

ORIGINAL RESEARCH

Generating peak-aware pseudo-measurements for low-voltage feeders using metadata of distribution system operators

Manuel Treutlein^{1,2}  | Marc Schmidt²  | Roman Hahn²  | Matthias Hertel¹  |
Benedikt Heidrich¹  | Ralf Mikut¹  | Veit Hagenmeyer¹ 

¹Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

²Netze BW GmbH, Stuttgart, Germany

Correspondence

Manuel Treutlein.

Email: manuel.treutlein@partner.kit.edu

Funding information

Helmholtz-Gemeinschaft, Grant/Award Number: Energy System Design, Helmholtz.AI; Netze BW GmbH

Abstract

Distribution system operators (DSOs) face challenges such as restructuring distribution grids for climate neutrality and managing grid consumption and generation. Measurements within the grid are crucial for DSOs, yet many low-voltage (LV) grids lack measurement devices. To address this, an approach is proposed to estimate pseudo-measurements for non-measured LV feeders using regression models. The models are based on feeder metadata, which includes the number of grid connection points, installed power of equipment, and billing data in the downstream LV grid. The authors also incorporate weather, calendar, and timestamp data as model features and use the existing measurements as the model target. For evaluation, a dataset of 2323 LV feeders is used and peak metrics for magnitude, timing, and shape of consumption and feed-in are introduced, inspired by the BigDEAL challenge. The authors employ XGBoost, a multilayer perceptron (MLP), and a linear regression (LR) model, finding that XGBoost and MLP outperform LR. The results demonstrate that this approach effectively adapts to varying conditions and generates realistic load curves from feeder metadata. Additionally, the authors elaborate on feeders where pseudo-measurements exhibit deficiencies. This method could be extended to other grid levels such as substations and contribute to research in load modelling, state estimation, and LV load forecasting.

KEYWORDS

artificial intelligence and data analytics, asset management, big data, data models, decision trees, distribution networks, load forecasting, meta data, metering

1 | INTRODUCTION

The increase of photovoltaic (PV) systems, heat pumps and electric vehicles (EVs) implies challenges and new requirements for the planning and operation of distribution grids [1]. Distribution system operators (DSOs) have to cope with increased load and generation which results in higher peaks of active power. In order to prevent congestions, for example due to thermal overload of DSO equipment, it is essential for the DSO to know the loads induced by consumption and generation in the electric grid [2].

However, we observe that the rollout of measurement devices for DSO equipment such as low-voltage (LV) feeders is

expensive and time-consuming. Consequently, the LV feeders of DSOs are often divided into a set A of measured feeders and a set B of non-measured feeders as it is depicted in Figure 1.

For closing the measurement gap, it can be beneficial to estimate the active power of non-measured LV feeders and substations. The estimated power values are subsequently named pseudo-measurements. In the present paper, we propose an approach based on feeder metadata such as the installed power of heat pumps or PV systems along with the available measurements. Table 1 provides an overview over the specific feeder metadata used in the present paper. However, the feeder metadata can be extended depending on the available data.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *IET Smart Grid* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

A model is trained to learn how a measured power curve of a LV feeder is related to the feeder metadata, the current timestamp and the weather conditions. During application, the model can predict the power of a non-measured feeder (a feeder from set B in Figure 1) with information about its metadata and the current timestamp and weather conditions. This approach does not require any measurements from the feeder for which the pseudo-measurements are generated. As long as the grid is not meshed, this pseudo-measurement approach is also applicable at substations by leveraging substation metadata.

Pseudo-measurements can serve as a basis for multiple use cases in planning and operation processes. For planning, they indicate the remaining capacity in the feeder and facilitate the integration of new consumers and producers. If the feeder capacity is too low, an installation of measurement devices for more accurate monitoring or a grid expansion can be triggered. Moreover, the energy transition requires studies for the long-term planning of the grid in view of increasing load and generation. In Germany, DSOs are obligated to plan their grid with respect to climate neutrality in the year 2045 [3]. For this purpose, it is important for the DSOs to have knowledge about the current degree of capacity utilisation. For operation, pseudo-measurements can improve the accuracy of a Distribution System State Estimation (DSSE) [4]. Furthermore, the

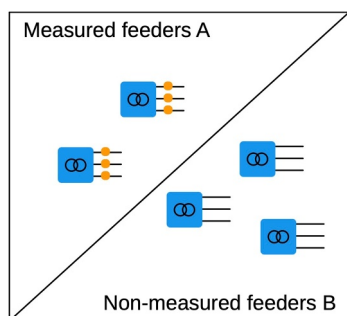


FIGURE 1 Classification of LV feeders into measured (A) and non-measured feeders (B). The symbols represent substations with three LV feeders. The present paper aims to estimate pseudo-measurements for non-measured feeders B by learning the relation between measurements of feeders in A and the feeder metadata (see Table 1).

TABLE 1 Overview over the feeder metadata which is used as model features in the present paper. Section 5.1 provides more details about the feeder metadata. The methodology in the present paper enables other feeder metadata depending on which data is available. Other model features which are not listed are weather and calendar features.

Group	Type	Metadata features	Unit
Grid connection points	Housing	housing units	# (Count)
Installed power of equipment	Consumer	storage heaters, heat pumps, electric heaters, EV chargers, hot water tanks, inductive power, flow-type heaters, public lighting, other consumers	kW
	Battery	batteries	kW
	Producer	PV systems, other producers	kW
Energy consumption data	Commerce and industry	g0 (general), g1 (workdays), g2 (evening), g3 (continuous), g4 (shop/hairdresser), g5 (bakery), g6 (weekend), l0 (farm)	kWh

control of consumers and producers in case of grid congestions such as thermal overload is becoming more important. In Germany, a new regulation allows the reduction of power consumption by devices in the LV grid under specific circumstances [5]. Hereby, pseudo-measurements help to detect grid congestions and can be used to set up control schedules.

Hence, in the present paper, we introduce a new approach to generate pseudo-measurements based on feeder metadata. We use a large real-world dataset from a DSO in Southern Germany which includes metadata which is rare in the current literature [6]. Furthermore, the measurement data is characterised by both load and feed-in. From 2323 feeders, 34.1% exhibit at least 10 times a net feed-in lower than -10 kW mainly due to prosumers with PV systems. Hence, we modify existing peak metrics to handle frequent zero-crossings as well as negative and positive peaks at one day.

Our key contributions in the present paper are.

- We present a new approach to generate pseudo-measurements for LV feeders based on feeder metadata.
- We evaluate the pseudo-measurements with respect to peak metrics inspired by the BigDEAL challenge [7] considering the magnitude, timing and shape of consumption and feed-in peaks.

The remainder of the paper is structured as follows. The related literature is presented in Section 2 and the problem statement is given in Section 3. The methodology in Section 4 introduces the machine learning (ML) approach of generating pseudo-measurements as well as the peak metrics. After introducing the dataset and the experiment in Section 5, we present the pseudo-measurements in Section 6. Finally, we provide a discussion in Section 7, the limitations in Section 8 and a conclusion in Section 9.

2 | RELATED WORK

In this section, we first give an overview on different approaches to estimate the load and generation of a non-measured feeder in the LV grid. Furthermore, we delimit the problem from adjacent research areas.

TABLE 2 Basic approaches in the literature which can be used to estimate the load and generation of non-measured LV feeders and other grid levels.

#	Approach	Type	Important data requirement	Study examples	Reviews
1	Smart meter exploitation	Bottom-up	AMI in downstream grid	[8–11]	[4]
2	Measurement disaggregation	Top-down	AMI in upstream grid	[8]	
3	Explicit modelling	Bottom-up	Deep domain knowledge + data about equipment in LV grid	[12, 13]	[14]
4	Synthetic load profiles	Bottom-up	Representative measurements of grid connection customers in any downstream grid + data about equipment in LV grid	[15–17]	[4, 14]
5	Leveraging grid measurements	Same grid level	Representative feeder measurements in any grid + data about equipment in LV grid	[8, 18–20]	

Abbreviations: Advanced Metering Infrastructure (AMI), low-voltage (LV).

The lack of measurements in the grid is a common problem for DSOs and researchers focussing on distribution grids. Table 2 shows different approaches to estimate the load. In particular, pseudo-measurements are essential for DSSE if the grid is unobservable [21]. But as emphasised in Section 1, there are also other DSO use cases for pseudo-measurements of LV feeders besides DSSE such as detecting a congestion at the feeder.

The first approach to generate pseudo-measurements in Table 2 is the *smart meter exploitation* in the downstream grid. In case of an (almost) complete coverage by and availability of Advanced Metering Infrastructure (AMI), the measurements at the grid connection customers can be aggregated to the feeder level (bottom-up) [8, 11]. Furthermore, information from smart meters can be used to adjust load profiles [10] or can be included directly in the DSSE optimisation problem [9]. However, the smart meter exploitation approach can be complex due to millions of distributed smart meters. More important, the data is often not available due to slow smart meter rollouts [22]. This also applies for the grid area used in the present paper, which is why smart meter data is not used. Furthermore, legal and technical constraints can be a reason for restricted access of DSOs to smart meter data [23].

Inversely, the second approach is the *measurement disaggregation* from the upstream grid (top-down). A simple approach is to use ratios given by the kVA rating of the transformers to distribute the measured load of medium-voltage (MV) feeders [8]. A disadvantage is the inaccuracy of the estimates, depending on the data used for building ratios. Disaggregation approaches are related to Non-Intrusive Load Monitoring (NILM) on the household level [24] which can also be extended to the grid level [25]. However, NILM on the grid level aims to detect the contribution of equipment described in the feeder metadata, for example, the proportion of power generated by PV systems. This is not the aim in the present paper.

The third approach *explicit modelling* is used in refs. [12, 13] to build pseudo-measurements for wind and solar generation. If domain knowledge about the generation of customers in the downstream grid is given, it is possible to model the estimated load by physical equations. Regarding the consumption of grid connection customers, a detailed bottom-up approach which combines the existing electrical devices with human activity patterns can be used [14]. The resulting

generation and load profiles can be aggregated to the higher grid level (bottom-up). The main disadvantage of this approach is the complexity of modelling, especially with diverse customers such as different types of commerce and industry.

We refer to the fourth approach as *synthetic load profiles* which can be aggregated to the higher grid level (bottom-up). In contrast to the third approach, the referenced methods need no detailed information about human activity or physical equations. Unlike the first approach, the load and generation profiles can be built with representative data from any downstream grid and we do not need AMI in the downstream grid of the non-measured feeder. In Germany, the load profiles provided by the German Association of Energy and Water Industries (BDEW) are often used [15]. However, they are criticized for being imprecise [26]. More advanced examples comprise a Variational Autoencoder (VAE) called Faraday which can provide synthetic household profiles conditioned on consumption devices such as EVs [17]. Nevertheless, the aggregation of load profiles does not incorporate the simultaneity factors of customers and devices in the downstream LV grid.

In contrast to the aforementioned approaches, the fifth approach consists in directly *leveraging grid measurements* of feeders which are representative for the non-measured feeder (same grid level). The study presented in the present paper belongs to this fifth approach. We refer to the required data about the equipment in the LV grid (see Table 2) as feeder metadata. In ref. [18], the feeder metadata comprises information about the residential and commercial customers and the public lighting. The authors in ref. [19] frame the problem of generating load profiles for non-measured MV/LV transformers as transductive transfer learning. Despite solar irradiance, social demographic data and the installed power of PV systems, they include load data from a similar transformer after clustering. Closest to the present paper, the authors in ref. [20] propose a method based on substation metadata, weather data and calendar data to estimate load of non-measured substations which can also contribute to DSSE [27]. However, the presentation does not contain an evaluation of their approach.

In general, it can be observed that using feeder metadata is rare in the literature, also due to missing public datasets [6]. If feeder metadata is used, it is often not detailed, for example, including only the number of households without knowledge

about large consumption devices like the number of EV charging stations. However, such information can be crucial for the resulting load at the feeder level [28].

Other research areas related to pseudo-measurements cover the generation of synthetic time series and load forecasting. For synthetic time series, the difference is that they are not for specific feeders but should include specific characteristics, for example, anomalies [29, 30]. Load forecasting requires measurements from the feeder itself [2]. The emerging Time Series Foundation Models (TSFMs) are focused on forecasting a given time-series, imputing missing values in a time-series or detecting anomalies in time-series [31]. However, the problem in the present paper consists of having no measurements at all for the majority of feeders. Nevertheless, load forecasting is a very active research field and methods used for data preparation, training and evaluation of time-series problems can also support regression problems like the one in the present paper [6]. For example, the generative load forecasting model conditional Invertible Neural Network can also be used to generate pseudo-measurements based on feeder metadata [32, 33].

In literature about DSSE, historical load data of a LV feeder is often given, but recent or future load data is missing [4, 34]. Generated recent or future load estimates of the LV feeders (and other grid levels) are also called pseudo-measurements. However, analogous to load forecasting, the methods need historical data of the feeder itself. Hence, they are not suited for the problem statement in the present paper.

In addition to the methods, the tailored evaluation of load estimations plays a crucial role in distribution grids. Since DSOs need to plan and operate the grid with respect to the highest loads, evaluating with respect to the peak values is becoming more important. The BigDEAL challenge [7] in the field of load forecasting addresses this requirement as well as the authors in ref. [35] for DSSE. In the present paper, we adapt the peak metrics of ref. [7] considering peak magnitude, timing and shape to cope with feeders exhibiting both consumption and feed-in.

All in all, we observe that feeder metadata together with feeder measurements is rarely used in the literature to generate pseudo-measurements. In addition, none of the studies listed at the fifth approach (leveraging grid measurements) in Table 2 contain as much metadata as in the present paper (see Table 1). Furthermore, all of the studies of the fifth approach focus on the substation or transformer level instead of the feeder level and do not evaluate explicitly on peak performance. Moreover, the aim in refs. [18, 19] is to generate load profiles instead of complete time-series as in the present paper. However, they induce simplifications, for example, by neglecting weather effects.

Therefore, we propose an ML method generating high resolution pseudo-measurements based on feeder metadata. The results are evaluated on a large real-world dataset with three models and nine metrics with a special focus on daily peak values.

3 | PROBLEM STATEMENT

Figure 1 illustrates a basic problem for DSOs in the LV grid. Even though an increasing number of LV feeders are equipped with measurement devices, there are still many feeders without

any measured values. Therefore, our aim is to estimate pseudo-measurements for the non-measured feeders. It is important that the feeders in A are representative for the whole grid ($A \cup B$) so that the estimation for the whole grid delivers reliable pseudo-measurements.

We divide the LV feeders into a set A and a set B depending on the existence of measurements. Feeders in set A can be used for training, whereas during inference, the unknown active power is estimated for feeders in B . The aim of the model training is to learn the parameters θ of the function f with

$$\hat{y}_{ij} = f_{\theta}(M_{ij}, W_{ij}, C_j), \quad (1)$$

where $M_{ij} \in \mathbb{R}_+^k$ is the metadata (see Table 1) of the feeder i for timestamp j , $W_{ij} \in \mathbb{R}^l$ is the weather data related to the feeder i for timestamp j and $C_j \in \{0, 1\}^m$ is calendar data for timestamp j . The variables k , l and m represent the dimension of the respective vector. The pseudo-measurement is denoted with \hat{y}_{ij} .

The metadata M_{ij} of a feeder i change rarely compared to the weather data W_{ij} . Reasons for changes are mostly new registered equipment in the downstream grid of i . The measurement period can differ significantly between different feeders i due to a continuous rollout of measurement devices. Furthermore, the measurements in the measured period can exhibit measuring gaps. We denote the set of timestamp indices where measurements for a feeder i are existing with N_i .

A given function f_{θ} not only allows creating pseudo-measurements for the past and current timestamps. A forecast of the measurement can be received by passing predicted weather data W_{ij} to f_{θ} . Furthermore, different mid-term (months - few years) and long-term (many years) scenarios with varying metadata M_{ij} or weather data W_{ij} can be conducted, for example, to simulate new installations and ramp-ups of electrical equipment.

4 | FEEDER METADATA-BASED PSEUDO-MEASUREMENTS

This section describes the methodology to generate pseudo-measurements for feeders without measurements based on feeder metadata. It includes data requirements and the model pipeline which is visualised in Figure 2. Furthermore, the peak metrics which are adapted from the BigDEAL challenge [7] are introduced.

4.1 | Data requirements

The required data to train and operate a pseudo-measurement model based on metadata comprises different data sources. In order to provide a target variable for the model training, a sufficiently large amount of measured LV feeders in the grid is needed.

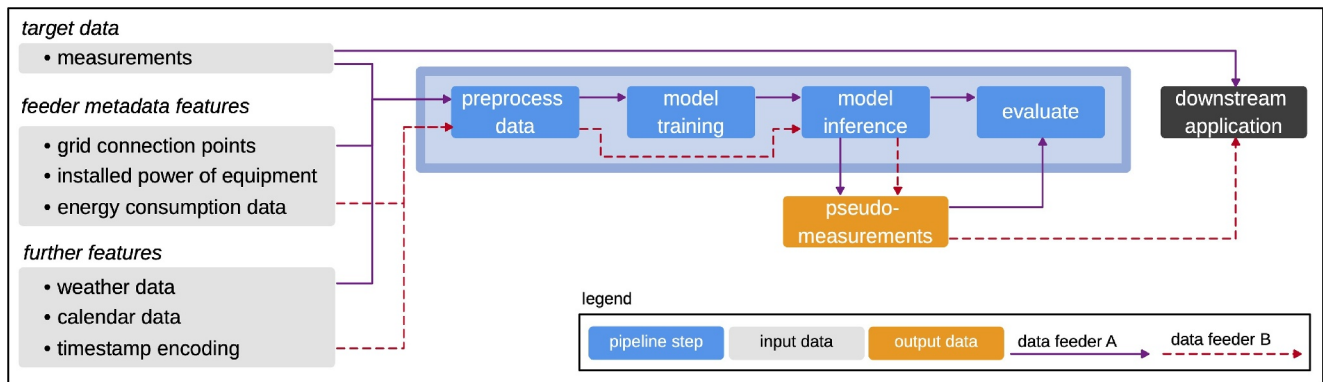


FIGURE 2 Model pipeline with the pipeline steps, input data and output data. The data flows are differentiated with respect to measured feeders *A* and non-measured feeders *B* as described in Section 3.

The essential feature data for our proposed model is metadata collected by the DSO. The concrete metadata can differ depending on the available data. We divide the feeder metadata into three groups.

The number of housing units or the number of commerce and industry units belong to the first group of grid connection points. The second group comprises installed power of equipment which describes producers such as PV systems or combined heat and power (CHP) and consumers such as heat pumps or household units. It is necessary that the equipment is allocated to the respective LV feeder and is registered with the installed electrical power. The third group of feeder metadata provides information about the energy consumption. Note that energy consumption data does not mean that the grid node has measurements. Rather, it is highly aggregated consumption data, for example, annual consumption of grid connection points in the LV grid. This can improve the model because the grid connection points or installed power of equipment can differ based on different usage behaviour of houses and industries at the LV feeder. The consumption data is used as a feature, therefore it needs to be known for both measured (training) and non-measured (inference) feeders.

In addition to feeder metadata, appropriate weather data should be provided since generation and consumption of installed equipment depend highly on the weather. Examples of relevant variables include temperature in case of heating equipment or global radiation in case of PV systems.

Moreover, calendar information about holidays or other events can improve the model accuracy, because the behaviour of people differs between workdays, weekends and holidays. Finally, a timestamp encoding is needed depending on the required temporal resolution of the pseudo-measurements.

4.2 | Model pipeline

Figure 2 illustrates the model pipeline with the four pipeline steps from the provided input data to the downstream application which consumes the model output data. We differentiate the data flow between feeders which belong to the set *A* and feeders belonging to set *B* (compare with Figure 1).

The flow for the feeders in *A* represents the pipeline training phase. The data given for feeders in *A* comprises the measurements, metadata and further features which includes weather data, calendar data and timestamp encoding. First, all the data is preprocessed. Due to the diverse data sources, the preprocessing includes aligning the different time resolutions between groups of data or aggregating several metadata features to one feature. The second step is the model training. After applying a train-test split to all feeders in *A*, ML regression models for tabular data are trained on the training data using the measurements as target variable and all other data from the first step as model input features. In contrast to LV load forecasting, the model does not use autoregressive features of the feeder because they are not available for feeders in *B* during the inference phase. In the third step, we conduct a model inference with the test feeders in *A* which produces the pseudo-measurements. Because the ground truth is given for this data, we can evaluate the performance in the fourth step.

The flow for the feeders in *B* represents the pipeline inference phase. Analogue to *A*, the data given for feeders in *B* comprises metadata and all further features. However, there are no measurements for feeders in *B*. Therefore, the preprocessing step only preprocess the features. Afterwards, we use the model trained during the pipeline training phase and conduct the step of model inference for all feeders in *B*. This produces pseudo-measurements for all feeders in *B*. The pseudo-measurements of feeders in *B* are used together with the real measurements of feeders in *A* in the downstream application of the DSO.

There is no need to pass an ID of the LV feeder to the model, because the feeder is defined by the characteristics of its metadata. Furthermore, the model does not forecast a time-series, because the feeders are in general not measured and there is no data available. However, if the features represent future meta- and weather data, the architecture can also be used to forecast.

4.3 | Metrics for evaluation

In order to evaluate the pseudo-measurement model for individual feeders, we use the following nine metrics which are

grouped into three all-observation metrics and six peak metrics.

Let y_{ij} be the measured value of feeder i at timestamp j and \hat{y}_{ij} the respective estimated pseudo-measurement. We denote the mean absolute error (MAE) with

$$\text{MAE}_i = \frac{1}{|N_i|} \sum_{j \in N_i} |y_{ij} - \hat{y}_{ij}| \quad (2)$$

and the root mean squared error (RMSE) with

$$\text{RMSE}_i = \sqrt{\frac{1}{|N_i|} \sum_{j \in N_i} (y_{ij} - \hat{y}_{ij})^2} \quad (3)$$

as *all-observation* metrics because the model is evaluated against all existing measurements of a feeder. Hereby, N_i is the set of timestamp indices of a feeder i where measurements are available and $|N_i|$ the cardinality. Furthermore, we denote

$$\text{MAE}_{i,\text{norm}} = \frac{\text{MAE}_i}{\max\{y_{ij} \mid j \in N_i\} - \min\{y_{ij} \mid j \in N_i\}} \quad (4)$$

as the MAE normalised by the min–max range of the feeder i .

Additional to the all-observation metrics, we introduce the metrics focussing on peaks. The metrics in the present paper are inspired by the BigDEAL forecasting challenge [7] and adapted to incorporate special requirements of DSOs facing both consumption and feed-in at the LV feeders.

In the BigDEAL challenge, the applied metrics evaluate the performance regarding the daily peak magnitude, timing and shape for consumption only. The daily peak magnitude is evaluated with the mean absolute percentage error (MAPE) and the daily peak timing with a non-weighted MAE (qualifying match) and a weighted MAE (final match). Regarding the daily peak shape, the authors in ref. [7] evaluate the deviation between the normalised loads 2 hours before and after the peak. The normalisation is based on the (estimated) peak load.

Compared to ref. [7], we make five adaptations in total:

- (1) The metrics are evaluated for positive and negative peaks to consider both the consumption and feed-in estimation.
- (2) Two thresholds are introduced so that only feeders with at least 10 daily peaks exceeding ± 10 kW are evaluated.
- (3) Regarding the daily peak magnitude, we use MAE instead of MAPE.
- (4) Regarding the daily peak shape, we use a min-max normalisation instead of a peak normalisation.
- (5) Regarding the daily peak timing, we use the non-weighted MAE as in the BigDEAL qualifying match.

The adaptations are necessary, because unlike in ref. [7], the data used in the present paper is characterised by both consumption and feed-in and many LV feeders exhibit zero-crossing measurements (1). With adaptation (2), days and feeders with low consumption and feed-in are excluded from the peak evaluation. To improve comparability with the

all-observation metrics, we use MAE instead of MAPE for the daily peak magnitude (3). Thereby, the unit kW of the daily peak magnitude is preserved. The many zero-crossings make peak normalisation for the daily peak shape unsuitable (4). With adaptation (5), we facilitate the interpretation and preserve the unit hour. We refer to the peak metrics in the present paper as Peak Magnitude MAE (PMag), Peak Time MAE (PTime) and Peak Shape Error (PShape) differentiated between consumption and feed-in. More precisely, the consumption and feed-in is the resulting net consumption and net feed-in at the feeder after the presumption of all grid connection customers belonging to the feeder.

We introduce the sets of pairs P_i^C and P_i^F to describe the timestamp indices of the peaks for the daily consumption (C) and the daily feed-in (F). For a feeder i , $P_i^C = \emptyset$ if there are less than 10 days where the measurement exceeds $+10$ kW. Otherwise, a tuple $(j_1, j_2) \in P_i^C$ for a day d if the measurements are greater or equal $+10$ kW at least once on d . The first tuple value j_1 is the timestamp index of the maximum in the measurement (ground truth) on day d . The second tuple value j_2 is the timestamp index of the pseudo-measurement maximum on day d . Analogously, we define P_i^F for feed-in peaks lower or equal -10 kW.

Hence, we denote

$$\text{PMag}_i^T = \frac{1}{|P_i^T|} \sum_{(j_1, j_2) \in P_i^T} |y_{ij_1} - \hat{y}_{ij_2}| \quad (5)$$

and

$$\text{PTime}_i^T = \frac{1}{|P_i^T|} \sum_{(j_1, j_2) \in P_i^T} |j_1 - j_2| \quad (6)$$

where $T \in \{C, F\}$ is the peak type, $|P_i^T|$ is the cardinality and $|j_1 - j_2|$ the absolute difference between two timestamps in hours.

Finally, we define the PShape with

$$\text{PShape}_i^T = \frac{1}{|P_i^T|} \sum_{(j_1, j_2) \in P_i^T} \sum_{x \in Q} |s_{ix} - \hat{s}_{ix}| \quad (7)$$

where $Q = \{x \mid x \in (j_1 - 2, \dots, j_1 + 2)\}$ is the set of timestamps 2 hours around the peak in the measurement j_1 (ground truth) with

$$s_{ix} = \frac{y_{ix} - \min\{y_{ij} \mid j \in Q\}}{\max\{y_{ij} \mid j \in Q\} - \min\{y_{ij} \mid j \in Q\}} \quad (8)$$

and

$$\hat{s}_{ix} = \frac{\hat{y}_{ix} - \min\{\hat{y}_{ij} \mid j \in Q\}}{\max\{\hat{y}_{ij} \mid j \in Q\} - \min\{\hat{y}_{ij} \mid j \in Q\}}. \quad (9)$$

representing the (pseudo) measurement min-max normalised around the peak.

5 | EXPERIMENTAL SETUP

First, this section describes the data set including the data cleaning and preparation steps. Second, we describe the structure of the conducted experiments.

5.1 | Data set and feature engineering

We use active power measurements of LV feeders together with metadata from the DSO *Netze BW* in the southwest of Germany. This data is not open due to privacy regulations. The weather data originates from the German Meteorological Service (DWD) [36].

Table 3 summarises basic characteristics of the dataset with 2323 LV feeders. The approximately 52 million measurements in 15 min resolution represent an aggregation from minute resolution data by mean. The measurement period ranges from February 23, 2022 to March 28, 2024 with different starting points for feeders due to the continuous rollout. Therefore, a feeder has on average 22402 measurements which is around 233 days. Meshed feeders are not included in the dataset.

We use filters in order to detect feeders with implausible data which results in the 2323 feeders after applying. Regarding the measurements, a feeder needs in total at least one day of measurements and the measurement must exceed ± 5 kW at least one time. Regarding the feeder metadata, feeders are removed from the dataset where the installed power of a single consumption or generation category exceeds the physical limit of the feeder. Furthermore, feeders without metadata are removed as well as feeders where the feed-in exceeds the installed power of producers by more than 5 kW. If a feed-in is observed between midnight and two a.m., but the only producer are PV systems, the feeder is also removed.

TABLE 3 Different characteristics of the dataset used in the experiments after applying the data cleaning and the feature engineering specified in Section 5.1.

Characteristic	Value
Number of feeders	2323
Number of measurements	52,039,067
Number of metadata categories	21
Avg. measurements per feeder	22,402
Lowest measurement	-146.69 kW
Highest measurement	178.59 kW
First measurement	February 23, 2022
Last measurement	March 28, 2024
Avg. measurement period per feeder	233 days

Figure 3 shows different quantiles of the measurements for all feeders in a daily load profile. It is visible that both consumption and feed-in during the day are present. The median ranges between 0 kW and 10 kW, whereas the 90%, 99% and 99.9% quantiles are close to 25 kW, 50 kW and 75 kW respectively. The feed-in shows a PV curve, where the 0.1% quantile is close to -100 kW. Figure A1 shows the monthly distribution. All months exhibit long tails for both consumption and feed-in. The distribution in the winter is broader compared to the summer. The highest consumption occurs in the winter with a maximum of 178.59 kW. The highest feed-in is in spring and summer with a minimum of -146.69 kW.

We do not include an outlier detection and removal of measurements, since the measurements seem reasonable. Moreover, correctly predicting high values is essential for DSOs and outlier removal involves the risk of removing extreme but correct data.

Before the data is fed into the model, we aggregate similar feeder metadata such as different types of *heat pumps*. In total, this results in 21 categories for feeder metadata summarised in Table 1. We include a feature for the number of *housing units* which belongs to the group of grid connection points. Regarding the installed power of equipment, there are nine features for consumers, one for *batteries* and two for producers. As energy consumption data, we include eight features for different types of commerce and industry named *g0 - g6* and *l0*. They correspond to the customer groups defined in ref. [15]. The features are derived from monthly or yearly billing data and describe the average energy consumption per day in kWh. The yearly billing data of households is not given in the metadata.

All of the feeder metadata categories except the number of *housing units* are visualised in Figure 4. The left part shows the *installed power of equipment*. *PV systems* are installed at 70% of all feeders and show the highest average installed power per feeder. Some categories such as *batteries* are present at many feeders but with relatively low installed power, other categories such as *flow-type heaters* have reverse characteristics. The right part shows the average *energy consumption data* of commerce and industry. The category *g1* for industry which produces mainly on workdays during 8 a.m. to 6 p.m. is dominating. The *housing units* are the only category from the feeder metadata group *grid connection points* as defined in Figure 4.

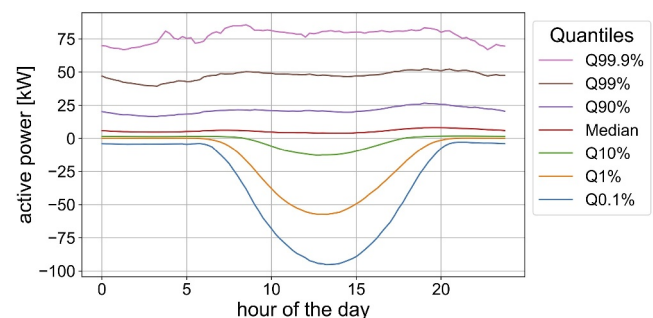


FIGURE 3 Daily profile of the measurements from all feeders for different quantiles in 15 min resolution after applying the data cleaning and feature engineering specified in Section 5.1.

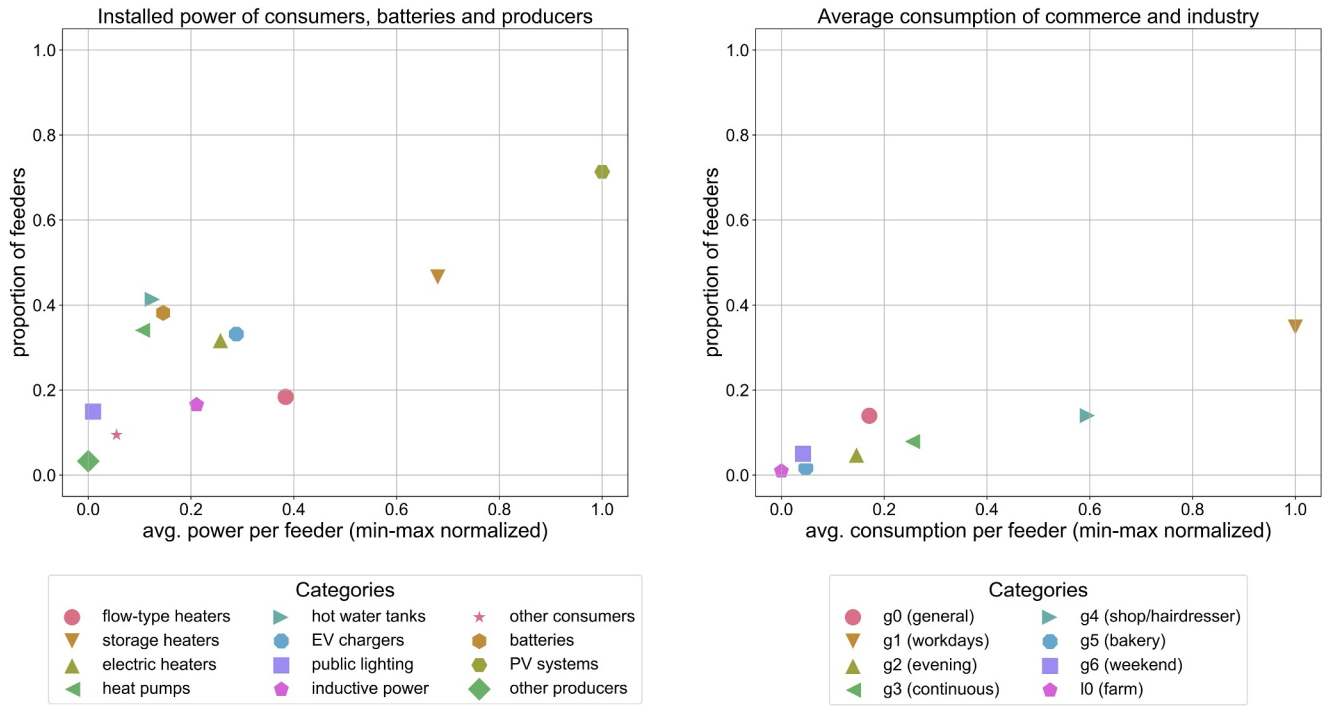


FIGURE 4 Distribution of metadata at LV feeders divided into consumers, batteries and producers (left) as well as the average consumption data of commerce and industry (right). The x-axis is scaled based on the categories *solar* and *g1* respectively. The y-axis describes the fraction of feeders where the category is greater than zero.

Regarding the weather we include four features, namely the global radiation, air temperature, precipitation and the snow height.

In order to model the timestamp information j , we encode the day of the year, the day of the week and the minute of the day with a cyclical encoding using a sine and a cosine function. This results in six features describing the timestamp. We further include two binary features describing if the day is a holiday or workday.

In total, we use 33 features which cover various effects on the target variable representing the active power measurement in kW at the LV feeder.

5.2 | Conducted experiments

The generation of pseudo-measurements is evaluated with three models based on an experiment with a 5-fold cross-validation. This results in 15 sub-experiments with 15 different models as shown in Figure 5. The cross-validation makes it possible to evaluate the method for each model on all feeders, because every feeder is exactly once in the test data.

In each of the 15 sub-experiments, we use a train-test split of the feeders with a ratio of 80%–20%. Applied to Figure 1, the 80% represent the feeder set A and the 20% represent B . Within one sub-experiment, the data of one feeder can be either in the training or in the test data, but not both. We do not split after time as it is common in load forecasting, because the primary problem statement is to estimate values of non-measured feeders and not future values of the same feeder.

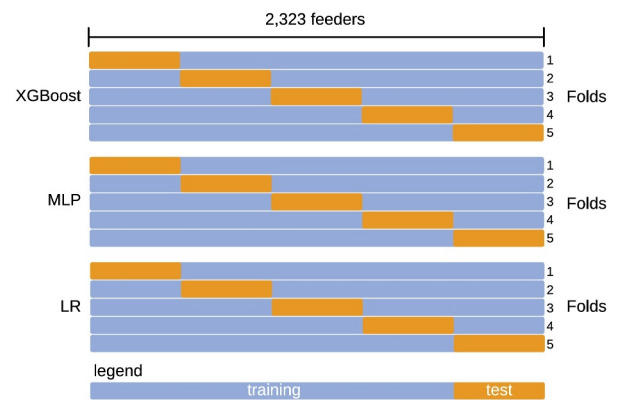


FIGURE 5 Visualisation of the used cross-validation.

All metrics and visualisations reported in Section 6 are solely based on feeders in test data. Regarding Figure 1, all of the feeders used in the experiments belong to the set A . With the retention of feeders in A as test data, we can determine the expected performance of the pseudo-measurements for feeders in B .

The used models are LR [37], MLP [37] and eXtreme Gradient Boosting (XGBoost) [38]. We choose these models to include a linear model, a neural network and a tree-based model. Tree-based models such as XGBoost are state-of-the art for tabular regression problems such as in the present paper [39].

Regarding the XGBoost and the MLP model 12.5% of the training data is used as validation data for early stopping of the training routine to prevent overfitting. The hyperparameters are

chosen based on pre-experiments on similar data and can be seen in Table A1. When using the MLP, we apply a min-max normalisation to the features and targets. The LR is used with elastic net regularisation.

All experiments are conducted with Azure Machine Learning on a compute cluster with 16 cores, 112 GB RAM and 224 GB disk.

6 | RESULTS

First, we present the all-observation metrics and the peak metrics evaluated for all cross-validation runs and all models. Afterwards, excerpts of pseudo-measurements for selected feeders are shown for the XGBoost model of fold 1.

6.1 | All-observation metrics

Table 4 includes different statistical measures for the metrics MAE, MAE_{norm} and RMSE evaluated for the models XGBoost, MLP and LR.

TABLE 4 Metrics evaluated for all timestamps and all feeders (all-observation metrics). The metric is evaluated on the feeders in the test data of a cross-validation fold and the results of all folds are combined into this table.

Metric	Statistical measure	XGBoost	MLP	LR
MAE [kW]	Mean	4.73	4.85	6.07
	Std	3.35	3.78	4.01
	Min	0.70	0.76	1.24
	25%	2.47	2.51	3.71
	Median	3.83	3.84	4.84
	75%	5.83	5.90	7.16
	Max	33.98	64.51	90.55
MAE _{norm}	Mean	0.13	0.13	0.17
	Std	0.24	0.22	0.21
	Min	0.01	0.02	0.03
	25%	0.06	0.07	0.08
	Median	0.09	0.09	0.12
	75%	0.13	0.14	0.19
	Max	7.24	6.34	5.38
RMSE [kW]	Mean	6.19	6.33	7.85
	Std	4.35	4.77	5.19
	Min	0.88	0.96	1.79
	25%	3.17	3.25	4.62
	Median	5.09	5.05	6.12
	75%	7.86	7.89	9.51
	Max	45.41	70.80	91.02

Note: Best values of the three models for a statistical measure are bold.

XGBoost and MLP show superior performance compared to LR for all three metrics. The average MAE over all feeders for XGBoost is 4.73 kW which is 1.34 kW better compared to 6.07 kW for LR. The performance of XGBoost is slightly better for the mean of the all-observation metrics compared to MLP.

All metric distributions for all models are heavily right skewed which is stated by *median < mean*.

The model performance differs between different train-test-splits which is shown in Table A2. For example, the standard deviation of the median RMSE regarding the five folds is between 0.22 kW and 0.49 kW for all models combined.

6.2 | Peak metrics

In contrast to the all-observation metrics, the peak metrics are only evaluated for feeders with daily consumption and feed-in peaks exceeding ± 10 kW at least 10 times. Table 5 highlights the P_{Mag}, P_{Time} and P_{Shape} metric for the consumption (1921 feeders evaluated) and feed-in (793 feeders evaluated).

Analogous to the all-observation metrics, the P_{Mag} for consumption and feed-in is heavily right skewed throughout all models except for the feed-in P_{Time}. For example, the median of the P_{Mag}^C for the MLP is with 8.55 kW lower compared to the mean of 11.51 kW.

The P_{Mag} for both consumption and feed-in is significantly higher compared to the MAE for all models. While the median MAE of the XGBoost model is 3.83 kW, the median of the P_{Mag}^C is 9.02 kW and the median of P_{Mag}^F is 11.44 kW. This implies that the peak estimation of the models is worse than the estimation of the complete time-series.

When comparing consumption and feed-in, the P_{Mag} for consumption is lower compared to feed-in with respect to the mean and the quantiles given in Table 5 for all models. For example, the mean of P_{Mag}^C of XGBoost is with 11.66 kW lower than the mean of P_{Mag}^F with 13.13 kW.

P_{Time} is significantly better for feed-in compared to consumption. While the mean of the P_{Time}^F is 1.17 h for XGBoost, the mean of P_{Time}^C is 5.07 h.

P_{Shape} is better for feed-in compared to consumption. While the mean of the P_{Shape}^F is 0.29 for XGBoost, the mean of the P_{Shape}^C is 0.35.

The P_{Shape}^F shows smaller fluctuations compared to P_{Shape}^C. We observe for all models combined that the 25% and 75% quantile is between 0.28 and 0.30 for P_{Shape}^F and between 0.31 and 0.41 for P_{Shape}^C.

The P_{Mag} and P_{Time} reveal some feeders with very high deviations of the peaks in magnitude and time. For example, the highest errors for P_{Time}^C can be observed for feeders with storage heaters, exhibiting a peak in the night after midnight while the pseudo-measurement estimates the peak in the evening.

TABLE 5 Peak metrics evaluated for all feeders and daily consumption and feed-in peaks. The metric is evaluated on the feeders in the test data of a cross-validation fold and the results of all folds are combined into this table.

Metric	Statistical measure	XGBoost	MLP	LR
PMag ^C [kW] (consumption)	Count	1921	1921	1921
	Mean	11.66	11.51	14.06
	Std	9.09	9.46	10.76
	Min	0.76	0.98	0.80
	25%	5.85	5.59	6.57
	Median	9.02	8.55	11.24
	75%	14.48	14.35	18.39
	Max	88.02	117.71	117.08
PMag ^F [kW] (feed-in)	Count	793	793	793
	Mean	13.13	12.52	21.36
	Std	7.50	7.68	13.24
	Min	1.06	1.81	1.59
	25%	7.98	7.32	11.83
	Median	11.44	10.78	18.54
	75%	16.60	15.14	28.10
	Max	55.56	67.21	73.71
PTime ^C [h] (consumption)	Count	1921	1921	1921
	Mean	5.07	5.14	8.02
	Std	2.82	2.80	1.39
	Min	0.54	0.56	2.10
	25%	2.97	3.03	7.16
	Median	4.40	4.45	7.94
	75%	6.54	6.67	8.90
	Max	16.33	16.57	13.35
PTime ^F [h] (feed-in)	Count	793	793	793
	Mean	1.17	1.27	1.13
	Std	0.69	0.81	0.43
	Min	0.50	0.50	0.54
	25%	0.95	0.97	0.97
	Median	1.05	1.09	1.08
	75%	1.19	1.29	1.21
	Max	12.04	11.24	9.15
PShape ^C (consumption)	Count	1921	1921	1921
	Mean	0.35	0.34	0.38
	Std	0.05	0.05	0.04
	Min	0.19	0.18	0.23
	25%	0.32	0.31	0.35

TABLE 5 (Continued)

Metric	Statistical measure	XGBoost	MLP	LR
	Median	0.35	0.34	0.38
	75%	0.38	0.37	0.41
	Max	0.60	0.54	0.55
PShape ^F (feed-in)	Count	793	793	793
	Mean	0.29	0.29	0.29
	Std	0.02	0.02	0.02
	Min	0.23	0.23	0.23
	25%	0.28	0.28	0.28
	Median	0.29	0.29	0.29
	75%	0.30	0.30	0.30
	Max	0.40	0.37	0.39

Note: Best values of the three models for a statistical measure are bold.

6.3 | Selected feeders

Figure 6 depicts two feeders F_1 and F_2 with the respective feeder metadata and an excerpt of 1 week for the measurements and generated pseudo-measurements as well as temperature and global radiation. The chosen week is a week in the winter of January 2024 with a temperature anomaly at Wednesday, January 17 to Thursday, January 18. The global radiation is between 0 and $333 \frac{\text{W}}{\text{m}^2}$. F_1 represents the 10% quantile and F_2 the 90% quantile of the MAE_{norm} metric in a pre-experiment. More details are given in the appendix: Figure A3 shows weekly quantile profiles for the feeders F_1 and F_2 and in addition for three other feeders $F_3 - F_5$. Figure A2 includes the individual metrics for $F_1 - F_5$ with respect to the complete time-series of the feeders (not only the excerpt).

Feeder F_1 is characterised by 13 housing units and diverse metadata including heating systems, EV chargers, batteries and PV systems. In the measurements, we observe a load between -35.94 kW and 33.18 kW and feed-in for at least 4 days with power smaller than 0 kW . The lowest load during the night can be observed during the highest temperature of 9°C on January 17 and 18. Likewise, we observe that the pseudo-measurements are low in the evening of January 17 and the morning of January 18 not exceeding 20 kW . On January 19, we detect the highest load whereas the feed-in peak is less compared to January 16 and January 20 with similar global radiation around $300 \frac{\text{W}}{\text{m}^2}$.

The MAE_{norm} of 0.04 indicates a good fit of the base load and the PMag^C of 7.14 kW is low compared to the fold average of 11.78 kW . While the PTime^C of 7.38 h is high compared to the average, the PShape^C of 0.32 is low in comparison to the average performance. We recognise for January 19 and January 20 that the consumption peak in the measurement is in the morning and the pseudo-measurement consumption peak is in the evening.

The PMag^F of 18.46 kW is higher than the average PMag^F of XGBoost fold 1 with 12.58 kW . We observe that the highest

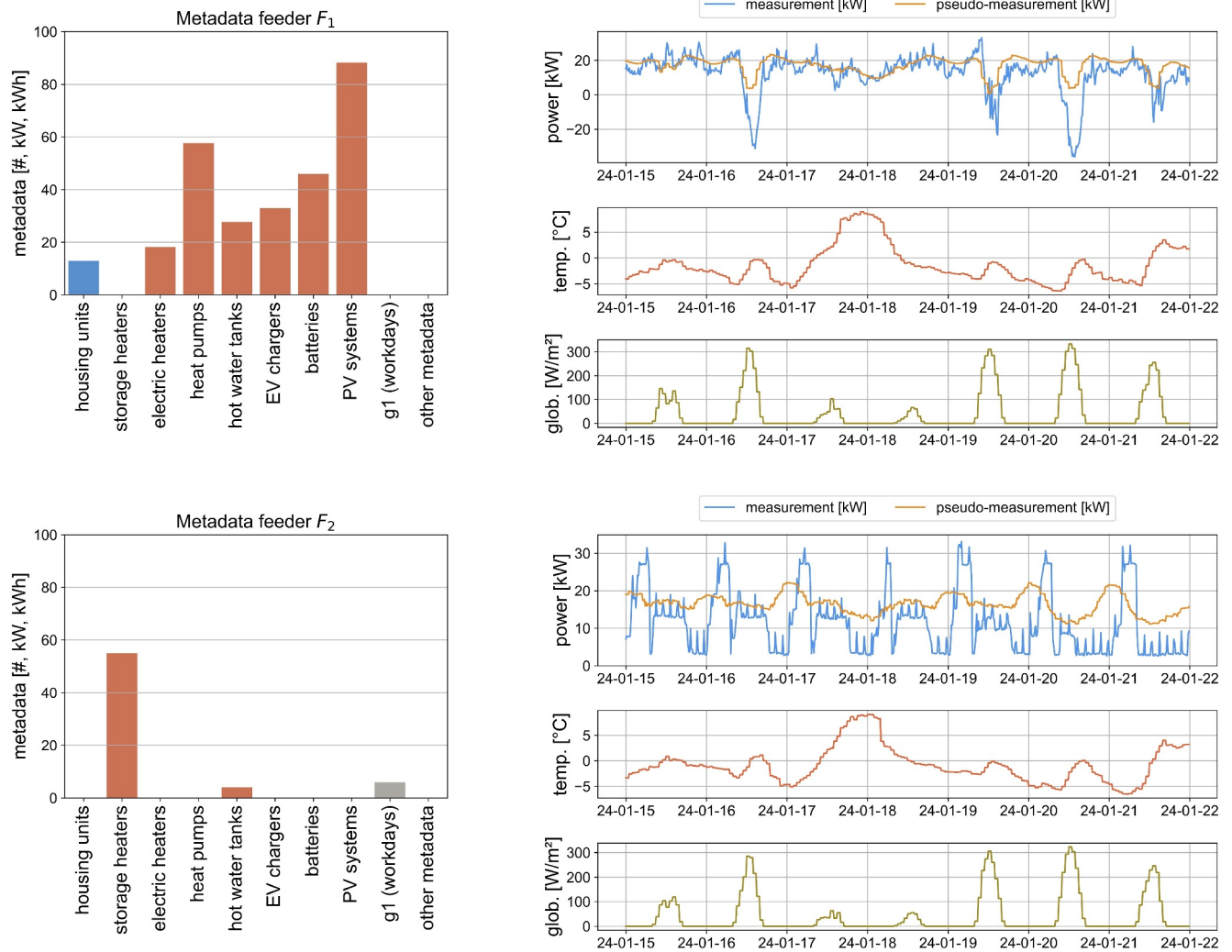


FIGURE 6 Two LV feeders with the feeder metadata on the left side and the measurement and pseudo-measurement on the right side together with the temperature and global radiation of one week in January 2024 (Monday - Sunday). The feeders represent the 10% (F_1) and 90% (F_2) quantile of the metric MAE_{norm} in the XGBoost fold one in a pre-experiment. Metadata with unit count (#) in blue, unit kW in red and unit kWh in grey. The metric values for the feeders are given in Table A3.

error for the feed-in peak estimation is on January 20 with 41.71 kW deviation at 3 pm. The $PTime^F$ and $PShape^F$ of 1.10 h and 0.29 are close to the respective fold averages.

Feeder F_2 is characterised by industry and commerce with storage heaters and a low installed power of hot water tanks. In the measurements, we observe only positive load values with high peaks during the night, an increased load during the working hours from Monday, January 15 to Friday, January 19 and regular load spikes every few hours. The load at the weekend on January 20 and 21 is lower. However, the nightly peaks are still present.

The MAE_{norm} of 0.14 indicates a high MAE with respect to the min-max range of the feeder. In the week shown in Figure 6, we observe that the pseudo-measurements show a higher base load compared to the measurements. At the weekend, we observe the lowest pseudo-measurements. In the nights with the lowest temperature below -5°C , the pseudo-measurements are the highest exceeding 20 kW. In the night with the highest temperatures of the week, it can be observed

that the pseudo-measurements are lower compared to the other nights.

The peak metrics are only evaluated for the consumption due to the constant positive load. The $PMag^C$ of 10.28 kW is lower compared to the fold average of 11.78 kW. In contrast to the measurements, the pseudo-measurements show lower consumption peaks in the night. The $PTime^C$ of 6.56 h and $PShape^C$ of 0.38 are higher than the fold average.

Figure A3 presents a week for the same feeder F_2 but during the summer. We observe that the measurements do not exhibit peak loads during the night and the pseudo-measurements show a lower base load compared to the winter week in Figure 6.

7 | DISCUSSION

First, we discuss the results with respect to the performance of the presented models. Afterwards, we investigate the design of the peak metrics.

7.1 | Model performance

In the following, we discuss the performance of the models in comparison with each other and with respect to different folds of the cross-validation. We further argue about possible model reasons for the pseudo-measurement results based on the XGBoost model of fold 1 as a representative of the examined models.

7.1.1 | Superiority of non-linear model

As already stated in the results section, XGBoost and MLP outperform LR. The models exhibit lower metrics in Tables 4 and 5 compared to LR except for $PTime^F$, $PShape^F$ and some max and min rows. However, the differences of the models regarding $PTime^F$ and $PShape^F$ are small and the statistical measures max and min represent only one feeder out of 2323, respectively. Therefore, we argue that the solution to the problem statement in Section 3 requires a non-linear model.

We treat the problem statement as a regression problem on tabular data. It is still unclear whether tree-based models or neural networks are better suited for this [40]. In our experiments, we cannot identify a superiority of the XGBoost model with respect to MLP or vice versa. It is noteworthy that the MLP with only few parameters consisting of one hidden layer and 20 neurons performs comparatively to the XGBoost model.

7.1.2 | Challenges in estimating pseudo-measurements

The results show that the distributions of the metrics evaluated on all feeders are right-skewed indicating that the generated pseudo-measurements of some feeders do not fit well to the measurements. Therefore, we give four reasons for the deviations.

First, we observe feeders with assumed data quality issues which are difficult to identify and remedy. An example of this is storage heaters which are registered but not deregistered after decommissioning.

Second, there are feeders exhibiting measurements which are hard to estimate. We often observe this for measurements exhibiting short but high consumption peaks, for example, due to heating systems or commerce and industry. An explanation could be that it is better for models to estimate close to the base load to optimise their quadratic loss function instead of estimating peaks. However, it should be noted that the challenge to estimate peaks is also related to the third and fourth reason given next.

Third, the features given to the pseudo-measurement model cannot explain all measurements. For example, the feature housing units does not make a statement if the house is a single-family home or a small apartment. This could be the reason for the high MAE_{norm} metric of 0.19 of feeder F_4 in the

appendix indicating a bad estimation for the base load. Another example is that the only feature for PV systems, the installed power, does not cover details about tilt or azimuth angles of the PV. This could explain some high values for the metric $PTime^F$, for example, for F_1 .

Fourth, the behaviour of the customers in a LV grid is stochastic. An example is EV charging, leading to sudden peaks in the measurements which could be the reason for the peak of feeder F_1 at Friday morning in Figure 6. However, the pseudo-measurements represent point estimations which cannot capture the inherent uncertainty of the measurements.

7.1.3 | High diversity of metadata and measurements

During the data analysis and experiments, we observe a high diversity both in the feeder metadata as well as in the active power measurements for different seasons. This is visible by the diverse weekly load profiles in Figure A2 and the different weeks of feeder F_2 in Figures 6 and A3. The high diversity is a reason why the performances of the same feeders differ regarding the metrics, for instance the $PMag^C$ compared to $PMag^F$ of feeder F_1 in Table A3.

7.1.4 | Performance with respect to peak metrics

Tables 4 and 5 illustrate that the $PMag$ metrics are worse compared to the MAE which is evaluated for all observations. Reasons are stated within this Section 7.1 in the paragraph about the challenges in estimating pseudo-measurement. In general, the difficulty of estimating the $PMag$ depends heavily on the feeder, the feeder metadata and the time of the year.

Regarding the $PTime$, the metric is better for feed-in compared to consumption with respect to mean and median but also regarding the standard deviation. Possible explanations are higher stochastic of consumption compared to feed-in or several consumption peaks on the same day with similar magnitude. This is depicted in Figure 6 for feeder F_1 on Friday and Saturday where the peak for this day is estimated at the evening instead of the morning.

The $PShape$ metric shows better results for the feed-in compared to the consumption. A reason could be that the shape of a feed-in is mainly induced by PV which resembles a bell curve upside down on cloud-free days (compare F_1 and F_3 in Figure A2). In contrast, the shapes of the consumption are more diverse.

7.1.5 | Performance during inference

During inference, the model predicts pseudo-measurements for non-measured feeders in B according to the methodology presented in Figure 2. In addition to conducting a cross-validation with large training and test data with feeders in A ,

it is important to estimate and monitor the convergence of the method for feeders in B when the model is deployed into production. For this purpose, it can be beneficial to compare the distribution of the features in A with the distribution of features in B to detect and correct a covariate shift. We emphasise that if a feature value for a feeder in B is significantly higher than all values of this feature in A , the model may not provide a reasonable prediction. Furthermore, monitoring software such as [41] may help to detect model performance degradation for feeders in B by training an extra model on the loss of the pseudo-measurement model.

7.1.6 | Accounting for line losses

A load or feed-in experiences different line losses depending on the spatial location in the low-voltage grid and further circumstances. The pseudo-measurement model learns the average line loss of consumption and generation in the low-voltage grid of measured feeders in set A . The feeder-specific line losses during model prediction can deviate from this learnt average line losses. However, the authors in ref. [42] report for a comparable set of low-voltage feeders that the line losses are mostly below 1.3% with a median of 0.45%. Hence, line losses are almost negligible in view of other model inaccuracies.

7.1.7 | Classification into existing literature

We cannot directly compare our metric results to other literature, because the papers leveraging grid measurements with feeder metadata, introduced in Section 2, aim to generate load profiles and not complete time-series [19, 20]. Additionally, different temporal resolutions, metrics, grid areas and grid levels make the comparison difficult. In particular, we observe in our pre-experiments that the temporal resolution and the grid level (MV/LV substation vs. LV feeder) influences the model accuracy. This is probably due to smoothing effects as stated in ref. [6].

7.2 | Peak metric design

In this section we discuss advantages and constraints of the presented peak metrics.

7.2.1 | Information gain through peak metrics

The peak metrics add valuable information to assess the model performance with respect to DSO requirements which often necessitate good peak estimations. PMag, PTime and PShape are suited to assess if a model is appropriate for use cases such as preventing overload with control operations. During the experiments, it is important to refine the peak metrics to account for measurements with many zero-crossings and feeders

with low capacity utilisation. Further refinements of the peak metric design could include adaptations to PShape which shows small deviations between feeders and models, especially for the feed-in.

7.2.2 | Combining peak metrics with all-observation metrics

Analysing peak metrics in isolation can lead to wrong assumptions when evaluating results. Consider a measurement with a base load and regular peaks. A pseudo-measurement model which estimates a base load close to the peak values of the measurements and not estimating any peaks receives a good peak metric value. However, both the base load and the peaks are estimated badly. Therefore, it is important to also include all-observation metrics in the analysis.

7.2.3 | Challenge of setting thresholds

The peak metrics require numerous thresholds for evaluation which must be specified. In particular, the peak metrics are based on daily thresholds which is meaningful because the load follows daily patterns. However, peaks occurring around midnight can lead to high errors of PMag and PTime even though the estimation is close to the peak with respect to magnitude and timing.

We also use a threshold of ± 10 kW to exclude days from the evaluation which do not exhibit a clear consumption or feed-in peak. This is, for example, the case for days with low global radiation. The threshold can also be used to evaluate only critical peaks, for example, peaks exceeding 80% of the cable capacity. Furthermore, feeders with less than 10 days with peaks exceeding ± 10 kW are excluded because these feeders can result in high outliers, for example, a PTime of 0 h when the timing of the only peak is met exactly.

Despite the problem of setting thresholds, missing thresholds can also lead to difficult interpretations of a peak metric result. Feeder F_2 performs very differently during winter and summer as shown in Figures 6 and A3. However, this is not reflected in a peak metric averaging all days regardless of the season.

8 | LIMITATIONS

In the following, we present the limitations of the dataset, the model and the approach in general.

8.1 | Data preparation and dataset

The DSO providing the data operates distribution grids mainly in rural areas. Therefore, distribution grids in urban areas are represented less.

Applying filters to the data is necessary, but the concrete thresholds are difficult to define, for instance for feeders with unrealistic high installed equipment power. Furthermore, it is only possible to detect implausible combinations of measurements and feeder metadata if measurements are available for that feeder. However, most feeders are not measured according to the problem statement in Section 3. This can lead to a bias between training and inference data sets.

Other challenges regarding the data quality comprise missing information about grid topology changes which affects the feeder metadata or the deregistration of installed equipment after decommissioning.

We do not apply anomaly removal techniques, because the distinction between anomalies and valid extreme values is ambiguous. However, extreme values are important to estimate peaks and to not deform the test data. Additionally, the aggregation to 15 min smooths short-time anomalies.

Since the metrics in Tables 4 and 5 are calculated after computing the metrics feeder-wise, feeders with long measurement periods could be underrepresented. However, by our approach we can ensure an equal representation of the distribution of metadata at feeders within the metrics.

8.2 | Model

We do not conduct an extensive hyperparameter optimisation for the models of all cross-validation folds. Hence, we do not state if XGBoost or MLP outperform each other. However, we tested a variety of hyperparameter combinations.

We do not include deep learning approaches. However, we do not expect significant improvements compared to XGBoost because tree-based models are state-of-the-art for tabular regression problems [39].

8.3 | Limitations of the approach

The presented approach requires well-maintained metadata assigned to LV feeders. This excludes the application of the approach in meshed grids where metadata contributes to several feeders. However, LV grids are operated predominantly as weakly meshed network with few loops [43].

We suspect that optimising the model regarding all measurement values with quadratic loss functions makes it hard to correctly predict peaks. Nevertheless, several adaptations are possible to mitigate this problem as stated in Section 9.

9 | CONCLUSION

The lack of active power measurements in the LV grid is an urgent problem for DSOs. In the present paper, we introduce a new approach to generate pseudo-measurements for LV feeders which are not equipped with measurement devices. The measurement estimations are based on feeder metadata

and adapt to different weather, calendar and timestamp conditions. To the best of our knowledge, there is no other publication applying and analysing the approach in the given depth.

We extensively evaluate the approach with different models in a 5-fold cross-validation on real-world data with 2323 feeders. We show that the approach can produce realistic estimations of measurements. Existing performance outliers are discussed in detail, which occur due to data quality issues, sudden peaks, missing features or stochastics in the data.

Furthermore, we give insights into the generated pseudo-measurements based on tailored peak metrics. With the help of the introduced peak metrics, we can evaluate the model performance with respect to the magnitude, timing and shape for peaks induced by consumption or feed-in. It is visible that the MAE only evaluating peak magnitudes is higher compared to the MAE over all observations.

In the future, the new approach can also be evaluated on other grid levels such as substations, other target variables such as the current and on different temporal resolutions. Additionally, other datasets with different characteristics and from different countries could be examined. Furthermore, an extensive benchmarking of other approaches to generate estimations for non-measured LV feeders could give valuable insights, for example, by including the aggregation of smart meter data. Moreover, a comparison between load forecasting and pseudo-measurements as well as using the pseudo-measurements in a state estimation can be examined in the future.

The feeder metadata is an essential part of the model features. Therefore, it is interesting in the future to evaluate the importance of individual or groups of feeder metadata for the pseudo-measurement outcome. Possible approaches for this include Explainable Artificial Intelligence (XAI) methods such as SHAP [44] or varying groups of input features.

Possible improvements of the presented approach include optimising the models with respect to the peak metrics or even estimating only peak values. Additionally, probabilistic models could give worst-case load estimates by providing quantiles and thereby estimating the aleatoric uncertainty. The epistemic uncertainty could be reduced by more diverse feeder metadata, for example, yearly billing data of households, social demographic and economic data of the underlying grid or features derived from smart meter data. Furthermore, it could be necessary to include features for market prices if the consumption adapts to market prices. Finally, post-processing pseudo-measurements could improve the pseudo-measurements with respect to specific metrics. The peak metrics in the present paper provide a basis to evaluate if model adaptations contribute to meet DSO requirements.

Moreover, it is beneficial to generate synthetic or anonymised data with similar characteristics which are not subject to any legal restrictions and can be published with the results. This would further improve the traceability of the results and enable other researchers to investigate other research questions.

AUTHOR CONTRIBUTIONS

Manuel Treutlein: Conceptualisation; investigation; methodology; software; validation; visualisation; writing - original draft. **Marc Schmidt:** Conceptualisation; Data curation; funding acquisition; project administration; resources; software; supervision; writing - review and editing. **Roman Hahn:** Conceptualisation; software; writing - review and editing. **Matthias Hertel:** Conceptualisation; supervision; writing - review and editing. **Benedikt Heidrich:** Conceptualisation; supervision; writing - review and editing. **Ralf Mikut:** Conceptualisation; supervision; writing - review and editing. **Veit Hagenmeyer:** Funding acquisition; supervision; writing - review and editing.

ACKNOWLEDGEMENTS

The authors thank Netze BW GmbH, in particular all team members of the project low voltage prognosis, for the data and infrastructure required for this study. Furthermore, we would like to thank the Helmholtz Association for the support by the Program “Energy System Design” and the Helmholtz Association's Initiative Helmholtz AI. Open Access funding enabled and organized by Projekt DEAL.








CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the present paper. Furthermore, we declare to agree to the Wiley *Best Practice Guidelines on Research Integrity and Publishing Ethics*. The present paper does not contain any studies involving humans or animals performed by any of the authors.

DATA AVAILABILITY STATEMENT

The data which is used for the experiments is internal data from the distribution system operator Netze BW. Unfortunately, it cannot be published along with this article due to corporate and privacy restrictions.

ORCID

Manuel Treutlein  <https://orcid.org/0009-0006-1071-341X>
Marc Schmidt  <https://orcid.org/0000-0002-9127-1708>
Roman Hahn  <https://orcid.org/0009-0002-2115-3711>
Matthias Hertel  <https://orcid.org/0000-0002-0814-766X>
Benedikt Heidrich  <https://orcid.org/0000-0002-1923-0848>
Ralf Mikut  <https://orcid.org/0000-0001-9100-5496>
Veit Hagenmeyer  <https://orcid.org/0000-0002-3572-9083>

REFERENCES

- Cakmak, H.K., Hagenmeyer, V.: Using open data for modeling and simulation of the all electrical society in eASiMOV. In: 2022 Open Source Modelling and Simulation of Energy Systems (OSMES), pp. 1–6. IEEE, Aachen, Germany (2022). <https://doi.org/10.1109/OSMES54027.2022.9769145>
- Haben, S., Voss, M., Holderbaum, W.: Core Concepts and Methods in Load Forecasting: With Applications in Distribution Networks. Springer International Publishing, Cham (2023). <https://doi.org/10.1007/978-3-031-27852-5>
- Deutscher Bundestag: Gesetz Über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz – EnWG) (2024). [Online]. https://www.gesetze-im-internet.de/enwg_2005/. Accessed 09 Mar 2024
- Dehghanpour, K., et al.: A survey on state estimation techniques and challenges in smart distribution systems. IEEE Trans. Smart Grid 10(2), 2312–2322 (2019). <https://doi.org/10.1109/TSG.2018.2870600>
- Bundesnetzagentur: Festlegung zur Durchführung der netzorientierten Steuerung von steuerbaren Verbrauchseinrichtungen und steuerbaren Netzanschlüssen nach § 14a EnWG (2023). [Online]. https://www.bundesnetzagentur.de/DE/Beschlusskammern/1_GZ/BK6-GZ/2022/BK6-22-300/BK6-22-300_Beschluss.html?nn=993170. Accessed 10 Feb 2025
- Haben, S., et al.: Review of low voltage load forecasting: methods, applications, and recommendations. Appl. Energy 304, 117–798 (2021). <https://doi.org/10.1016/j.apenergy.2021.117798>
- Shukla, S., Hong, T.: BigDEAL Challenge 2022: forecasting peak timing of electricity demand. IET Smart Grid 7(4), 442–459 (2024). <https://doi.org/10.1049/stg2.12162>
- Arritt, R., Dugan, R.: Comparing load estimation methods for distribution system analysis. In: 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013), pp. 1–4. Institution of Engineering and Technology, Stockholm (2013). <https://doi.org/10.1049/cp.2013.0869>
- Rousseaux, P., et al.: A new formulation of state estimation in distribution systems including demand and generation states. In: 2015 IEEE Eindhoven PowerTech, pp. 1–6. IEEE, Eindhoven, Netherlands (2015). <https://doi.org/10.1109/PTC.2015.7232471>
- Krsman, V., Tesanovic, B., Dojic, J.: Pre-processing of pseudo measurements based on AMI data for distribution system state estimation. In: Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2016), pp. 1–6. IET, Belgrade (2016). <https://doi.org/10.1049/cp.2016.1099>
- Zhao, J., et al.: Robust medium-voltage distribution system state estimation using multi-source data. In: 2020 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–5. IEEE, Washington (2020). <https://doi.org/10.1109/ISGT45199.2020.9087787>
- Rankovic, A., Saric, A.T.: Modeling of photovoltaic and wind turbine based distributed generation in state estimation. In: 2012 15th International Power Electronics and Motion Control Conference (EPE/PEMC). IEEE, Novi (2012). LS2b.2-1-LS2b.2-6. <https://doi.org/10.1109/EPEPEMC.2012.6397410>
- Morsy, B., AlSadat, M., Pozo, D.: State estimation with exogenous information for grids with large renewable penetration. In: 2020 International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), pp. 1–6. IEEE, Moscow (2020). <https://doi.org/10.1109/REEPE49198.2020.9059144>
- Proedrou, E.: A comprehensive review of residential electricity load profile models. IEEE Access 9, 12114–12133 (2021). <https://doi.org/10.1109/ACCESS.2021.3050074>
- VDEW: Repräsentative VDEW-Lastprofile. Frankfurt am Main, Germany (1999). [Online]. <https://www.bdew.de/energie/standardlastprofile-strom/>. accessed 09 03 2024
- Singh, R., Pal, B., Jabr, R.: Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement. IET Gener., Transm. Distrib. 4(1), 50–59 (2010). <https://doi.org/10.1049/iet-gtd.2009.0167>
- Chai, S., Chadney, G.: Faraday: synthetic smart meter generator for the smart grid. arXiv: 2404.04314 [cs] (2024). pre-published
- Adinolfi, F., et al.: Pseudo-measurements modeling using neural network and Fourier decomposition for distribution state estimation. In: IEEE PES Innovative Smart Grid Technologies, Europe, pp. 1–6. IEEE, Istanbul (2014). <https://doi.org/10.1109/ISGTEurope.2014.7028770>
- Salazar, M., et al.: Data driven framework for load profile generation in medium voltage networks via transfer learning. In: 2020 IEEE PES

- Innovative Smart Grid Technologies Europe (ISGT-Europe), pp. 909–913. IEEE, The Hague (2020). <https://doi.org/10.1109/ISGT-Europe47291.2020.9248753>
20. Ivasenko, O.: Netzzustandsschätzung und Leistungsvorhersage mit KI. In: *KI in Verteilnetzen*, Konstanz, Germany (2022). [Online]. <https://ai4grids-symposium.htwg-konstanz.de/de/presentationen/symposium1>. Accessed 09 Mar 2024
 21. Sanchez, R., et al.: Observability of low voltage grids: actual DSOs challenges and research questions. In: 2017 52nd International Universities Power Engineering Conference (UPEC), pp. 1–6. IEEE, Heraklion (2017). <https://doi.org/10.1109/UPEC.2017.8232008>
 22. Vitiello, S., et al.: Smart metering roll-out in Europe: where do we stand? Cost benefit analyses in the clean energy package and research trends in the green deal. *Energies* 15(7), 2340 (2022). <https://doi.org/10.3390/en15072340>
 23. Kroener, N., et al.: State-of-the-Art integration of decentralized energy management systems into the German smart meter gateway infrastructure. *Appl. Sci.* 10(11), 3665 (2020). <https://doi.org/10.3390/app10113665>
 24. Schirmer, P.A., Mporas, I.: Non-intrusive load monitoring: a review. *IEEE Trans. Smart Grid* 14(1), 769–784 (2023). <https://doi.org/10.1109/TSG.2022.3189598>
 25. Wang, S., et al.: Regional nonintrusive load monitoring for low voltage substations and distributed energy resources. *Appl. Energy* 260, 114–225 (2020). <https://doi.org/10.1016/j.apenergy.2019.114225>
 26. Gros, D., et al.: Parametrization of stochastic load profile modeling approaches for smart grid simulations. In: 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), pp. 1–6. IEEE, Torino (2017). <https://doi.org/10.1109/ISGTEurope.2017.8260180>
 27. Keller-Giessbach, D., Löwen, A.: Das Potenzial von künstlicher Intelligenz im Verteilnetz – mit geringem Kostenaufwand Netzüberwachung intelligenter machen. In: Doleski, O.D. (ed.) *Realisierung Utility 4.0 Band 1*, pp. 839–858. Springer Fachmedien Wiesbaden, Wiesbaden (2020). https://doi.org/10.1007/978-3-658-25332-5_52
 28. Vasile, P., et al.: Abschlussbericht zum NETZlabor E-Mobility-Chaussee. Netze BW GmbH, Stuttgart (2022). [Online]. <https://www.netze-bw.de/unsernetz/netzinnovationen/netzintegration-elektromobilitaet/e-mobility-chaussee>. Accessed 09 Mar 2024
 29. Turowski, M., et al.: Modeling and generating synthetic anomalies for energy and power time series. In: *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, pp. 471–484. ACM, Virtual Event (2022). <https://doi.org/10.1145/3538637.3539760>
 30. Turowski, M., et al.: Generating synthetic energy time series: a review. *Renew. Sustain. Energy Rev.* 206, 114–842 (2024). <https://doi.org/10.1016/j.rser.2024.114842>
 31. Liang, Y., et al.: Foundation models for time series analysis: a tutorial and survey. *arXiv: 2403.14735 [cs]*, 6555–6565 (2024). pre-published. <https://doi.org/10.1145/3637528.3671451>
 32. Heidrich, B., et al.: Using conditional Invertible Neural Networks to perform mid-term peak load forecasting. *IET Smart Grid* 7(4), 460–472 (2024). <https://doi.org/10.1049/stg2.12169>
 33. Treutlein, M.: Erzeugung von Pseudomessdaten für Ortsnetzstationen mittels diskriminativer und generativer Modelle. M.S. thesis. Netze BW GmbH, Stuttgart (2023). [Online]. Available: intern
 34. Carcangiu, S., et al.: Forecasting-aided monitoring for the distribution system state estimation. *Complexity* 2020, 1–15 (2020). <https://doi.org/10.1155/2020/4281219>
 35. Zufferey, T., Hug, G.: Impact of data availability and pseudo-measurement synthesis on distribution system state estimation. *IET Smart Grid* 4(1), 29–44 (2021). <https://doi.org/10.1049/stg2.12004>
 36. Reinert, D., et al.: DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System. German Meteorological Service, Offenbach am Main, Germany (2024). [Online]. https://dwd.de/DWD/forschung/nw/fepub/icon_database_main.pdf. Accessed 09 Mar 2024
 37. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011). [Online]. <https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Accessed 09 Mar 2024
 38. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. ACM, San Francisco (2016). <https://doi.org/10.1145/2939672.2939785>
 39. Grinsztajn, L., Oyallon, E., Varoquaux, G.: Why do tree-based models still outperform deep learning on tabular data? *arXiv: 2207.08815[cs, stat]* (2022). pre-published
 40. Gorishniy, Y., et al.: Revisiting deep learning models for tabular data. In: Ranzato, M., et al. (eds.) *Advances in Neural Information Processing Systems*. vol. 34, pp. 18932–18943. Virtual Event: Curran Associates, Inc., (2021). [Online]. https://proceedings.neurips.cc/paper_files/paper/2021. Accessed 09 Mar 2024
 41. NannyML (release 0.12.1), <https://github.com/NannyML/nannyml>, NannyML, Belgium, OHL, (2023)
 42. Ma, C., et al.: Evaluation of energy losses in low voltage distribution grids with high penetration of distributed generation. *Appl. Energy* 256, 113–907 (2019). <https://doi.org/10.1016/j.apenergy.2019.113907>
 43. Lourenço, E.M., Bosco Augusto London Junior, J. (eds.) *Power Distribution System State Estimation*. IET, Stevenage, England (2022). <https://doi.org/10.1049/PBPO183E>
 44. Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions. *arXiv: 1705.07874 [cs, stat]* (2017). pre-published

How to cite this article: Treutlein, M., et al.: Generating peak-aware pseudo-measurements for low-voltage feeders using metadata of distribution system operators. *IET Smart Grid*. e12210 (2025). <https://doi.org/10.1049/stg2.12210>

APPENDIX A

Model	Hyperparameter	Value	Model	Hyperparameter	Value
XGBoost	n_estimators	1500	MLP	hidden_layer_sizes	(20)
	eta	0.05		activation	\tanh
	subsample	0.7		random_state	2
	colsample_bylevel	0.5		early_stopping	True
	early_stopping_rounds	30		validation_fraction	0.125

TABLE A1 Hyperparameters used for the models XGBoost and MLP (without default values).

TABLE A2 Mean and standard deviation of the 5 folds from the cross-validation for all metrics and models.

Metric	XGBoost	MLP	LR
MAE [kW]	4.73 ± 0.21	4.85 ± 0.24	6.07 ± 0.29
MAE _{norm}	0.13 ± 0.02	0.13 ± 0.02	0.17 ± 0.02
RMSE [kW]	6.19 ± 0.22	6.33 ± 0.26	7.85 ± 0.49
PMag ^C [kW]	11.68 ± 0.56	11.55 ± 0.81	14.07 ± 0.72
PMag ^F [kW]	13.24 ± 0.87	12.67 ± 1.23	21.43 ± 1.66
PTime ^C [h]	5.09 ± 0.27	5.17 ± 0.41	8.02 ± 0.19
PTime ^F [h]	1.17 ± 0.04	1.26 ± 0.03	1.13 ± 0.02
Shape ^C	0.35 ± 0.00	0.34 ± 0.00	0.38 ± 0.00
Shape ^F	0.29 ± 0.00	0.29 ± 0.00	0.29 ± 0.00

TABLE A3 Metric overview for the LV feeders which are shown in detail in Figure 6, Figure A2 and Figure A3. The metric is calculated for the whole measurement period of a feeder. All feeders are part of the test data in the XGBoost fold 1. The peak metrics are differentiated between consumption (C) and feed-in (F).

Feeder	MAE	MAE _{norm}	RMSE	PMag ^C	PMag ^F	PTime ^C	PTime ^F	PShape ^C	PShape ^F
Average XGBoost fold 1	4.57	0.10	6.22	11.78	12.58	5.11	1.18	0.35	0.30
F ₁ (Figures 6 and A2)	4.62	0.04	6.75	7.14	18.46	7.38	1.10	0.32	0.29
F ₂ (Figures 6, A2, A3)	4.92	0.14	6.15	10.28	–	6.56	–	0.38	–
F ₃ (Figure A2)	6.39	0.06	8.62	5.80	11.17	1.66	1.60	0.28	0.30
F ₄ (Figure A2)	6.47	0.19	7.24	6.34	–	2.49	–	0.32	–
F ₅ (Figure A2)	12.41	0.12	14.96	30.78	19.34	3.75	3.26	0.34	0.33

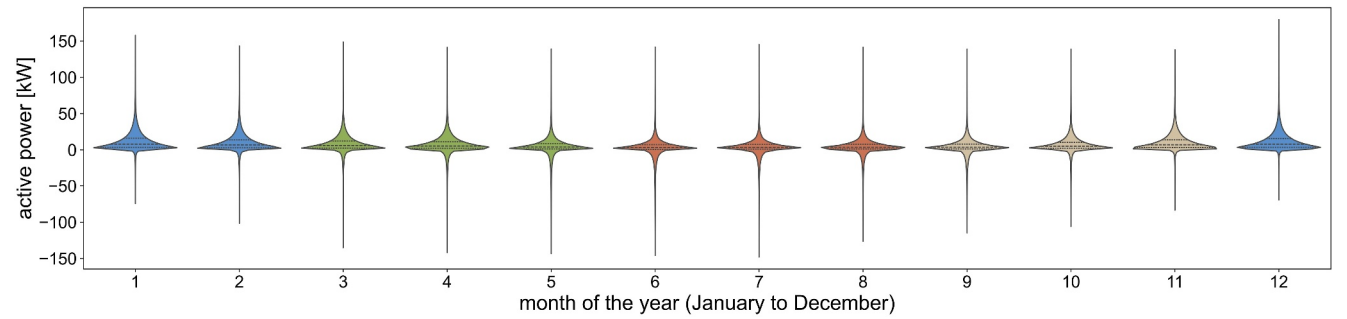


FIGURE A1 Monthly distributions of measurements from all feeders aggregated to 15 min resolution. The colours indicate spring (green), summer (red), autumn (brown) and winter (blue). The dotted lines denote the 25% quantile, median and 75% quantile.

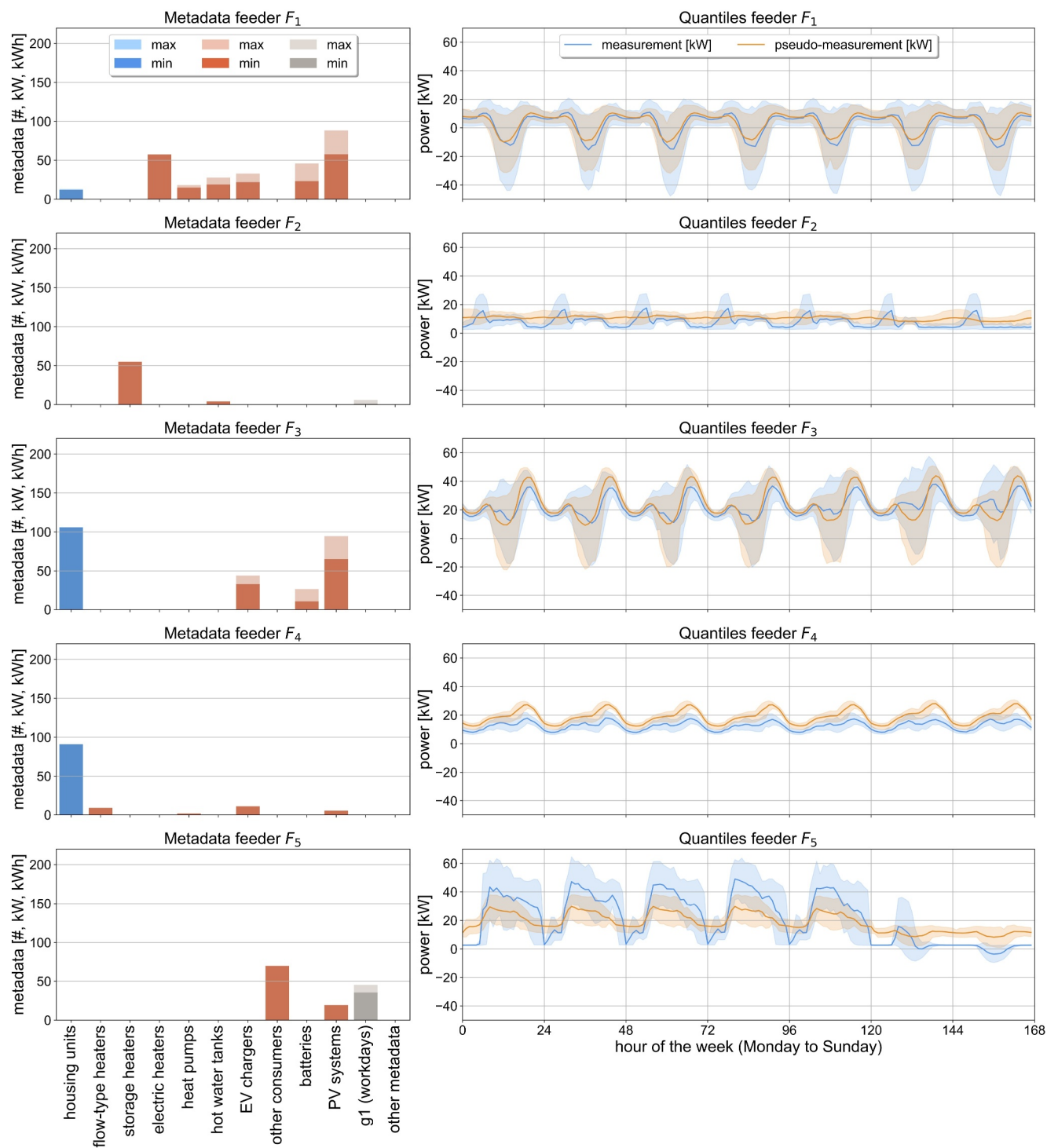


FIGURE A2 Selected feeders F_1 - F_5 with feeder metadata, measurements and pseudo-measurements from the test data of XGBoost fold 1. Metadata with unit count (#) in blue, unit kW in red and unit kWh in grey. The label min is the lowest value of the metadata and max the highest during the evaluation period from the end of 2022 until March 2024. The time series of measurements and pseudo-measurements are condensed to a weekly profile with median, 10% and 90% quantile for visualisation.

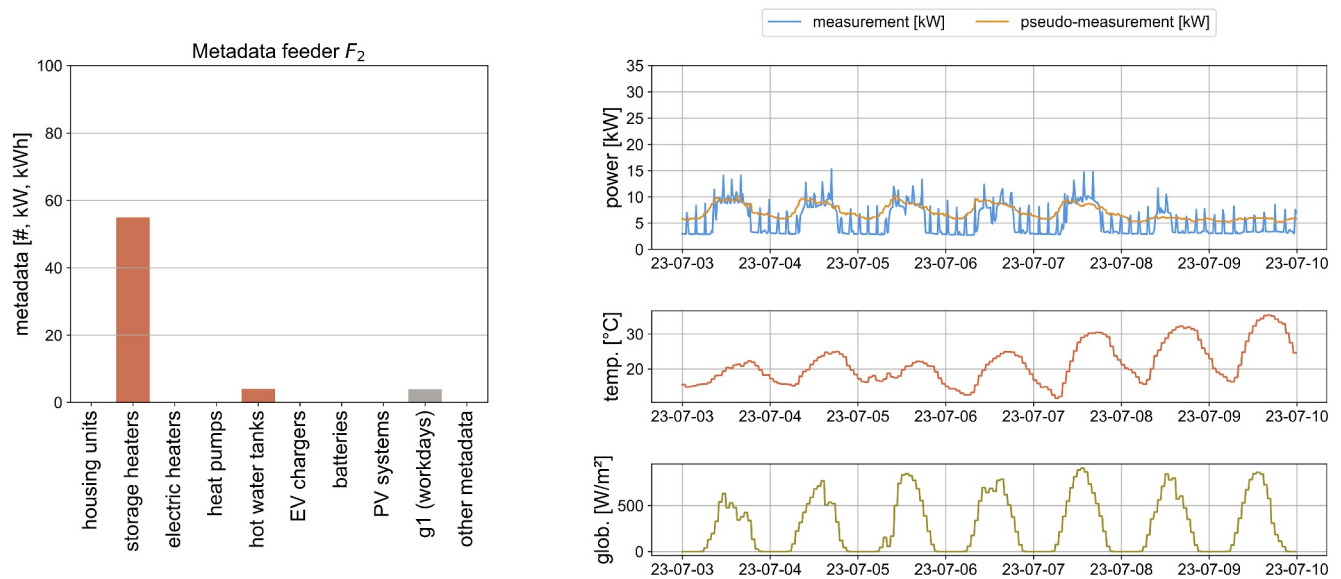


FIGURE A3 Feeder F_2 with the feeder metadata on the left side and the measurement and pseudo-measurement on the right side together with the temperature and global radiation of one week in July 2023 (Monday - Sunday). The figure complements the Figure 6 where the feeder F_2 is shown for a week in winter. Metadata with unit kW in red and unit kWh in grey. The metric values for the feeders are given in Table A3.