(Position Paper) Characterizing the Behavior of Small Producers in Smart Grids

A Data Sanity Analysis

Stefan, Maria; Gutierrez Lopez, Jose Manuel; Barlet-Ros, Pere; Prieto, Eduardo ; Gomis, Oriol ; Olsen, Rasmus Løvenstein

Published in:
Procedia Computer Science

DOI (link to publication from Publisher):
10.1016/j.procs.2020.02.264

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 168 (2020) 224–231

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme:
Leveraging AI and Machine Learning for Societal Challenges, CAS 2019

(Position paper) Characterizing the Behavior of Small Producers in Smart Grids

A Data Sanity Analysis

Maria Stefan[a]*, Jose Gutierrez[a,d], Pere Barlet[b], Eduardo Prieto[c], Oriol Gomis[c] and Rasmus L. Olsen[a]

[a]*Department of Electronic Systems, Aalborg University, Denmark*
[b]*Department of Computer Architecture, Universitat Politecnica de Catalunya, Spain*
[c]*Department of Electrical Engineering, Universitat Politecnica de Catalunya, Spain*
[d]*2operate, Denmark*

Abstract

Renewable energy production throughout low-voltage grids has gradually increased in electrical distribution systems, therefore introducing small energy producers - prosumers. This paradigm challenges the traditional unidirectional energy distribution flow to include disperse power production from renewables. To understand how energy usage can be optimized in the dynamic electrical grid, it is important to understand the behavior of prosumers and their impact on the grid's operational procedures.
The main focus of this study is to investigate how grid operators can obtain an automatic data-driven system for the low-voltage electrical grid management, by analyzing the available grid topology and time-series consumption data from a real-life test area.
The aim is to argue for how different consumer profiles, clustering and prediction methods contribute to the grid-related operations. Ultimately, this work is intended for future research directions that can contribute to improving the trade-off between systematic and scalable data models and software computational challenges.

Keywords: energy optimization; data sanity; user modeling and applications; clustering; prediction methods

* Corresponding author. Tel.: +45 52227605.
E-mail address: marst@es.aau.dk

10.1016/j.procs.2020.02.264

1. Introduction

The installation of diverse industrial and domestic renewable and green energy generators and the subsequent decentralized architecture guide the power grid progress towards a smart grid strategy. This is a direct result of the commitment to 100% renewable energy production in Denmark by 2050 as part of a national climate change mitigation plan, which is influenced by international interests (i.e. EU's clean energy package and regulations [1]). Currently, the power grid supports up to 43% power generation from wind turbines [2] [3] and can operate with up to 50% power supplied by combined heat and power (CHP) plants [4]. Consequently, as the penetration rate of renewable energy sources (RES) intensifies, the low-voltage power grid becomes active as consumers evolve into prosumers, influencing daily grid operation and management for the distributed system operators (DSOs).

Currently, the incoming metering data is used only for billing purposes [5]. The increasing complexity of the grid will require an automated solution in which different anomalies can be detected with minimum delay time. It is expected that with the high penetration of RES there will be an increase in reported issues by the consumers [6]. Measurement data from the Advanced Metering Infrastructures (AMI) can be utilized in this case to characterize the consumption patterns and predict future possible issues. At the same time, the aim is to make use of the available data in order to prepare for a scenario with 100% renewable energy.

This research topic aims to utilize the available billing data measurements in order to understand the con- sumers'/prosumers' behavior. Moreover, the objective is to propose solutions to some of the foreseeable problems, by integrating electrical engineering knowledge into a computer software solution. In this way, the contribution comes from analyzing the quality of the available electrical grid data. It is shown that it would be useful for the DSOs to consider this data for analysis, in order to automate their most frequent grid management procedures.

2. Machine learning for electrical grid data - Related work

Various machine learning techniques have been previously used in the power grid domain to provide the DSOs with the right tools for grid planning, monitoring and forecasting. Understanding the energy behavior at the low-voltage grid can be done by clustering households by specific attributes, defined by some analytic techniques:

- Extracting the electricity demands according to different times of the day, season and weekdays;
- Classification according to the chosen attributes (from low to large variability);
- Reliability testing: sample robustness assessed using a bootstrapping method as in [7].

For forecasting purposes: Electricity short term load forecasting (STLF) applied to historical customer data is addressed in [8], by means of data cleaning (smooth out irregular electricity consumption patterns, such as holidays), error correction methods and ANN (Artificial Neural Networks) with historical weather data. Demand is very random over short periods of time, day-to-day profiles, therefore a demand forecast model is needed in the management control system, as explained in [9]. The model was obtained through data pre-processing, correlation clustering and discrete classification NN (Neural Network).

For monitoring purposes: The study in [10] is used to obtain forecast density estimation by searching for analogs in the historical data. It can be utilized for in-memory computing in distributed systems, by saving computational time for a high number of smart meters and by providing scalability.

For planning purposes: Short term state forecasting and operation is addressed in [11], in the form of: A) Optimized distributed energy resources allocation, based on the location of energy resources in the distribution network. The amount of required load adjustment is minimized to match with the network constraints. This service can be utilized for energy balancing. And B) Voltage estimation using historical smart meter data and estimates of the net demand. Probabilistic estimates of low-voltage profiles are obtained, assuming that the smart meters cannot measure voltage or power quality.

The clustering methods for analyzing time-series data streams also provide insight into the customers' privacy, by identifying specific behaviors [12].

3. A data sanity study for low-voltage electrical grids

3.1. Data system and data flow

The information exchange in the electrical grid corresponding to the RemoteGRID project [5] [13] is depicted in Fig. 1. The three actors defined in Fig. 1. show the relation between the DSO, AMI provider and IT distribution.

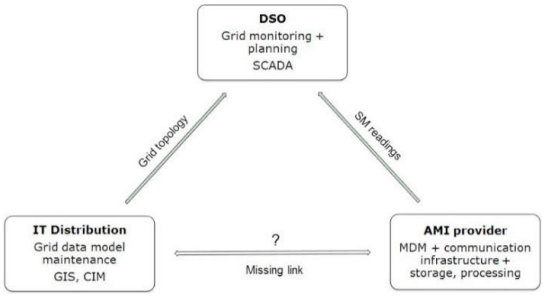


Fig. 1. Information exchange among the meter provider, IT distribution and the DSO.

The AMI provider [14] is in charge of the Meter Data Management (MDM) module for data storage and processing and of the AMI communication infrastructure. It provides the DSO with smart meter (SM) readings in the form of time-series data. The DSO [15] uses these historical readings to check for potential anomalies, (monitoring) and for grid planning. Simultaneously, the DSO is provided with grid topology information which is managed via a SCADA system (Supervisory Control and Data Acquisition). The IT distribution company [16] is in charge of data integration between GIS and SCADA, as defined by the CIM standard. The available GIS data has to be regularly modeled and converted to CIM to be correctly imported to the SCADA system, according to updates in the topology (i.e. new customers with PVs and wind turbines).

The flow of the present-day data system lies in the lack of interaction between the IT distribution and the AMI provider. This missing link may result in data inaccuracy, posing operational challenges to the DSOs. For example, the quality of the GIS-modeled low-voltage network may not be sufficient due to missing customer-related data. In this case, the DSO relies on knowledge of the number of customers connected to the transformers in the specific secondary substations (medium-voltage), instead of the low-voltage grid topology information.

3.2. Data types

The two data types received by the DSO from the AMI provider and from the IT distributor are described as:

- **GIS data:** The grid topology comes in the form of geographic information, containing the connectivity information between the medium and the low-voltage part of the grid. This includes nodes (secondary substations, cable boxes and consumers) and their interconnecting cables. An example of the topology information is provided by Fig. 2., where substations are represented by the red triangles, cable boxes by the blue squares and consumers by green dots. The red dotted lines show the AC connections among the secondary substations. The low-voltage grid connections are marked by the different colored lines, each color depicting the different groups of consumers fed by each of the substations.

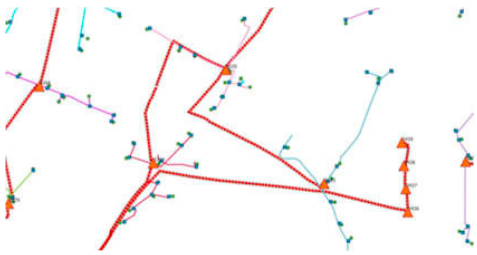


Fig. 2. Medium and low-voltage grid topology sample.

- **Time series data:** Active and reactive energy measurements are provided for a period of one year, with a granularity of 15 minutes, which is defined by the current metering infrastructure.

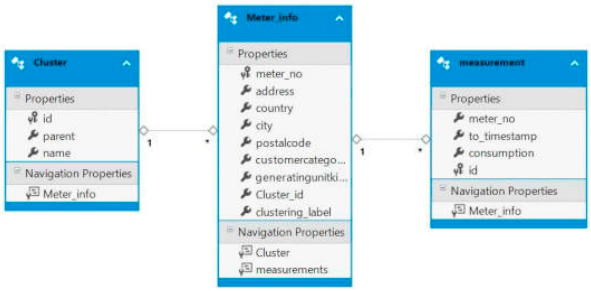


Fig. 3. Data model for the time-series measurements.

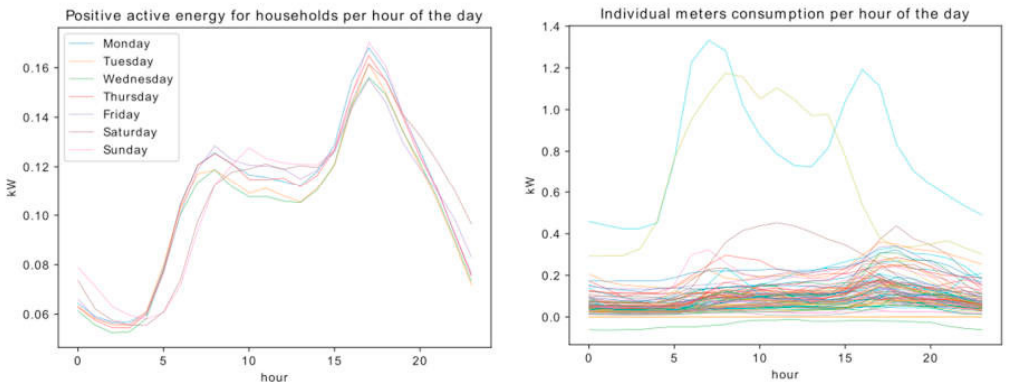
The data model for the time-series measurements is shown in Fig. 3., comprising of three descriptive tables. The *measurements* table contains the meter ID ("meter no"), measurement timestamp and consumption values (active positive energy). The meter ID is used as foreign key element for the *Meter_info* table, which contains general information about the individual metering points: address, generating unit kind (solar cells, windmills, other) and customer category name (household, company, school, other). The meter no field is used as foreign key for the *Cluster* table, which is used to store information about consumption classification.

The meter ID field (obtained from the distribution company involved in the study [9]) was used to link the two data types via address geocoding, making it possible to perform statistical analysis on the time-series measurements, based on the meters' geographic information.

3.3. User profile analysis - labeled data

Two types of labelled customers have been identified - households and companies, which will be used as starting point in the analysis. Some of the companies are labelled with PVs, but there is no information about RES at household level.

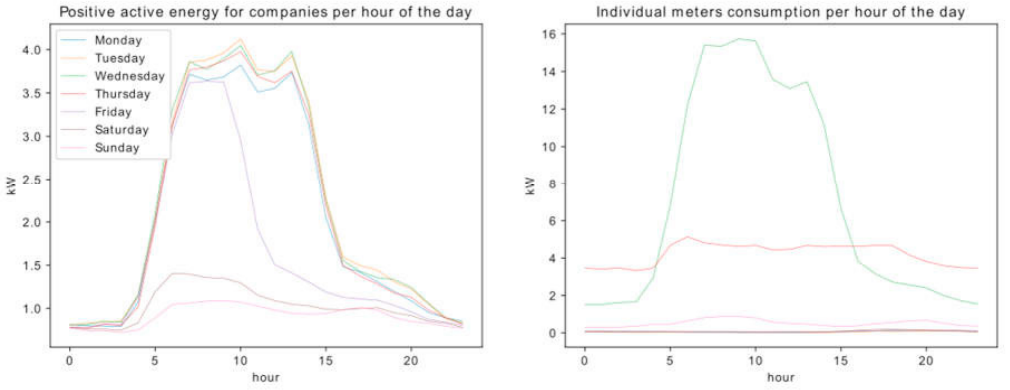
- **Household profiles**: The plot is Fig. 4. shows the household consumption statistics (in kWh). Subfig. 4.(a) depicts the average consumption of all households for the whole year per hour of the day, for each day of the week. All data is taken into account for all seasons of the year, without filtering out holidays. This is done in order to obtain a general overview over the households' consumption trends. It can be seen that the trend is as expected, with the highest consumption peaks in the morning (around 6-7 AM in the weekdays and later in the weekends) and in the afternoon (5-7 PM). Subfig. 4.(b) shows the individual household consumption per hour of the day. The data is again averaged for the whole year and for all the days of the week. As it can be noticed, there are two households whose consumption trends are different than the average.



(a) Households' average consumption patterns per hour. (b) Individual households' consumption patterns per hour.

Fig. 4. Positive active energy plots for labelled households

- **Company profiles:** Similarly, the company consumption profiles (kWh) are presented in Fig. 5. The average consumption per year for every weekday is shown in Subfig. 5 (a). The trend represents a typical working week in Denmark, starting early in the morning (6-8 AM) and ending at about 4 PM in the weekdays and earlier on Friday. Also, the lowest consumption is registered in the weekends and after working hours. Individual company consumption per day is depicted in Subfig. 5.(b), averaged over one year. Two companies seem to issue different trends in their patterns other than the rest, which can also be extracted from the statistical values.



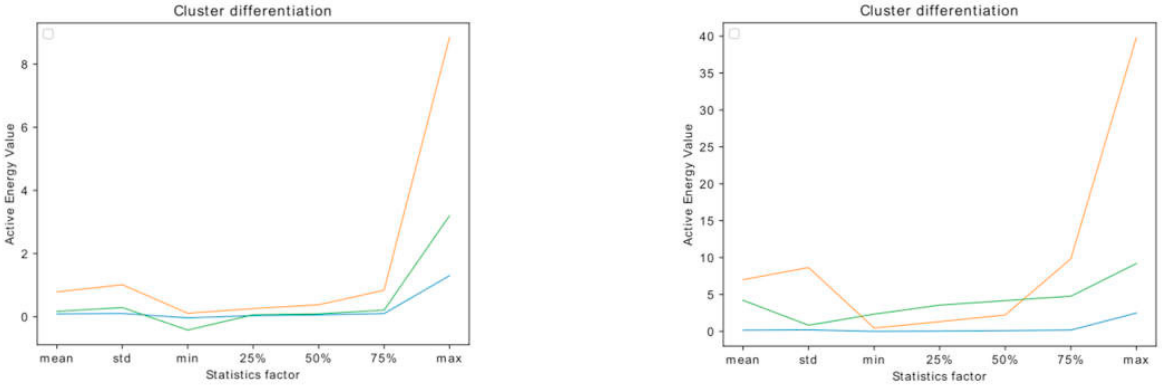
(a) Companies' average consumption patterns per hour. (b) Individual companies' consumption patterns per hour.

Fig. 5. Positive active energy plots for labelled companies

3.4. Customer classification

It can be concluded from the previous statistical analysis that there is some variation in the individual consumption patterns for households and companies. In order to help anticipating trends for the different metering points, with the purpose of detecting whether data is missing or erroneous, an automatic clustering method is applied for the two labeled data sets. This is done using the positive active energy values, obtaining three clusters per data set. The results are presented in Fig. 6., for households (Subfig. 6.(a)) and companies (Subfig. 6.(b)).

The results still show some variance in the data, particularly in the case of companies. As a result of automatic clustering, all companies with PVs have been assigned to same cluster, which is as desired. However, the implications of individual consumer behavior can be depicted from the large variance values obtained in Subfig. 6.(b). Therefore, in a data-driven software solution, an automatic anomaly detection would not be possible in such a case, meaning that other data analytics methods should be applied for this data set.



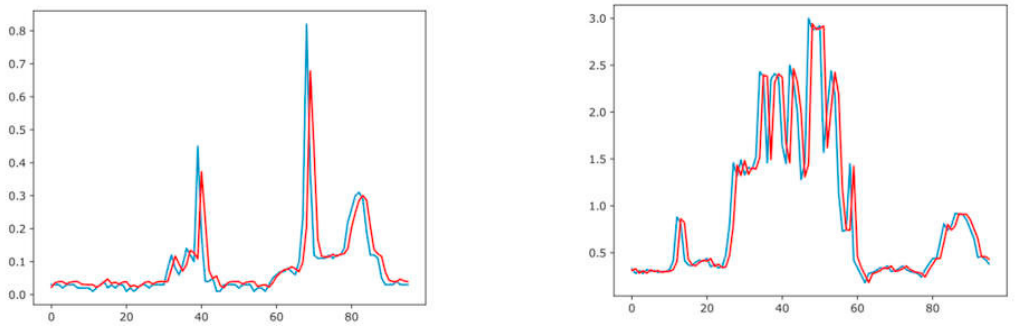
(a) Cluster differentiation for households. (b) Cluster differentiation for companies.

Fig. 6. Clustering method based on active energy consumption values for labelled households and companies

3.5. Customer behavior prediction

The above-mentioned clustering technique was utilized in order to classify the low-voltage grid customers into categories defined by their energy consumption patterns. Based on this classification, a step forward is taken in the analysis towards forecasting models. A basic ARIMA model is used to illustrate the predicted consumption patterns for one of the clusters obtained from the household labels and for the cluster containing the companies with PVs.

The plots in Fig. 7. represent the consumption values in kWh per number of samples (96 samples correspond to one day) for the two chosen clusters. It can be noticed from the plots that the predictions (red curves) follow the household and company profiles, resulting in MSE values of 0.009 and 0.141, respectively. The low error values add up to the potential of using prediction models based on clustering, however, in Subfig. 7.(a) and 7.(b) the prediction curve is shifted from the actual measurements (blue curve), as a result of the ARIMA model fitting.



(a) ARIMA prediction for a cluster containing households; MSE = 0.009.

(b) ARIMA prediction for the cluster containing the companies with PVs; MSE = 0.141.

Fig. 7. ARIMA predictions based on clustering for households and companies

4. Discussion

The data analysis presented in Section 3 had the purpose of exploring the potential uses of the available consumption data from the low-voltage grid. The analysis of the time-series data was based on statistical results, making it possible to perform consumer classification/clustering out of the available active power measurements. From this, it can be concluded that the automatic clustering can be utilized as data pre-processing method, due to its ability to correctly place the six different individual user consumption patterns into six corresponding clusters.

The results obtained from clustering still depict variance in the data, due to the subjective behavior of the small producers (consumers with PVs). Further analysis was performed by using a simple ARIMA model for behavior prediction. It can be noticed than even if the prediction follows the consumption patterns, ARIMA is not an accurate model in this case, as the predicted values are just a shifted version of the actual measurements. The model could be improved by taking into account seasonality in the data and/or by introducing weather dependencies in the model. For example, in the case of consumers with PVs, a potential parameter of influence is the solar irradiation.

The data analysis study brings out other possibilities for the DSOs for using the available data, than only for billing calculations. Understanding the data is essential for understanding the behavior of the grid's residential consumers, which is rather subjective. This study is important for the DSOs when taking into consideration future electrical grids consisting of 100% renewable, due to challenges unaccounted for in the traditional low-voltage electrical grids:

- *Mobile prosumers* - it is expected that with the increasing proliferation of electrical vehicles (EVs), the amount of mobile users will also increase. Monitoring the users will then become even more challenging due to their mobility and their subjective behavior, with more inconsistency in the data. The resulting distribution grid is anticipated to develop a recurrent number of anomalies and imbalance in the distributed power, which can be addressed by some of the analytical methods presented in Table 1.
- *Scalability* - the amount and diversity in the data incoming from the different distribution energy resources calls for a scalable and flexible data analysis solution. This scalability may also refer to a collective group of operators (heat, water, transmission system operators) who need to use data for similar purposes as the DSOs.

- *Prosumers' privacy* - with a more accurate insight into the users' electricity consumption and generation, the privacy issue evolves into being more sensitive. The trade-off lies between how much knowledge is needed to provide the required and stable electricity supply and the barrier towards accessing sensitive user data. One exception for breaking the privacy rule is the case of customers who are suspected of fraud.

Given these challenges, the aforementioned study can bring contributions to some of the DSOs daily operations by customer profiling, clustering and predictions. The different concerns regarding accuracy in the data are presented in Table 1, along with the corresponding analytical methods that can help overcome them.

Table 1. Contributions to grid operations brought by the analytic methods.

DSO Operation	Data accuracy concerns	Analytic methods
Anomaly detection	• missing/inaccurate data due to model inconsistency between the time-series and the GIS information • customers suspected of fraud (i.e. stealing energy) • faults in the grid, possible cable faults or power outages	Profiling and Prediction
Power balancing	• unexpected change of pattern for a group of customers not necessarily belonging to the same substation	Clustering
Planning	• necessary grid reinforcements due to detected anomalies • re-routing of information in the grid as part of future grid planning and optimization	Clustering and Prediction
Monitoring	• keeping track of the specific consumers who are more prone to report anomalies	Profiling and Clustering

These methods are useful as a data sanity check-up in the different situations where the available data is not labelled, missing or inaccurate. The particular lifestyle of the low-voltage grid consumers can nonetheless be deducted even after profiling and clustering, due to the high variance in the data. This issue can be eliminated by performing a more refined classification, taking into account data seasonality.

The data sanity study was performed using only active energy measurements (consumption), though the developing AMI networks are capable of collecting more varied types of parameters, such as voltage and current traces. These values, combined with knowledge of the users' consumption behavior, open up for the possibility of performing more accurate data analysis, for example for anomaly detection.

The requirements for the future smart grids imply scalable computational solutions for automatic anomaly detection, real- time grid monitoring, power balancing and planning. The computational power in an automatic data-driven management system is challenged by the data variety, volume and granularity, particularly when trying to adapt and optimize the existing DSOs' operational system to real-time conditions.

5. Conclusion

This study underlines the need for efficient data-driven solutions in the low-voltage electrical grid operation, as the traditional grids evolve into smart grids. The data analysis presented in this work is meant to demonstrate how basic statistical analysis can bring a contribution towards the challenges imposed by new grid operating conditions and use cases which arise with the proliferation of smart grids. In this sense, predictions can be used for scenarios with new areas and entities in the low-voltage grid, in order to anticipate any operational constraints. The study also shows that low-voltage grid consumers can be characterized and classified by their consumption patterns in order to facilitate some of the basic grid operations, such as anomaly detection, power balancing, planning and monitoring.

It was found that due to the diversity in the users' consumption patterns, the active energy alone is not enough for designing an automatic information-based management system. Additionally, scalable solutions depending on the amount and variety of data require more information in the form of varied AMI parameters, weather-related variables or other machine learning techniques.

Future research directions should test and take into consideration a more scalable solution for real-time data management operations in electricity grids, all the while accommodating for the imminent computational issues that come with the scalability.

Acknowledgement

This work is financially supported by the Danish project RemoteGRID, which is a ForskEL program under Energinet.dk with grant agreement no. 2016-1-12399.

References

[1] D. E. Association, Smart grid in denmark 2.0 (2016). URL https://www.usef.energy/app/uploads/2016/12/Smart-Grid-in-Denmark-2.0-2.pdf
[2] M. H. Gottlieb, Danmark sætter ny rekord i vind (2018). URL https://www.danskenergi.dk/nyheder/danmark-saetter-ny-rekord-vind
[3] D. E. Agency, Månedlig og årlig energistatistik. URL https://ens.dk/service/statistik-data-noegletal-og-kort/maanedlig-og-aarlig-energistatistik
[4] D. E. Agency, Statistik, data, nøgletal og kort. URL https://ens.dk/service/statistik-data-noegletal-og-kort
[5] R. Sanchez, F. Iov, M. Kemal, M. Stefan, R. Olsen, Observability of low voltage grids: Actual dsos challenges and research questions, in: 2017 52nd International Universities Power Engineering Conference (UPEC), 2017, pp. 1–6. doi:10.1109/UPEC.2017.8232008.
[6] M. Stefan, J. G. Lopez, R. L. Olsen, Exploring the potential of modern advanced metering infrastructure in low-voltage grid monitoring systems, 2018 IEEE International Conference on Big Data (Big Data) (2018) 3543–3548.
[7] S. Haben, C. Singleton, P. Grindrod, Analysis and clustering of residential customers energy behavioral demand using smart meter data, IEEE Transactions on Smart Grid 7 (1) (2016) 136–144. doi:10.1109/TSG. 2015.2409786.
[8] T. Boßmann, J. Schleich, R. Schurk, Unravelling load patterns of residential end-uses from smart meter data, 2015.
[9] C. Bennett, R. Stewart, J. Lu, Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system, 2014.
[10] M. Reis, A. Garcia, R. J. Bessa, A scalable load forecasting system for low voltage grids, in: 2017 IEEE Manchester PowerTech, 2017, pp. 1–6. doi:10.1109/PTC.2017.7980936.
[11] B. P. Hayes, M. Prodanovic, State forecasting and operational planning for distribution network energy management systems, IEEE Transactions on Smart Grid 7 (2) (2016) 1002–1011. doi:10.1109/TSG.2015. 2489700.
[12] V. Ford, A. Siraj, Clustering of smart meter data for disaggregation, in: 2013 IEEE Global Conference on Signal and Information Processing, 2013, pp. 507–510. doi:10.1109/GlobalSIP.2013.6736926.
[13] R. L. Olsen, Remotegrid - project description. URL https://www.en.remotegrid.dk/projectdescription/
[14] Kamstrup, Intelligent solutions for electricity utilities. URL https://www.kamstrup.com/en-en/electricity-solutions
[15] T.-M. E. E. A/S, Beskrivelse af forskningsprojekter i elnettet. URL https://www.tme-elnet.dk/om-os/forskningsprojekter
[16] DAX, Network model management (nmm). URL https://dax.dk/nmm-dansk-introduktion/