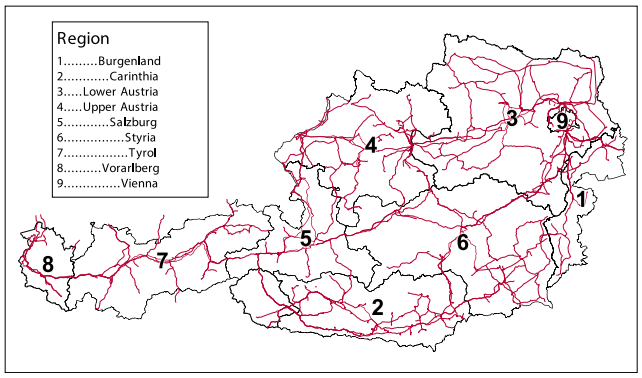


Fig. 5. Results. Electric lines in Austria as inferred from OSM.

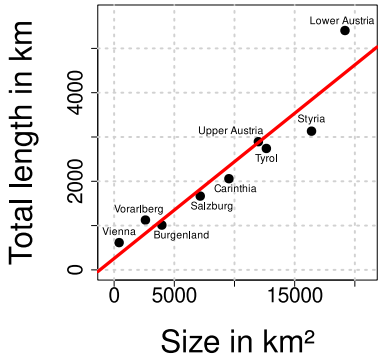


Fig. 6. Total system length wrt region size. $R^2 = 0.897$. The larger a region, the more electric systems are found in OSM.

Within the borders of Austria, we found 2048 line objects and a total of 20,646 km of electric systems, see the map in Fig. 5. Fig. 6 shows the total electric system length of a region (y-axis) w.r.t. the region's size (x-axis), including a regression. For the nine Austrian regions, the figure shows that the larger a region, the more electric system length is found. This matches our expectations as a first approximation all of Austria is supplied. For a detailed presentation of line objects and aggregated line length per voltage resp. region, we refer to Tables 8 and 9 in the Appendix.

5.2. Validation - Extra-high voltage

For validation of the extra-high voltage systems, we relied on static network data provided by the Austrian TSO (Austrian Power Grid, 2021c). It encompasses a list of 104 electric systems of 220 kV resp. 380 kV, including their start and end location (at the granularity of community), length, resistance, and reactance. First, we checked whether all lines from the TSO are also present in our data (*completeness*); second, we conducted an inverse check to verify whether all lines from our data set are also available in the TSO's data set (*coverage*). Finally, we compared our results for the systems' length, resistance, and reactance with the values provided by the TSO.

Completeness & coverage: Ground truth encompasses 104 electric systems; thereof, six systems are extremely short (20–600 m) and connect two adjacent substations. As operation of adjacent substations is rather an organizational than a technical necessity – for example, one substation is operated by the TSO, the other by the regional DSO – we considered these substations as one, and consequently omitted these systems in our analysis, leaving 98 lines for validation. Considering completeness, we found 98 of these 98 systems in our data set **[*KI1 - Completeness 100% (98/98)*]**.[12] All their voltage specifications were correctly captured **[*KI2 - Voltage Specifications 100% (98/98)*]**. With regard to coverage, we found an additional four systems (beyond the 98 from ground truth) in our data set **[*KI3 - Coverage +4% (4/98)*]**. According to Google Street View, all of them are existing in reality and connect power plants with substations. They might belong to another operator and are thus not included into the TSO's data used as a ground truth.

Electric parameters: For the 98 systems included into validation, we compare the length inferred from OSM using QGIS (Fig. 7(a)), as well as our values for resistance R (Fig. 7(b)), and reactance X (Fig. 7(c)) – calculated according to formulas above – with the parameters provided by the TSO's ground truth. In all figures, the x-axis reflects the ground truth values and the y-axis our calculated values. Ideally, the results yield a straight line with a slope of one, i.e., our values match ground truth, and indeed our parameters meet reality well. The average estimation errors are 0.76 km (length) **[*KI4 - System Length 0.76 km (est. err.)*]**, 0.17Ω (resistance), and 0.76Ω (reactance) **[*KI5 - Resistance R & Reactance X - 0.17 Ω, 0.76 Ω (est. err.)*]**. Fig. 7(a) shows a clear outlier for length (Inferred: 93 km, TSO: 61.6 km). It appears to be an error on behalf of the TSO as the linear distance between the line's endpoints is already more than 61.6 km, a governmental document supports our result (Land Steiermark - A13 Umwelt und Raumordnung, 2021), and the calculated electric parameters, both reliant on system length, are correct when including our results for length into calculation.

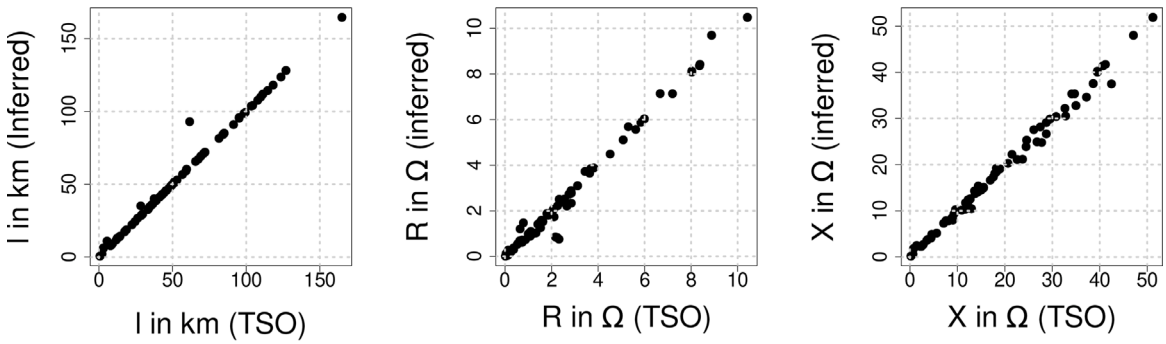
5.3. Validation - High voltage

For 110 kV systems, such final data for validation including system length, resistance and reactance is unavailable, but validation of the extra-high voltage systems, as conducted in Section 5.2, has shown high congruence of our inferred data with reality and suggests that our way to infer these results from the input data is correct. For 110 kV systems, we thus validate the data as inferred from OSM, serving as an input for our parameter calculations, and assume that correct input will lead to correct output following the already confirmed calculation methods.

At the voltage level of 110 kV, validation is done in a twofold way: First, we validate our data against land use plans to verify completeness and coverage of the OSM-inferred lines. It also allows to check geographical positioning of the lines, an essential prerequisite to correctly calculate their length (see first step in Fig. 4), and the systems' electric parameters afterwards. Second, we verify the OSM tags *cables* and *wires* against Google Street View. The *cable* tag shows the number of parallel electric systems, and is thus essential for splitting lines into individual systems (see second step in Fig. 4); *wires* refers to the number of parallel conductors of a system, and is fundamental for the calculation of the electric parameters resistance and reactance (see third step in Fig. 4).

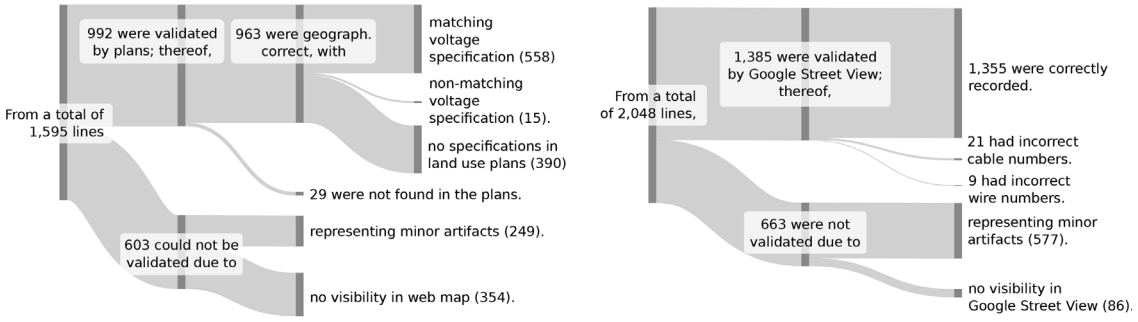
Completeness & coverage: We imported the available regional land use plans into QGIS, and manually compared whether an OSM-inferred line is also present in the land use plans and vice versa. With this approach, we were able to compare seven out of nine regions; thereof, the land use plans of three regions also provided voltage specifications, allowing an even more detailed comparison. For two regions, validation was not possible as their land use plans did not include any information on the power grid infrastructure. An overview on the feasible types of validation per region is found in Table 11 in the Appendix.

[12] From our results, we highlight key indicators by KI and an increasing integer to support readability. We use this indicators whenever referring to this result in the boxes on key findings at the end of the sections or in the final overview tables, see Table 7.



(a) System length l. Est. error: 0.76km (b) Resistance R. Est. error: 0.17 Ω (c) Reactance X. Est. error: 0.76 Ω

Fig. 7. Inferred system length, resistance, and reactance (y-axis) wrt to TSO specifications serving as a ground truth (x-axis). The straight line with a slope of 1 reflects high congruence of our values with reality.



(a) Validation of 110kV lines with Land Use Plans. (b) Validation of 110kV lines with Google Street View.

Fig. 8. Sankey diagrams. Validation of high voltage lines against ground truth. The total amount of lines differ as two regions do not include power grid artifacts in their land use plans and were thus omitted in the validation (left figure).

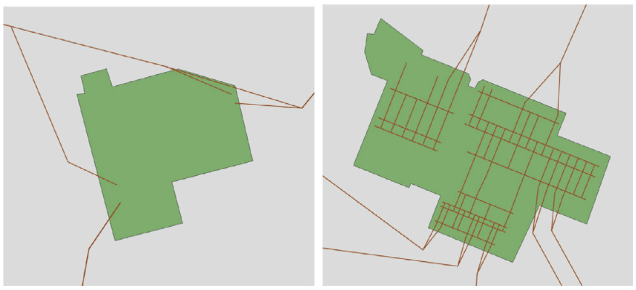


Fig. 9. Substation in *Burgenland* (left) and *Carinthia* (right). The level of details varies among substations, and is dependent on the preferences of the OSM user digitalizing the object.

Regarding completeness, our validation reveals that all lines of the land use plans have also been found in our OSM-based data set **[*KI6 - Completeness: 100%*]**. Regarding coverage (see also Fig. 8(a)), from a total of 1595 line objects associated with 110 kV found in OSM, we were able to validate 992 lines against ground truth (land use plans). The other 603 lines were not included into validation, either due to representing minor artifacts of substations in *Carinthia*, *Upper Austria*, and *Tyrol* that are too detailed for comparison (249) (see Fig. 9), or lacking land use plans in the web maps in these regions (354). From the 992 included into validation, 963 were geographically correct. If voltage specifications were available (573), they were correct for 558 lines objects **[*KI7 - Voltage Specifications 97% (558/573)*]**, and incorrect for 15. For 390 geographically correct lines, we could not check the voltage due to lacking specifications in the land use plans of the respective regions serving as ground truth. An overview per region is found in Table 10 in the Appendix. Beyond, our OSM data set reveals an additional 29 lines that are not available in the land use plans. We were able to determine the existence of 19 of these lines using Google Street View, i.e., they exist in reality; for the other 10, we were lacking proper sight for validation and are unable to verify whether they exist or not **[*KI8 - Coverage +3% (29/992)*]**. For readability, Fig. 8(a) presents the just described results also as a Sankey diagram.

Electric parameters: For each line in our data set, we looked for a distinctive street junction in OSM, searched for the respective location in Google Street View, and manually counted the cables and wires of the lines leading to the following results (see also Sankey diagram in Fig. 8(b)): From the 1385 lines that could be validated against Google Street View, 1355 were correctly captured by OSM about their cables and wires **[*KI9 - OSM Parameters 98% (1355/1385)*]**. In 21 cases, the cables were wrong and would consequently result in wrong numbers of electric systems when splitting the lines (see second step in Fig. 4); in nine cases, an incorrect number of wires would result in wrong electric parameters (resistance R, reactance X, see third step in Fig. 4). As in the validation against the land use plans, we had to

exclude 663 lines from our validation. 577 due to being minor artifacts (see again Fig. 10), and 86 lines without visibility in Google Street View. For region-wise results, we refer to Table 12 in the Appendix.

Key Findings

OSM data reveals a completeness of 100% for extra-high and high voltage **[*KI1, KI6*]**, i.e., all lines/systems from ground truth were also available in our data. Most notably, we found an additional 33 lines **[*KI3, KI8*]** and could verify their existence in reality on *Google Street View* for 23 thereof, emphasizing OSM's congruence with the power grid's actual deployment and independence from ground truth (see discussion in Section 4.6). Beyond, voltage specifications were widely correct, though, extra-high voltage **[*K2*]** appears to be slightly better documented than high voltage **[*K7*]**. For extra-high voltage systems, we were able to calculate the parameters length, resistance, and reactance with high accuracy **[*K4,K5*]** For comparison, the estimation errors for resistance R and reactance X are lower than their change that is associated with a (common) 15 °C temperature shift. For high voltage, we could not directly compare the resulting parameters resistance R and reactance X, but the input data that is necessary for their calculation. OSM tags *cables* and *wires* were correct in 98% of the cases **[*KI9*]**. The correct number of cables allows adequate splitting of lines into individual electric systems (see second step in Fig. 4), and correct wires allow calculation of resistance R and reactance X (see third step in Fig. 4).

6. Topology: Substations

For a full representation of the Austrian power grid, the adversary also requires knowledge on the substations connecting the electric systems with each other, but also to connect the grid with electric power generation and consumption. In this section, we investigate whether it is feasible to infer all substations at high and extra-high voltage from OSM (RQ2), gaining a complete picture of the Austrian grid topology. In particular, we describe how substations were inferred from OSM (6.1), and validated against ground truth (6.2 for extra-high, 6.3 for high voltage).

6.1. Approach and results

We selected all OSM objects with the *power* tag's value *substation*, and assigned them a buffer zone with a radius of 1000 m. If there is a line with 110 kV, 220 kV, or 380 kV within this buffer, we exported the respective substation. From this list, we excluded substations that are identified as plants by their name, and those without a name and voltage level. In total, we found 376 substations spread over all regions of *Austria*, see Fig. 10.

6.2. Validation - Extra-high voltage

For validation of the substations running at extra-high voltage (220 kV and 380 kV), we relied on an overview map of the Austrian TSO (Austrian Power Grid, 2021a). It encompasses 55 substations; thereof, 54 are within the national borders of Austria. We found 53 of these 54 substations in our data set **[*KI10 - Completeness 97% (53/54)*]**. A manual analysis suggests that the missing substation Sattledt is digitalized in OSM; however, it is only represented by two stubs without any further labeling and thus not included into our data set. Beyond, we found an additional substation, namely Villach, that is attributed to the TSO in OSM, but not yet represented in the TSO's overview map **[*KI11 - Coverage +2% (1/54)*]**. The substation has been brought online in 2021, but it had been digitalized in OSM by volunteers already during its construction phase.

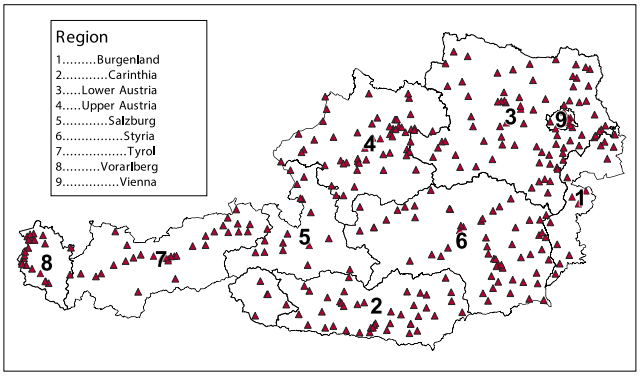


Fig. 10. Results. Substations in Austria as inferred from OSM.

6.3. Validation - High voltage

For validation of the 110 kV substations, we relied on data provided by the regional DSOs. The level of details varies: For three regions (*Burgenland* (Werner, 2012b), *Lower Austria* (Netz Niederösterreich, 2021), *Vorarlberg* (Vorarlberger Übertragungsnetz GmbH, 2018)), we found overview maps including the substations' name and approximate geographic location. For five regions, we found total numbers on substations (*Carinthia* (Kärnten Netz, 2021), *Upper Austria* (Abart et al., 2021), *Styria* (Steiermark, 2021), *Tyrol* (TINETZ, 2021), *Vienna* (Wiener Netze, 2021)).

Validation with overview maps: We manually compared the overview maps with our data set to see whether our OSM-inferred substations are also present in the official maps and vice versa. From a total of 149 substations in these overview maps, 131 were found in our OSM-based data set **[*KI12 - Completeness 87% (131/149)*]**, 18 were not (*Burgenland*: 1, *Lower Austria*: 14, *Vorarlberg*: 3). Beyond, OSM reveals additional substations that do not appear in the overview maps (*Lower Austria*: 4, *Vorarlberg*: 3) **[*KI13 - Coverage +5% (7/149)*]** The latter substations could be verified to be existent by Google Maps, serving the railway system (2), a nearby power plant (3), an industrial zone (1), or a practical exclave that is supplied by energy from neighboring *Germany* (1). These substations might not be included in the overview maps as they are operated by other organizations than the regional DSOs providing the ground truth. Our results also show that our approach does not misclassify other objects (e.g., low voltage substations supplying the last mile) as distribution substations.

Validation with DSO specifications For all regions, the total numbers of substations in our data are lower than those provided by the DSOs (*Carinthia*: 46 vs. 50, *Upper Austria*: 60 vs. 86, *Styria*: 60 vs. 63, *Tyrol*: 46 vs. 47, *Vienna*: 13 vs. 46). It remains, however, unclear whether the provided numbers include the substations of the transmission grid resp. whether all substations operated by the TSO or all in the respective region are included. Beyond, the definition of the numbers could even vary among the different regions; accordingly, the respective numbers should be treated with caution. Yet, our approach appears to underestimate the number of substations, and substations of the distribution grid are less likely to be documented in OSM than their counterparts in the transmission grid, running at higher voltages. The gap is particularly significant for the urban *Vienna*, presumably as many of the substations are indoor and thus not recorded in OSM. For a more detailed version of our results, we refer to Table 13 in the Appendix.

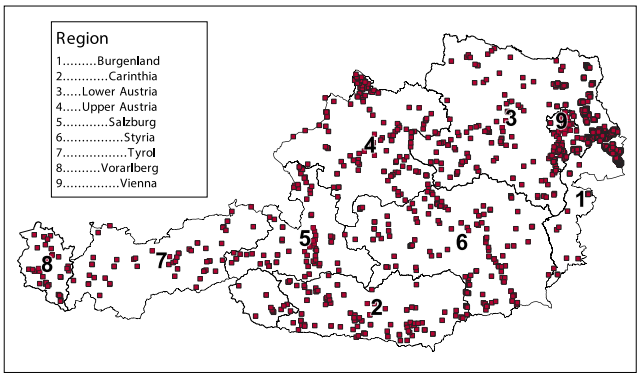


Fig. 11. Results. Power plants in Austria as inferred from OSM.

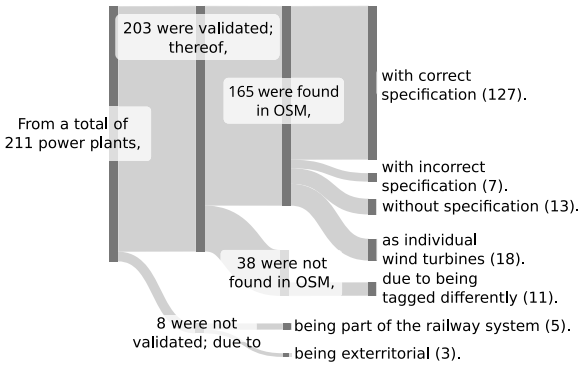


Fig. 12. Validation of plants with overview map (Sankey diagram).

Key Findings

Completeness of substations is high, though, substations operating at extra-high voltage are **[*KI10*]** again better documented than those operating at high voltage **[*KI12*]**. Regarding coverage, we found an additional extra-high voltage substation that was still under construction **[*KI11*]** and seven high voltage substations **[*KI13*]**, emphasizing again OSM's up-to-dateness in comparison to the official materials of the operators, see also Section 9 for more discussion. Considering region-wise numbers, substations in urban areas appear to be underreported due to being indoor facilities.

7. Production: Power plants

Electric power generation occurs in power plants of different size, referred to as capacity and measured in the unit of Watts, and types, e.g., hydro, solar, gas, nuclear. While the first describes the maximum power that might be delivered to the power grid by a plant, a plant's primary source defines its generation profile over time. For example, solar power feeds the power grid when the sun is shining, hydropower is dependent on the water level typically providing more in spring and autumn than in other seasons, and nuclear plants provide almost constant output over the year due to being largely independent from weather conditions. In this section, we investigate whether it is feasible to infer all Austrian power plants, including their production capacity and primary energy source, in order to model their generation behavior (RQ3). In particular, we describe how power plants were inferred from OSM (7.1), and validated against ground truth (7.2).

7.1. Approach and results

In OSM, power plants are identified by the *plant* value; we exported these objects and included them into our data set. Parameters that are needed for power plants are its maximum capacity in Watts (tags *plant:output:electricity*) and their type (tags *plant:source*). In total, we found 1047 power plants within the national borders of Austria, see Fig. 11; their output is between 22 kW and 832 MW.

7.2. Validation

For validation, we relied on an overview map on power plants (OEsterreichs Energie, 2021), providing all power-generating facilities with an installed capacity of more than 10 MW. It is provided by a national lobbying and advocacy group for the energy sector, representing more than 140 actors, and can therefore be considered complete. The map provides the plant's capacity as well as its type.

Completeness & specifications: From a total of 211 power plants in the overview map (see also Sankey diagram in Fig. 12), 203 plants – representing a total generation capacity of 19 GW – were included into validation. Thereof, we found 165 with a total generation capacity of 16.5 GW in our data set **[*KI14 - Completeness 87% (16.5 GW/19 GW)*]**. Among these 165 power plants, all were matching concerning the power plant type, allowing to describe their typical operation behavior **[*KI15 - Type Specifications 100% (165/165)*]**. For 127, the specification of output was correctly recorded **[*KI16 - Output Specification 78% (127/165)*]**; however, seven showed incorrect numbers, another 13 lack a specification.

A particular challenge were wind parks as each turbine is digitalized individually in OSM. Aggregating neighboring turbines, we were able to identify 18 wind parks. Another eleven power plants were available in OSM but not tagged as *plant*, instead, they were classified as *weir* (4), *generator* (3), *dam* (1), or *industrial* (3). We only found them through a manual search.

A total of 38 plants were not available at all in OSM; this includes 20 wind parks. It seems as if wind parks remain more likely to be undocumented in OSM than other types of plants; the reason might be found in the high effort necessary to document wind parks.

Eight plants from ground truth were excluded from our validations; five as they are intended for the railway system, and – given that we only exported the plants within the national borders of the investigated country – another three as they are bi-national projects with their powerhouse in a neighboring country. A detailed overview per region is provided in Table 14 in the Appendix.

Key Findings

Our inferred data set contained 81% (165/203) of the power plants from ground truth, representing 87% **[*KI14*]** of the national generation. While the type of power plants (e.g., hydro, wind) were documented correctly in all cases **[*K15*]**, output specifications lag behind **[*K16*]**. A particular challenge are wind parks. More than half of the plants that were unavailable in our data sets were such wind parks, presumably as each and every turbine has to be documented individually, posing significant effort to the volunteers. Another challenge are plants that were not tagged as *plants* in OSM, but rather as weir, generator, dam or industrial. Thus, we suggest to also include other objects that bear the word plant in the local language in their name.

8. Consumption: Communities

Data on consumption is only available in an aggregated form per region, sector, and year. For power grid modeling, however, consumption

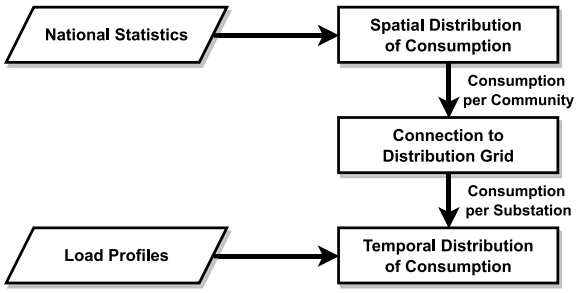
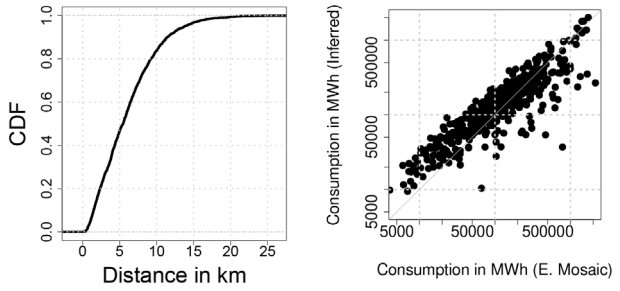


Fig. 13. Methodology. Region-wise consumption is spatially and temporally distributed, and connected with the power grid.



(a) Community-substation distance. (b) Consumption. Est. error: 105,114MWh

Fig. 14. Results. Distance between communities and closest substation, and annual consumption per substation district.

has to be distributed to the granularity of communities, and then assigned to substations, serving as hubs for consumption. In this section, we investigate whether it is feasible to spatially distribute these aggregated numbers using statistics on population and commercial activity. Beyond, we investigate whether it is feasible to connect consumption with substations representing reality well (RQ4). In particular, we describe how consumption is modeled, using national statistics and load profiles (8.1), and validated against ground truth (8.2).

8.1. Approach and results

The challenges are twofold: First, annual consumption data is only provided by national statistics[13] in an aggregated form; it must be geographically and timely spread for power grid modeling. Second, consumption must relate to a substation of the distribution power grid — the latter serving as an entry/exit point.[14] The following paragraphs describe our approach, see also Fig. 13:

1. *Spatial Distribution of Consumption:* In Austria, sector-wise power consumption is only available at the granularity of NUTS2 regions (equivalent to the nine Austrian regions). We used sector-wise gross value added (GVA) – an indicator for economic activity and available at the level of the more granular NUTS3 regions – as a proxy to distribute the aggregated consumption figures to NUTS3 regions. Then, we further distributed these consumptions to individual communities using inhabitants. The only exception was the consumption of private households, we distributed it directly by the number of inhabitants.
2. *Connection to the Distribution Grid:* For connection, we assigned the communities to the closest 110 kV substation; therefore, we determined the community centers by the address of the municipal office as found on German Wikipedia. This approach is supported by the fact that electric losses increase with line length, and operators aim to minimize losses. All consumption that is connected to the same substation, forming a substation district, are summed up.
3. *Temporal Distribution:* To be included in a grid model, annual consumption must be finally spread over time. Therefore, publicly available load profiles (Power Clearing & Settlement Austria, 2021), provided by the Austrian regulatory authority, are available and allow to calculate consumption per 15-min interval.

Fig. 14(a) provides a CDF of the distance between the communities and the assigned substation. The average distance is 6.1 km, the maximum distance 24.5 km. The average annual consumption per substation is 195,897 MWh, the maximum is 2,410,912 MWh.

[13] https://www.statistik.at/.

[14] Connection of power plants is dispensable as they are, in comparison to the more fine-granular consumption, directly supplied with a high or extra-high voltage line. These lines are available in the data sets of Section 5.

8.2. Validation

It remains infeasible to compare the final results, i.e., consumption of substation districts at individual points in time, as this data is unavailable. Load profiles are however intended for the stakeholders of the energy sector, and provided by a national authority; thus, we assume them to be trusted. Instead, we validate the input provided to the load profiles, i.e., the connection of communities with the power grid, and the consumption per substation district.

Validation of substation districts: Lacking ground truth on substation districts, we were only able to conduct plausibility checks. According to previous work, distances between consumption in communities and substations in rural areas are considered to be up to 30 km (Hülk et al., 2017). In our case, all values, see Fig. 14(a), are below this number. The average distance is 6.1 km, the maximum distance is 24.5 km **[*KI17 - Distance 6.1* km]**. For 97% of the communities the distance is even less than 15 km. From these numbers, we conclude our results on connection between consumption and substations to be plausible. Besides, derivation of substation districts are among the few aspects that had already been validated in previous work (Hülk et al., 2017), see Section 3; the authors were able to partially compare their results with a German DSO, and concluded that the assignment of substations based on physical vicinity is applicable.

Validation with Energy Mosaic Austria: For validation of each substation district's annual consumption, we relied on the Energy Mosaic Austria (Abart-Heriszt et al., 2019a) providing the annual electric consumption for each Austrian community. For each substation, we summed up the consumption of the connected communities, and compared these results with our approach. Fig. 14(b) shows the results from the Energy Mosaic, our ground truth, on the x-axis, and our results on the y-axis. Ideally, the result should be a straight line with slope 1.

Our results appear to meet reality as the average mismatch in consumption is 105,114 MWh **[*KI16 - Annual Consumption per Substation 105* GWh *(est. err.)*]**. Yet, Fig. 14(b) shows that our approach overestimates substations with low annual consumption (the data points tend to be above the gray line for low values of x), and underestimates those with high annual consumption (the data points tend to be below the gray line for high values of x). The reason therefore might be found in distributing industrial and retail consumption from NUTS3 regions to communities by inhabitants. This way, rural communities are assigned (inexistent) industrial and retail consumption, while cities are assigned less consumption than they have in reality and might be overcome by more detailed land cover data, e.g., Schultz et al. (2017) and GIScience Research Group, Institute of Geography, Heidelberg University (2021).

Table 6
Power grid modifications documented in grid development plans (GDP) vs. OSM.

Change (incl. description)	GDP	OSM
Line: Construction of a line or extension by additional systems	9	9
Line: Replacement of wires for optimization/due to aging	20	0
Cables: Construction of a cable/integration in a substation	6	0
Substation: Construction of a substation	5	5
Substation: Integration of line in a substation	8	8
Substation: Extension by additional transformers/switch gears	29	17
Substation: Replacement due to optimization/aging	18	0

Key Findings

The average distance between a community and a substation is 6.1 km **[*K17*]**, and distance is in all cases below 30 km which is considered an upper bound to limit losses as they increase with distance. This implies that substations were evenly spread among the whole country, despite some of them remain undocumented in OSM (see discussion in Section 6). The average mismatch in consumption is 105 GWh **[*K18*]**, a rough equivalent of 13 wind turbines in generation and thus comparable to the quality of the documentation of power plants in OSM. Thereby, consumption per substation is typically overestimated in rural areas, and underestimated in urban areas, and a consequence of distributing consumption by inhabitants. This mismatch appears to have limited impact on the (simulation of the) transmission grid as the consumption is only attributed to a close-by substation within the region. Yet, this mismatch could be decreased by using more detailed land use data, indicating the location of industrial and retail facilities. However, current land use data in OSM is too coarse-grained to do so; but more sophisticated approaches might yield better results.

9. Power grid modifications

The power grid undergoes continuous, though slow modifications that need to be incorporated into an up-to-date model reflecting reality well. By now, we have revealed that nowadays Austrian power grid is well represented in OSM; in this section, we investigate whether modifications of power grid artifacts like lines or substations are documented timely in OSM (RQ5). The gained insights allow to determine whether OSM-based model data can be kept up to date. Therefore, we rely on grid development plans (GDPs). They are annual reports describing the TSO's plans for grid extension in a coarse-grained way. The data is, however, not directly sufficient for documentation in OSM as certain details, e.g., precise geographic coordinates or the number of wires/cables, are lacking. From the plans of 2012–2021 (Austrian Power Grid, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021), we filtered the projects that have been realized before 2022, and analyzed the respective OSM objects' history to see whether (and when) the described changes are reflected in the OSM database. In total, we found 45 projects with a total of 95 changes affecting power engineering infrastructure, and classified them into seven categories, see Table 6.

We found that all cases of newly constructed power lines (9), newly built substations (5), and the integration of existing lines into substations (8) have been documented by volunteers in OSM **[*KI19 - Modifications 100% (21/21)*]**; 76% (16/21) of the modifications have already been documented by the time of the estimated finalization date; the others were documented 13 months to 6 years afterwards **[*KI20 - Update before Finalization 76% (16/21)*]**. Fig. 16 in the Appendix provides a detailed timetable presenting project realization and related changes in OSM. In contrast, modifications related to the replacement of wires (20) or substation equipment (18), as well as any changes affecting underground cables (6) remain opaque for OSM. With regard to the extension of substations by additional transformers or switchgears, we expected the substations' areas to increase in OSM. The results are, however, mixed. We found such changes that appear to be related to an extension; others appear unrelated – they had either been done multiple years after the extensions and/or affect only small areas – or do not change at all, i.e., changes in a substation's size can only be considered as an indication for the substation's extension.

Key Findings

Modifications of aerial lines and substations appear to be documented by volunteers in OSM **[*KI19*]**, typically no later than project finalization **[*KI20*]**, but replacement of existing equipment remains opaque to OSM. Also, all modifications related to underground cables remain unnoticed by OSM, suggesting, again, better documentation of extra-high than high voltage levels in Austria. For extra-high voltage, cables are practically inexistent for technical reasons; in Austria, there are only two 380 kV cables of 4.5 km resp. 5.5 km length, but high voltage cables represent a significant part of the distribution grid in urban areas.

10. Adoption to other countries

In this section, we expand our work beyond the borders of Austria, and investigate whether our approach appears also worthwhile for modeling other countries' power grid (RQ6). In a first step, we investigate the other European countries[15] that are, like Austria, connected to the Synchronous Grid of Continental Europe. In a second step, we discuss other countries, including those outside of Europe.

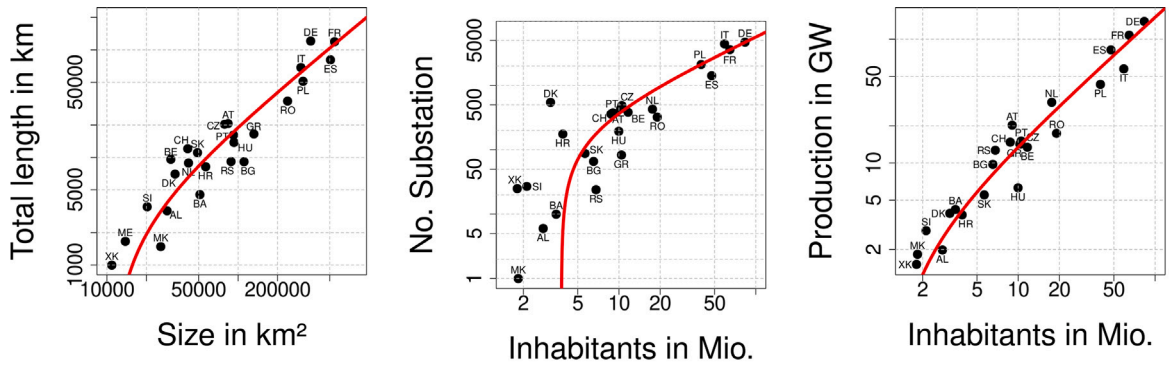
Power lines and electric systems: Assuming country-wide supply, we expect the total line length to be roughly linear with a country's size. This relation has already turned out to be true for the nine regions of Austria, see Fig. 6. In order to check this assumption for the other European countries, we exported their lines from OSM and calculated their individual length in QGIS. Then, we calculated the overall system length by multiplying the individual lengths with the amount of parallel electric systems as indicated by OSM's *cable* tag, summing up the respective results for all lines with a voltage level of 50 kV or more.

Fig. 15(a) shows the calculated length per country (y-axis) with respect to the country's size (x-axis),[16] including a linear regression. R^2 is lower (0.870) than for Austria's regions (0.897), see also Fig. 6; however, it is still high considering the different legacies of the national power grids (e.g., coming apparent in the different voltage levels used in distribution), a difference that is not prevalent among the nine Austrian regions **[*KI21 - Electric Systems* $R^2 = 0.870$]**.

For calculation of electric parameters, it is necessary to know the respective wire types. These wire types are standardized in EN 50182 (European Standards, 2013) which is via European Union legislationin effect in all but four of the investigated countries, allowing to infer the necessary parameters for calculation of resistance and reactance. The four exceptions are either candidate countries (Albania, Bosnia, and Herzegovina, Montenegro) or potential candidate countries (Kosovo) of the European Union, i.e., overall, the standard will also be obligatory there in the future. The national grid operators – assured to a good future – might keep to these standards already today.

[15] Albania, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, France, Germany, Greece, Hungary, Italy, Kosovo, Luxemburg, Montenegro, Netherlands, Northern Macedonia, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Switzerland. Moldavia and Ukraine only joined after Russia's full-scale invasion in 2022 and are thus not included.

[16] Denmark is only partially connected to the Synchronous Grid of Continental Europe; thus, we used the size and inhabitants of the supplied area instead of the country's total size.



(a) Total system length wrt country size. $R^2 = 0.870$ (b) Substations wrt inhabitants. $R^2 = 0.916$ (c) Power plant production wrt inhabitants. $R^2 = 0.938$

Fig. 15. Total system length, number of substations and electric production (y-axis) for 24 European countries are largely defined be the country's size resp. inhabitants (x-axes). The data and regression curve is presented in logarithmic scale; otherwise, the many countries with sizes between 10,000 km² and 100,000 km² would overlap.

Topology - Substations: We expect the number of a country's total substations to be roughly linear with a country's inhabitants. As more energy must be provided to its inhabitants, more substations are needed, especially as more substations also reduce distances and, consequently, also transport losses. We exported all substations from OSM, that have a line with a voltage of more than 50 kV within a radius of 1000 m. Fig. 15(b) shows the number of substations (y-axis) with respect to the countries' inhabitants (x-axis) including a linear regression. R^2 of the regression is high (0.916), suggesting that substations in other countries are covered to a similar extent as in Austria **[*KI22 - Substations* $R^2 = 0.916$.]**

Production - Power plants: As a first approximation, we expect a country's total production to be roughly linear with a country's inhabitants as lifestyle among these European countries is similar. Consequently, we exported all countries' power plants from OSM and calculated the national production, i.e., the sum of all power plants' output. Fig. 15(c) shows a country's production on its *y*-axis and their inhabitants on the x-axis, including a linear regression. R^2 is again high (0.938), suggesting that generation in other countries are recorded in a similar extent as in Austria **[*KI23 - Production* $R^2 = 0.938$]**.

Consumption - Communities: For modeling consumption, we distributed aggregated consumption spatially and temporally relying on national statistics and load profiles. Thus, we investigated the availability of this data at *Eurostat*, responsible for provision of statistical information and harmonization of methods among EU member states and candidates, and found the following data sets: (I) total electric consumption on a national level (NUTS1) in its simplified energy balances (eurostat, 2022c), (II) GVA per NUTS3 region and sector (eurostat, 2022a), and (III) the number of inhabitants on community level (eurostat, 2022b).

This deviates from Austria in certain aspects, and might have consequences for the resulting data's quality: (I) In Austria, electric consumption is available at the finer granular NUTS2 level instead of the national level. For small countries (e.g., Albania, Luxemburg, Kosovo, Northern Macedonia, Slovenia) with size and population comparable to Austrian regions, the impact on modeling appears to be negligible. For medium and particularly large countries, spatial distribution of consumption by sector-wise GVA might introduce deviation from reality. A more sophisticated approach, e.g., like proposed by Priesmann et al. (2021) might overcome this drawback; however, requires more effort. (II) Load profiles, that are provided by official authorities in Austria, are lacking for many countries, challenging the temporal distribution of consumption. As a work around, existing load profiles from Germany or Austria might be adapted for these countries (e.g., depending on climate, time zone, etc.). This is insofar a practicable way as today's load profiles for Austria are also a derivation of German load profiles. Nonetheless, it requires more effort and might introduce additional deviations from reality.

Beyond Europe: Outside of Europe, OSM documents power grid features in countries such as China, India, Japan, Russia, and the United States. However, data quantity varies; for instance, Russia exhibits fewer lines per area – the cause of which remains uncertain – whether reflecting actual infrastructure deployment or OSM coverage limitations. Conducting plausibility checks, as done for European countries (see regressions on system length, substations, and generation in Figs. 15(a), 15(b), and 15(c)), is challenging due to diverse living styles, utility supply, and operational grid rules. Language barriers hinder access to national statistics for most non-European countries, making it difficult to assess data adequacy and collection methods. While there is potential to model these countries, uncertainties persist, necessitating a comprehensive validation process with local expertise and language skills for each region.

Key Findings

Investigating 24 European countries, a country's total length of electric systems, the number of substations, and total electric production – as inferred from OSM – is highly determined by its size resp. inhabitants **[*KI21,KI22,KI23*]**, and implies that all these countries are documented in OSM at a similar level as Austria, the latter serving as our example in the previous sections. This suggests that our approach is worthwhile for adoption to other countries. A challenge might be the availability of detailed statistical data on consumption or load profiles, but both might be overcome with more sophisticated modeling of publicly accessible data. OSM also documents power grid artifacts for countries beyond Europe, i.e., China, India, Japan, Russia, and the United States. Due to the different legacy, investigations on the quality of documentation remains subject for future work.

11. Discussion

Data quality from open source intelligence: Table 7 summarizes our key findings, and emphasizes high congruence between the inferred data and reality. We did not only find (almost) all power grid artifacts, and estimated the relevant parameters with high accuracy, but also found consistently more artifacts in OSM than in the ground truths. These artifacts were eventually confirmed to be existent. This fact emphasizes

Table 7
Key findings highlighting inferred power grid artifacts, ground truth in brackets and key figures.

Extra-high voltage systems	
KI1. Completeness (TSO network data): All systems are available in OSM, emphasizing OSM's comprehensiveness.	100%
KI2. Voltage Specifications (TSO network data): All systems are assigned the correct voltage level in OSM, emphasizing OSM's accuracy.	100%
KI3. Coverage (TSO network data): OSM reveals an additional four (really existent) systems, emphasizing OSM's up-to-dateness and independence from ground truth.	+4%
KI4. System Length (TSO network data): The estimation error is 1.4% of the mean, i.e., OSM allows derivation of lengths, necessary for electric parameter calculation, with high accuracy.	0.76 km
KI5. Resistance R & Reactance X (TSO network data): Their estimation errors are lower than their change that is associated with a (common) 15 °C temperature shift.	0.17 Ω 0.76 Ω
High voltage lines	
KI6. Completeness (Land use plans): All lines are available in OSM, emphasizing OSM's comprehensiveness, also at the high voltage level.	100%
KI7. Voltage Specification (Land use plans): 97% (558/573) of the voltage specifications in OSM are consistent, indicating slightly lower accuracy for high than extra-high voltage lines.	97%
KI8. Coverage (Land use plans) OSM reveals an additional 29 lines of which 19 are confirmed by Google Street View, again emphasizing OSM's up-to-dateness and independence from ground truth.	+3%
KI9. OSM Parameters (Google Street View): For 98% (1355/1385) of the lines, the parameters *cables* and *wires* are correctly recorded, i.e., they guarantee correct splitting into systems and parameter calculation.	98%
Extra-high voltage substations	
KI10. Completeness (TSO overview map): With a single exception, all 53 substations are recorded in OSM, again emphasizing OSM's comprehensiveness.	98%
KI11. Coverage (TSO overview map): OSM reveals an additional substation, highlighting the OSM user's swiftness in digitalizing new objects.	+2%
High voltage substations	
KI12. Completeness (DSO overview maps): 131 out of 149 substations are found in OSM, indicating lower comprehensiveness than for the extra-high voltage level.	87%
KI13. Coverage (DSO overview maps): OSM reveals an additional seven (really existent) substations, i.e., our approach does not classify inadequate objects as substations (false positives).	+6%
Generation	
KI14. Completeness (Plant overview maps): 16.5 GW of Austria's generation capacity (19 GW) is found in OSM, again showing OSM's comprehensiveness.	87%
KI15. Type specifications (Plant overview maps): All power plants are assigned the correct type (e.g. hydro, wind, etc.), allowing to infer their typical production behavior over time.	100%
KI16. Output Specification (Plant overview maps): 127 of 165 power plants provide consistent specification of their output.	78%
Consumption	
KI17. Distance: The average distance between communities and substations is 6.1 km, implying an evenly spread over the country and nation-wide supply.	6.1 km
KI18. Annual Consumption per Substation (Energy Mosaic): The estimation error is the rough equivalent to 13 wind turbines in generation, i.e., mismatch in consumption is comparable to the mismatch in generation.	105 GhW
Power grid modifications	
KI19. Line and Substations Modifications: The construction of new lines or substations as well as the integration of existing lines into substations has always been documented in OSM.	100%
KI20. OSM Update before Finalization: 76% (16/21) of the power grid modifications are already documented in OSM before their finalization in the real world; the others take up to six years.	76%
Adaption to European countries	
KI21. Electric Systems: The documentation's extent of electric systems in OSM is highly dependent ($R^2 = 0.870$) on a country's size.	0.870
KI22. Substations: The number of substations in OSM is highly dependent ($R^2 = 0.916$) on a country's inhabitants.	0.916
KI23. Generation: Total generation in OSM is highly dependent ($R^2 = 0.938$) on a country's inhabitants.	0.938

OSM's up-to-dateness; some of these lines/substations were just brought online. Beyond, it supports OSM's independence from ground truth data. In general, the transmission grid operating at extra-high voltage is better covered than the distribution grid.

Detailed analysis of cyber and terroristic attacks: In manipulation-of-demand (MAD) attacks, an adversary increases power consumption of a high number of devices under its control (e.g., via a botnet). This sudden increase in load puts a strain on the power grid, either by outplaying reserve capacities (Dabrowski et al., 2017) or overwhelming individual lines or cascades of lines (Soltan et al., 2018), and eventually causes (partial) blackouts. In the past, a lack of power grid models representing real-world power grids has prevented independent researchers to investigate the impact of such or other attacks on actual countries. This is, however, important to understand which regions are particularly vulnerable to such attacks, or whether such attacks are more likely in certain load scenarios (e.g., in summer, in the night) than others (e.g., in winter, during the day).

Our work paves the way towards such analyses: We developed a method to infer power grid model data from publicly accessible sources. By the example of Austria, a medium-sized country in the European Union, we have shown that the inferred data reflects the actual Austrian power grid with high accuracy. This data could now be processed to serve as an input for power flow analysis, e.g., using a tool such as MATPOWER. This way, it is possible to infer the operational state of the Austrian power grid, in presence and absence of attacks, and to investigate the impact of such attacks on resilient operation. Soltan et al. (2018) have already shown, though by the example of an imaginary power grid and a limited number of scenarios (5) for Poland, that power flow analysis is an adequate tool to investigate the impact of MAD attacks. In comparison to these previous analyses for Poland, our data allows to investigate each 15-min block of a year. Summarizing, our work enables detailed analysis to investigate MAD as well as other cyber or terroristic attacks for real-world power grids.

Our OSINT method as an attack: Our work does not only enable the analysis of attacks in real-world power grids, both for independent researchers and adversaries, but also represents an attack itself. The energy sector refrains from providing its models, resp. model data, to the public for reasons of security. Our results show, however, that the power sector's security-by-obscurity strategy is inadequate. Anybody is able to obtain sufficient knowledge on the power grid from public sources to cause significant harm. Concluding, our results show that the energy sector's asset, namely the power grid data, is accessible by unauthorized parties and the protection mechanism (secrecy) does not reach the intended goal. In security, this is the very specification of an attack.

Adoption to other countries: Unlike computer science, power grid engineering is more dependent on location. Yet, we see high potential to also model the power grid of the other 24 European countries successfully, that are, like Austria, connected with the Synchronous Power Grid of Continental Europe, by far the largest grid on the planet by total load. In OSM, their lines, substations, and generation are documented to a similar extent as in Austria; beyond, Eurostat provides statistics for all these countries to infer consumption. The only drawback is that numbers on consumption are only available at the national level, and the lack of load profiles for many countries. The first might result in inaccurate spatial distribution of load, and might mandate more sophisticated approaches from literature (Priesmann et al., 2021). The latter hinders temporal distribution, but could be overcome by adapting load profiles from other countries, e.g., Germany. This is insofar legitimate as also the official Austrian profiles are derivated from the German ones.

For power grids outside of Europe, we were only able to check whether power grid data is principally available in OSM, and, indeed, we found such data in China, India, Japan, Russia, and the United States. Plausibility checks, as done for the other European countries, are, however, infeasible as living style, supply with utilities, and the grid's operational rules differ significantly from the European ones. For validation, we would thus have to repeat the entire process for each country, requiring local knowledge and language skills.

Automation: Using the geographic information system QGIS in combination with python scrips, our approach is automatable to a large extent. This includes data export and processing of lines, substations, power plants, and the distribution of consumption from Eurostat. The changes that are necessary among European countries are limited: First, the respective voltage levels must be identified as they slightly differ among countries. This remains a manual task, but this recherche is easily done by accessing the ENTSO-E map, web pages of national/regional providers, or Wikipedia. Second, wire types used in this country must be identified. This will remain a manual task, but, like the search for voltage levels, it must be done only once per country. Additionally, the respective standard (European Standards, 2013) provides country-specific wire types, and their unique labels foster an easy (Internet) search. Third, the data set on power plants might have to be manually improved. The effort should be moderate as plants are highly covered by media and are usually featured in the operators' Internet presence. From our experience, we estimate that adequate model data can be inferred within three weeks for a European country, assuming command of the local language and excluding data validation against ground truth.

Countermeasures and transparency: The only countermeasure on behalf of the power grid would be underground deployment, as shown by our experiments including grid development plans. This is obviously infeasible for (all) power plants due to their size. For power lines, partial underground deployment is common, particularly in urban regions and for lower voltage levels; however, complete underground deployment would multiply costs of construction and maintenance, and, most notably, make the power grid more fragile due to resonance effects. Thus, we advise the energy sector to adapt its strategy and to include this threat into national security strategies.

At the same time, the availability of this data increases transparency and contributes towards more fairness in the energy market. Because of energy transition, new stakeholders – typically, less powerful, and having less information about the grid infrastructure than the established companies – are entering the energy markets, potentially causing competitive imbalance. In the past, small engineering companies complained about lacking transparency (Szelgrad, 2014). Their projected plants were denied access to the power grid, while technically equivalent projects at the same location were granted access when planned by a subsidiary of the regional grid operator. According to regulation, the grid operator must guarantee fair access and can only deny it in case of technical or operational necessities; yet, the provider's decision remains a black box for the engineering companies. Our work makes it possible to retrace such decisions, potentially even serving in a legal process, and this way contributes towards more fairness in the energy market.

Updates and flexibility: Our results do not only confirm that society extensively documents knowledge on the power grid in digital form, but our approach also bears multiple advantages for models that built upon our data: (I) Sooner or later, modifications of the power grid are documented in OSM and thus, automatically update the model data. (II) Beyond, it allows more flexibility for the simulated scenarios: First, it is feasible to investigate the full year at the granularity of 15-min intervals. Second, it allows to modify these parameters to reflect other, (yet) hypothetical scenarios. For example, one might assume a (cyber or terroristic) attack against lines in a certain region. As we know the location of the lines from the OSM data set, we can easily remove them from the analysis and see how the power grid behaves without them. Alternatively, one could also investigate scenarios of energy transition. As we know the type of the current plants of Austria, we could remove all coal and gas plants, insert wind turbines and photovoltaic instead, and again, investigate their impact on power grid operation.

Limitations: First, underground cables and indoor substations remain hidden for the volunteers contributing to OSM. This barely affects the extra-high voltage level as underground/indoor deployments are rare exceptions, e.g., in highly densely populated areas or near airports. In Austria, there are only two extra-high voltage underground cables of 4.5 resp. 5.5 km. Both are well covered by media and could be added manually. In comparison, underground cables and indoor substations at the high voltage level are prevalent in urban areas. Second, wind parks are not fully covered, potentially as each and every wind turbine has to be documented individually in OSM. We expect that this becomes even more challenging with the advent of renewable energy sources (wind turbines, photovoltaics, etc.). They are typically smaller and more decentralized than their fossil counterparts, and require more effort for documentation in OSM. In the future, it might thus become necessary to model generation – at least partially – in a similar way as consumption, i.e., by statistical modeling. Finally, adoption to other European countries appears plausible, though consumption models might be coarse-grained, especially for large countries like Germany, France, or Poland, as aggregated numbers of power consumption are only available at the granularity of countries. Adaption to non-European countries, e.g., China, India, Japan, or Russia, remain an open question and are left to future work, especially as specific language skills and knowledge on national regulation is necessary. Plausibility checks as done for the European countries are inadequate for these countries due to the diverse living styles, utility supply, and operational grid rules.

Future work: It is possible to extend our current work to conduct even more sophisticated analyses than power flow analysis in the future. For example, optimal power-flow analysis (OSP) aims to find economically optimal states for power generation, assuming certain consumption in a power grid. Therefore, generation must be extended by a cost

function defining the costs of operating certain power plants. This data might be inferred from investment calculations, or portals for market transparency. Typically, costs of generation are heavily dependent on the power plant type, an information that is readily available in our data set. Then, also security-constrained optimal power flow (SCOPF) would become feasible. In addition to OSP, it requires that additional contingency constraints are fulfilled. For example, European operators follow the (n-1) criteria, stating that the power grid must sustain even in case a line fails. Limits on current carrying capacities are dependent on the wire type and can be inferred from the respective standard. These methods investigate the power grid in steady state. For the dynamic response in presence of disturbances, transient stability analysis is needed. This is, however, far beyond the scope of this work. First, open-source tools like MATPOWER are not able to perform such an analysis, one must use a fee-based tool instead. Second, it requires more detailed parameters like short circuit parameters of substations or inertia constants of generators.

12. Conclusion

In this paper, we investigated whether an Open Source Intelligence (OSINT)-based approach is capable to generate model data of real-world power grids. By the example of Austria, we inferred the grid's relevant parameters from OpenStreetMap (OSM) and national statistics, and extensively validated them against ground truths, namely Google Street View, governmental land use plans, and the power sector's aggregated information material. The gained model data reflects reality well, and its accuracy is certainly sufficient to generate power grid models identifying bottlenecks of the current system (e.g., to target cyber or terroristic attacks), understanding the impact of manipulation-of-demand (MAD) attacks on individual countries or its regions (e.g., to prepare for such scenarios), or investigating emerging scenarios like the increased deployment of photovoltaic cells (e.g., to develop public strategies). To summarize, our approach is currently the best way to obtain such model data if insider knowledge as held by grid operators is not available. Beyond, it can be mostly automated facilitating the generation of relevant model data in approx. three weeks, motivating its application to other countries. At a societal level, our approach fosters research and discussion that is independent of the power grid operators; among others, it allows double-checking decisions on grid connection, supporting market fairness. Beyond that, our method represents an attack itself and challenges the energy sector's security-by-obscurity approach. If we – researchers with limited resources – are able to infer such model data, an adversary can do so as well, thus becoming a threat to national security. We recommend including these aspects into national security concepts.

CRediT authorship contribution statement

Anja Klauzer: Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Markus Maier:** Writing – review & editing, Validation. **Lore Abart-Heriszt:** Validation. **Johanna Ullrich:** Funding acquisition, Conceptualization, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Johanna Ullrich reports financial support was provided by Christian Doppler Research Association. Anja Klauzer reports financial support was provided by Austrian Research Promotion Agency. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The paper contains a link to a github repository.

Acknowledgments

This material is based upon work partially supported by (1) the Christian-Doppler-Laboratory for Security and Quality Improvement in the Production System Lifecycle; the financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged; (2) SBA Research (SBA-K1), a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the province of Vienna. The COMET Programme is managed by FFG; (3) Project 877110 2big2fail and F0999895389 NurZu! funded by the Program "BRIDGE 1" (FFG); (4) Project DynAISEC FO999887504 funded by the Program "ICT of the Future" – an initiative of the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology;

Appendix A. Land use plans of the nine regions

- *Burgenland*: https://gis.bgld.gv.at/WebGIS/synserver
- *Carinthia*: https://gis.ktn.gv.at/webgisviewer/atlas-mobile/map/Raumordnung/Raumordnung
- *Lower Austria*: https://atlas.noe.gv.at/atlas/portal/noe-atlas
- *Upper Austria*: https://wo.doris.at/weboffice/synserver?project=weboffice&client=core&user=guest&basemapview=or_flaewi
- *Salzburg*: https://www.salzburg.gv.at/sagismobile/sagisonline
- *Styria*: https://gis.stmk.gv.at/wgportal/atlasmobile
- *Tyrol*: https://maps.tirol.gv.at/synserver?user=guest&project=tmap_master
- *Vorarlberg*: http://vogis.cnv.at/atlas/init.aspx?karte=planung_und_kataster
- *Vienna*: https://www.wien.gv.at/flaechenwidmung/public/

Appendix B. Detailed evaluation results per region

See Fig. 16 and Tables 8–14.

Table 8
Line objects inferred from OSM.

		Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
	Total	61	587	474	261	104	201	216	89	55	2048
assoc. w.	380 kV	14	12	85	13	26	17	24	3	15	209
	220 kV	4	85	70	30	21	7	65	30	11	323
	110 kV	45	531	324	218	63	183	134	67	30	1595

Table 9
Total electric system length in km.

	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
Total	1011	2059	5404	2894	1666	3131	2741	1125	616	20,646
380 kV	253	145	1182	435	355	479	508	227	89	3,672
220 kV	190	624	1157	700	569	714	1153	465	224	5,796
110 kV	568	1290	3065	1759	742	1938	1080	433	303	11,178

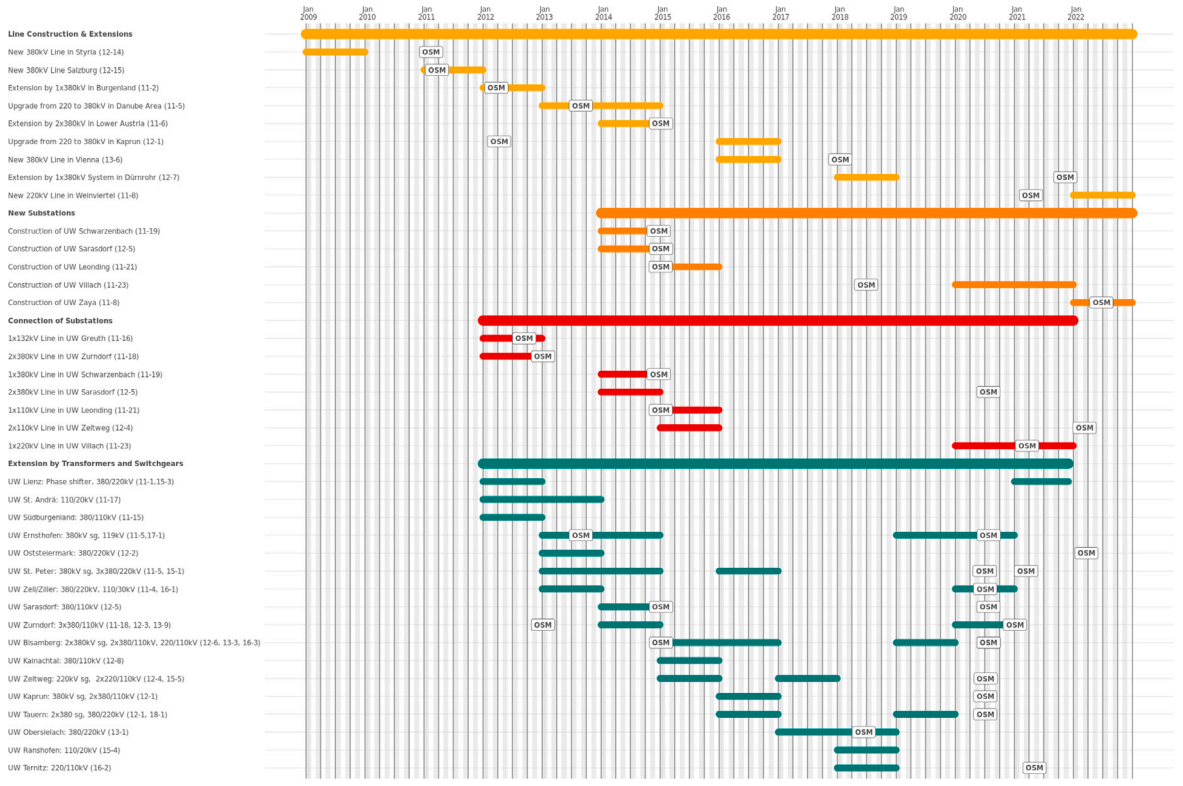


Fig. 16. Project Completion and OSM Updates: The bars reflect the project's completion as provided in the Grid Development Plans, provided by the Austrian TSO APG, the tags refer to changes in the OSM database by volunteers.

Table 10
Validation of 110 kV lines with land use plans.

		Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
Total		45	531	324	218	63	183	134	67	30	1595
Validated	Geographically correct (incl. voltage)	45 (–)	361 (350)	0 (–)	154 (152)	58 (56)	177 (–)	101 (–)	67 (–)	0 (–)	963 (–)
	Not found	0	9	0	7	5	6	2	0	0	29
Not val	Minor artifacts	0	161	0	57	0	0	31	0	0	249
	Lacking land use plans	0	0	324	0	0	0	0	0	30	354

Table 11
Land use plans. Feasibility of validation steps.

	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Coverage	✓	✓	✗	✓	✓	✓	✓	✓	✗
Completeness	✗	✓	✗	✓	✓	✓	✓	✓	✗
Voltage Specific.	✗	✓	✗	✓	✓	✗	✗	✗	✗

Table 12
Validation of 110 kV lines with Google Street View.

		Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
Total		61	587	474	261	104	201	216	89	55	2048
Validated	Correct cables & lines	59	183	345	220	92	176	139	86	55	1355
	Incorrect cables	0	1	14	1	0	4	0	1	0	21
	Incorrect wires	1	0	3	4	0	1	0	0	0	9
Not val	Minor artifacts	0	395	86	10	4	11	71	0	0	577
	No visibility in GSV	1	8	25	27	8	9	6	2	0	86

Table 13
Validation of substations with DSO overview maps/aggregated numbers.

	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
Total (OSM)	17	46	88	60	20	60	46	26	13	376
OSM only	0	–	4	–	–	–	–	3	–	–
OSM ∩ DSO	17	–	84	–	–	–	–	23	–	–
DSO only	1	–	14	–	–	–	–	3	–	–
Total (DSO)	18	50	98	86	–	63	47	26	46	–

Table 14
Validation of power plants with overview map.

		Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	Austria
Total (Overview Map)		15	25	34	34	29	28	23	18	5	211
Validated	Found in OSM	8	25	22	31	18	24	18	16	3	165
	(Corr. Out.)	0	22	10	24	14	22	17	16	2	127
	(Inc. Out.)	0	3	1	0	2	0	1	0	0	7
	(Lack. Out.)	0	0	1	7	2	2	0	0	1	13
	(Wind p.)	8	0	10	0	0	0	0	0	0	18
	Not in OSM	7	0	12	1	9	4	3	0	2	38
Not val	Railway	0	0	0	0	2	0	1	2	0	5
	Exterritorial	0	0	0	2	0	0	1	0	0	3

References

Abart, A., Köpplmayr, H., Dobetsberger, G., Gahleitner, B., Wolkerstorfer, K., Leitner, W., Niederhuemer, W., 2021. Technischer Bericht: Netzverträglichkeit von Erdkabeln im gelöscht betriebenen 110-kV-Netz in Oberösterreich. https://www.land-oberoesterreich.gv.at/Mediendateien/Formulare/Dokumente%20UWD%20Abt_US/EN_110kV_NetzausbauKabelreserveNetzOOE_20181124.pdf. (Accessed 17 October 2021).

Abart-Heriszt, L., Erker, S., Reichel, S., Schöndorfer, H., Weinke, E., S., L., 2019a. Energiemosaik Austria. Österreichweite Visualisierung von Energieverbrauch und Treibhausgasemissionen auf Gemeindeebene. https://www.energiemosaik.at. (Accessed 04 October 2021).

Abart-Heriszt, L., Erker, S., Stoeglehner, G., 2019b. The energy mosaic Austria—A nationwide energy and greenhouse gas inventory on municipal level as action field of integrated spatial and energy planning. Energies 12 (16), http://dx.doi.org/10.3390/en12163065, URL https://www.mdpi.com/1996-1073/12/16/3065.

Ämter der Steiermärkischen Landesregierung und der Burgenländischen Landesregierung, 2004. Umweltverträglichkeitsgutachten 380kV-Leitung Südburgenland - Kainachtal. URL https://www.verwaltung.steiermark.at/cms/dokumente/11682278_74834965/d7b3b1c2/UVPGutachten.pdf.

Austrian Power Grid, 2012. Netzentwicklungsplan 2012 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2013–2022.

Austrian Power Grid, 2013. Netzentwicklungsplan 2013 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2014–2023.

Austrian Power Grid, 2014. Netzentwicklungsplan 2014 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2015–2024.

Austrian Power Grid, 2015. Netzentwicklungsplan 2015 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2016–2025.

Austrian Power Grid, 2016. Netzentwicklungsplan 2016 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2017–2026.

Austrian Power Grid, 2017. Netzentwicklungsplan 2017 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2018–2027.

Austrian Power Grid, 2018. Netzentwicklungsplan 2018 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2019–2028.

Austrian Power Grid, 2019. Netzentwicklungsplan 2019 für das Übertragungsnetz der Austrian power grid AG (APG), planungszeitraum 2020–2029.

Austrian Power Grid, 2020. Netzentwicklungsplan 2020 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2021–2030.

Austrian Power Grid, 2021. Netzentwicklungsplan 2021 für das Übertragungsnetz der Austrian power grid AG (APG), Planungszeitraum 2022–2031.

Austrian Power Grid, 2021a. Locations and connections of APG substations. https://www.apg.at/api/sitecore/projectmedia/download?id=27e04cae-d929-4e56-b9af-992e25a1dea4. (Accessed 01 October 2021).

Austrian Power Grid, 2021c. Published power grid line specifications. https://www.apg.at/api/sitecore/projectmedia/download?id=703efbb9-bd69-49db-b2f6-bb676cac466b. (Accessed 01 October 2021).

Bruny, C., Jandrasits, A., Salomon, G., Messner, K., Kügler, I., 2016. Umweltverträglichkeitserklärung Ersatzneubau APG-Weinviertelleitung - Vorhabensbeschreibung. URL https://docplayer.org/53755568-Umweltvertraeglichkeitserklaerung-ersatzneubau-apg-weinviertelleitung.html.

Dabrowski, A., Ullrich, J., Weippl, E., 2017. Grid shock: Coordinated load-changing attacks on power grids. In: Annual Computer Security Applications Conference. ACSAC, pp. 303–314.

Egerer, J., 2016. Open Source Electricity Model for Germany (ELMOD-DE). Tech. Rep., DIW Data Documentation.

Egerer, J., Gerbaulet, C., Ihlenburg, R., Kunz, F., Reinhard, B., von Hirschhausen, C., Weber, A., Weibezahn, J., 2014. Electricity Sector Data for Policy-Relevant Modeling: Data Documentation and Applications to the German and European Electricity Markets. Tech. Rep., DIW Data Documentation.

European Standards, 2013. BS EN 50182:2001 - Conductors for overhead lines. Round wire concentric lay stranded conductors.

eurostat, 2022a. Gross value added at basic prices by NUTS 3 regions. https://ec.europa.eu/eurostat/databrowser/view/NAMA_10R_3GVA_custom_6680196. (Accessed 12 July 2022).

eurostat, 2022b. Local administrative units (LAU). https://ec.europa.eu/eurostat/c/portal/layout?p_l_id=345247&p_v_l_s_g_id=0. (Accessed 12 July 2022).

eurostat, 2022c. Simplified energy balances. https://ec.europa.eu/eurostat/databrowser/view/NRG_BAL_S_custom_6679924. (Accessedn 12 July 2022).

Evangelista, J.R.G., Sassi, R.J., Romero, M., Napolitano, D., 2021. Systematic literature review to investigate the application of open source intelligence (OSINT) with artificial intelligence. J. Appl. Secur. Res. 16 (3), 345–369. http://dx.doi.org/10.1080/19361610.2020.1761737, arXiv:https://doi.org/10.1080/19361610.2020.1761737.

Fichtinger, K., Kostner, M., 2013. Umweltverträglichkeitserklärung 380-kV Salzburgleitung Netzknoten St. Peter - Netzknoten tauern. URL https://www.yumpu.com/de/document/read/25968762/tb-380kv-ltg-nk-st-peter-nk-tauern-bereinigt-30012013pdf.

GIScience Research Group, Institute of Geography, Heidelberg University, 2021. OSM landuse landcover. https://osmlanduse.org/#12/8.7/49.4/0/. (Accessed 15 October 2021).

Glassman, M., Kang, M.J., 2012. Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). Comput. Hum. Behav. 28 (2), 673–682. http://dx.doi.org/10.1016/j.chb.2011.11.014, URL https://www.sciencedirect.com/science/article/pii/S0747563211002585.

Heitkoetter, W., Medjroubi, W., Vogt, T., Agert, C., 2019. Comparison of open source power grid models—Combining a mathematical, visual and electrical analysis in an open source tool. Energies 12 (24), 4728.

Hoersch, J., Hofmann, F., Schlachtberger, D., Brown, T., 2018. PyPSA-Eur: An open optimisation model of the European transmission system. Energy Strategy Rev. 22, 207–215. http://dx.doi.org/10.1016/j.esr.2018.08.012, arXiv:1806.01613.

Huang, B., Cardenas, A., Baldick, R., 2019. Not everything is dark and gloomy: Power grid protections against IoT demand attacks. In: 28th USENIX Security Symposium. USENIX Security 19, USENIX Association, Santa Clara, CA, pp. 1115–1132, URL https://www.usenix.org/conference/usenixsecurity19/presentation/huang.

van Huis, P., 2018/19. A birdie is flying towards you - Idenitfying the separatists linked to the downing of MH17. a bellingcat investigation, URL https://www.bellingcat.com/app/uploads/2019/06/a-birdie-is-flying-towards-you.pdf.

Hülk, L., Wienholt, L., Cußmann, I., Müller, U., Matke, C., Kötter, E., 2017. Allocation of annual electricity consumption and power generation capacities across multiple voltage levels in a high spatial resolution. Int. J. Sustain. Energy Plan. Manag. 13, 79–92.

Kärnten Netz, 2021. Der Weg des Stroms. https://kaerntennetz.at/uebersicht-verteilernetz-6054.htm. (Accessed 17 October 2021).

Keliris, A., Konstantinou, C., Sazos, M., Maniatakos, M., 2019. Open source intelligence for energy sector cyberattacks. In: Gritzalis, D., Theocharidou, M., Stergiopoulos, G. (Eds.), Critical Infrastructure Security and Resilience: Theories, Methods, Tools and Technologies. Springer International Publishing, Cham, pp. 261–281, URL https://doi.org/10.1007/978-3-030-00024-0_14.

Kiessling, F., Nefzger, P., Nolasco, J., Kaintzyk, U., 2003. Overhead Power Lines: Planning, Design, Construction, vol. 759, Springer.

Kirschner, M., Lugschitz, H., Panosch, W., 2007. Freileitungen–Vorschriftenlage, Leitungsplanung, technische Neuerungen. E & I Elektrotech. Inf. 124 (3), 40–50.

Land Steiermark - A13 Umwelt und Raumordnung, 2021. 380 kV Freileitung. https://www.verwaltung.steiermark.at/cms/beitrag/11682278/74834965/. (Accessed 19 September 2021).

Lee, R., Assante, M., Conway, T., 2016. Analysis of the cyber attack on the Ukrainian power grid. In: Electricity Information Sharing and Analysis Center. E-ISAC, Vol. 388.

Matke, C., Medjroubi, W., Kleinhans, D., 2016. SciGRID - An open source reference model for the European transmission network (v0.2). URL https://scigrid.de.

Medjroubi, W., Müller, U., Scharf, M., Matke, C., Kleinhans, D., 2017. Open data in power grid modelling: new approaches towards transparent grid models. Energy Rep. 3, 14–21.

Mueller, U., Wienholt, L., Kleinhans, D., Cussmann, I., Bunke, W., Pleßmann, G., Wendiggensen, J., 2018. The eGo grid model: An open source approach towards a model of German high and extra-high voltage power grids. In: Journal of Physics: Conference Series. Vol. 977, IOP Publishing, 012003.

Netz Niederösterreich, 2021. Anschluss großer Ökostromanlagen. https://www.netz-noe.at/SpecialPages/Kapazitatsauslastung.aspx. (Accessed 20 October 2021).

Netz Niederösterreich GmbH, 2018. Austria-Maria Enzersdorf: Overhead line construction - Construction notice 2018/S 082-185220. URL https://ted.europa.eu/udl?uri=TED:NOTICE:185220-2018:TEXT:EN:HTML&src=0.

OEsterreichs Energie, 2021. Locations of power plants within Austria. https://oesterreichsenergie.at/kraftwerkskarte. (Accessed 24 September 2021).

Ospina, J., Liu, X., Konstantinou, C., Dvorkin, Y., 2021. On the feasibility of load-changing attacks in power systems during the COVID-19 pandemic. IEEE Access 9, 2545–2563. http://dx.doi.org/10.1109/ACCESS.2020.3047374.

Oswald, B., 2007. 380 kV Salzburg line - impact of the possible (partial) cabling of the new section Tauern-Salzach. URL https://renewables-grid.eu/uploads/tx_nbrgilgext/Oswald_2007_380kV_Salzburgleitung_-_Auswirkungen_der_mo_glichen_Teil-Verkabelung_des_Abschnitts_Tauern-Salzach_neu.pdf.

Pastor-Galindo, J., Nespoli, P., Gómez Mármol, F., Martínez Pérez, G., 2020. The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. IEEE Access 8, 10282–10304. http://dx.doi.org/10.1109/ACCESS.2020.2965257.

Peles, S., 2021. GridKit. https://www.osti.gov/biblio/1340202. (Accessed 16 August 2021).

Power Clearing & Settlement Austria, 2021. Sythetic load profiles: Consumption forecasts. https://www.apcs.at/en/clearing/physical-clearing/synthetic-load-profiles. (Accessed: 17 October 2021).

Priesmann, J., Nolting, L., Kockel, C., Praktiknjo, A., 2021. Time series of useful energy consumption patterns for energy system modeling. Sci. Data 8 (1), 148.

Scharf, M., Nebel, A., 2016. osmTGmod load-flow model. https://github.com/wupperinst/osmTGmod. (Accessed 16 August 2021).

Schultz, M., Voss, J., Auer, M., Carter, S., Zipf, A., 2017. Open land cover from OpenStreetMap and remote sensing. Int. J. Appl. Earth Obs. Geoinf. 63, 206–213. http://dx.doi.org/10.1016/j.jag.2017.07.014, URL https://www.sciencedirect.com/science/article/pii/S0303243417301605.

Soltan, S., Mittal, P., Poor, H., 2018. BlackIoT: IoT botnet of high wattage devices can disrupt the power grid. In: 27th USENIX Security Symposium. pp. 15–32.

Steiermark, E., 2021. Sichere stromversorgung für die steiermark. https://www.lko.at%2Fmedia.php%3Ffilename%3Ddownload%253D%252F2015.01.19%252F1421657409344297.pdf. (Accessed 17 October 2021).

Szelgrad, M., 2014. Vorarlberg: PV-Unternehmer wehren sich gegen "unfairen Wettbewerb". https://report.at/energie/19143-vorarlberg-pv-branche-wehrt-sich-gegen-unfairen-wettbewerb. (Accessed 11 September 2021).

Tidy, J., 2022. Ukrainian Power Grid 'Lucky' to Withstand Russian Cyber-Attack. BBC, URL https://www.bbc.com/news/technology-61085480.

TINETZ, 2021. Netzkennzahlen. https://www.tinetz.at/unternehmen/ueber-uns/netzkennzahlen/. (Accessed 17 October 2021).

Vorarlberger Übertragungsnetz GmbH, 2018. Netzentwicklungsplan 2018.

Vorarlberger Übertragungsnetz GmbH, 2021. Netzentwicklungsplan 2021.

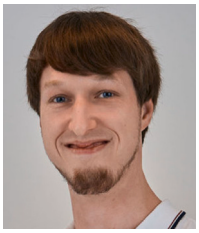
Weniger, B., 2019. Statische Neudimensionierung für Freileitungsmasten. http://www.weninger-zt.at/images/downloads/Download_Statikbeispiel%20HSP-Mast.pdf. (Accessed 11 August 2021).

Werner, F., 2012b. Energiesysteme im Umbruch - Erfahrung und Realisierung Windkraftwerksprojekte. https://fachportal.ph-noe.ac.at/fileadmin/fwz/etech/Energiesysteme/7_Erfahrung_Realisierung_Windkraftprojekte.pdf. (Accessed 01 October 2021).

Wiener Netze, 2021. Dachgleiche am 47. Umspannwerk für Wien und Umgebung. https://www.wienernetze.at/dachgleiche-am-47.-umspannwerk-f%C3%BCr-wien-und-umgebung. (Accessed 17 October 2021).

Yadav, A., Kumar, A., Singh, V., 2023. Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. Artif. Intell. Rev. 56 (11), 12407–12438.

Anja Klauzer is a researcher at SBA Research at the Networks and Critical Infrastructures Security Research Group. She received a Bachelor's degree of science in Geography and a Master's degree in Cartography and Geographic Information Science at the University of Vienna. Her research interest include geographic information, geospatial data and critical infrastructure in combination with GIS.



Markus Maier is a Ph.D. candidate at University of Vienna, Austria, Research Group Security and Privacy. He received his B.Sc. and M.Sc. in Software Engineering & Internet Computing from Technical University of Vienna. His research interests include network measurement and network data analysis.

Lore Abart-Heriszt is a Senior Scientist at the Institute for Spatial Planning, Environmental Planning and Land Management at the University of Natural Resources and Life Sciences. Her research focuses on contributions to spatial energy planning, including the modeling of energy consumption and greenhouse gas emissions in different spatial contexts as a basis for spatial planning policy decision-making processes. With Energiemosaik Austria, she has developed a municipal energy and greenhouse gas database for Austria and made it available online in order to contribute to the discussion on the spatial relevance of the energy transition and climate protection.

Johanna Ullrich is a key researcher at SBA Research, Austria, leading the Networks and Critical Infrastructures Security Research Group, and a researcher of the Christian Doppler laboratory for Security and Quality Improvement in the Production System Lifecycle (University of Vienna). She received a M.Sc. in Automation Engineering/Electrical Engineering and a Ph.D. sub auspiciis praesidentis from TU Wien. Her research focuses on network security, and security at the intersection of computer science and traditional engineering.