Automation of Smart Grid operations through spatio-temporal data-driven systems

Stefan, Maria

Publication date:
2019

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

AUTOMATION OF SMART GRID OPERATIONS THROUGH SPATIO-TEMPORAL DATA-DRIVEN SYSTEMS

BY
MARIA STEFAN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Automation of Smart Grid operations through spatio-temporal data-driven systems

Ph.D. Dissertation

Maria Stefan

Dissertation submitted: May 29, 2019

Dissertation submitted:	May 29, 2019
PhD supervisor:	Assoc. Prof. Rasmus Løvenstein Olsen Department of Electronic Systems Aalborg University
Assistant PhD supervisor:	Assoc. Prof. Jose Manuel Gutierrez Lopez Department of Electronic Systems Aalborg University
PhD committee:	Associate Professor Tatiana Kozlova Madsen (chair.) Aalborg University Professor Josep Solé-Pareta Universitat Politècnica de Catalunya Dr. Piotr Kiedrowski University of Science and Technology in Bydgoszcz
PhD Series:	Technical Faculty of IT and Design, Aalborg University
Department:	Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-445-4

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

Curriculum Vitae

Maria Stefan

Maria Stefan received her B.Sc. E.E with a specialization in Electrical, Electronics and Communications Engineering from the Polytechnic University of Bucharest in 2013. In 2015 she received her M.Sc. E.E. in Wireless Communication Systems from Aalborg University. Her M.Sc. thesis was the result of research conducted with Nokia Solutions and Networks Denmark. She was employed for a few months after completing the M. Sc. studies as research assistant at Aalborg University. Since 2016 she has been employed as a PhD Fellow in the Wireless Communication Networks section (WCN) in the Department of Electronic Systems at Aalborg University. In 2018-2019 she visited Universitat Politècnica de Catalunya in Barcelona, Spain as internship trainee, collaborating on the topic of data analysis and machine learning. The focus of her current research is on low-voltage electrical grids data analysis, processing and visualization, based on user experience studies.

Abstract

Traditional electricity grids are currently undergoing a transformation towards distributed generation, changing the state of the art operational processes for grid monitoring and maintenance. As Danish incentives for green energy production are being laid out, planning to have 100% renewable energy production by year 2050, consumers have begun to install renewable energy resources (RES) in the form of PVs, small wind turbines, heat pumps and electrical vehicles. The typical consumers become small producers (so-called prosumers), producing a bi-directional power flow. The face of the low-voltage electrical grid is therefore changing at a rapid pace, which poses operational challenges to the Distributed System Operators (DSOs) in terms of grid monitoring and maintenance.

Electrical grid operation is furthermore influenced by the deployment of Advanced Metering Infrastructures (AMI), consisting of a large amount of interconnected sensors/consumers. Modern AMI are capable of delivering many more measured parameters compared with traditional metering infrastructures, where the data is currently used only for billing. AMI opens the possibility to utilize the available information for more efficient grid monitoring, planning and can even be used for prediction and event-detection purposes. Novel data-driven analysis techniques are therefore required to explore the new AMI parameters, bringing the electrical grids research field towards digitalization.

The large and varied amount of data conveyed by the AMI has recently been referred to as Big Data, both in industry and in the research fields. This definition is furthermore enhanced by the modern communication infrastructures, which make it possible to stream the data from the AMI with a much finer granularity, known as real-time data.

The aim of this PhD study is to investigate how AMI data can contribute to a more efficient grid operation for the DSOs, by means of processing, analytics and visualization techniques. The conducted research has been based on a real electrical grid case scenario in collaboration with a Danish DSO from Thisted, located in the north-western part of the country. The focus has been on designing and implementing a visualization system based on the DSOs' needs and requirements. At the same time, the available geographic and time-series data was used to perform data accuracy studies and to propose a potential analytics platform for the DSOs.

User experience studies (UX) have been an important part of the work, especially for designing a simple and effective visual overview over the low-voltage electrical grid. In these studies, the users (DSOs) took part in on-site interviews and therefore helped shaping the user interface for the visualization prototype, which can be utilized for monitoring, planning and predictions. The contribution consists of enhancing the automation of the consumer level grid operations, by designing and implementing a decision support information system. Furthermore, the use of geographic information systems (GIS) contributed to spatial and situation awareness, especially relevant in a human-dependent operational environment.

Additionally, the available time-series measurements and GIS grid topology have been part of a study concerning the validity and integrity of data exchanged in the electrical grid. It has been found that due to the lack of a fully data-integrated system there are often inaccuracies in the data exchanged between the different parties, leading to erroneous use of information for the different operations. To provide the DSOs with smart functionalities, consumer behavior studies have been conducted. Based on their results, a classification of the low-voltage grid consumers has been proposed according to their energy consumption. It was shown that the created clusters are useful for grid planning even in the case of missing information, as well as for predicting how a certain customer might behave based on its profile.

Finally, the outcome of this work involves the optimization of the DSOs daily workflows by system redesign and minimizing the operating expenses (OpEx) by integrating smart analytical methods. The conducted research proves that even simple statistics and machine learning methods can bring intelligence to current power systems, helping with automatic anomaly detection and data accuracy diagnosis. Eventually, this together with other current as well as future research will contribute to the development of so-called Smart Grids.

Resumé

Det danske elnet undergår i denne tid en forandring mod en større grad af distribueret energigenerering, hvilket betyder nye procedurer for blandt andet overvågning og planlægning af nettet. Da der fra politsk side er et ønske om at have 100% produktion af vedvarende energi i 2050, er forbrugere allerede nu i gang med at installere forskellige vedvarende energiressourcer så som solceller, små vindmøller, varmepumper og elektriske køretøjer. Traditionelle forbrugere bliver dermed til småskala producenter ("prosumers"), der introducerer et tovejs flow af strøm. Lavspændingsnetværkets struktur ændrer sig derfor hurtigt og resultatet afspejles i operationelle udfordringer hos de energi operatører (DSO) der kontrollerer netværkets overvågning og vedligeholdelse.

Driften af elnettet er desuden påvirket af udbredelsen af Advanced Metering Infrastructures (AMI), der består af en stor mængde af sammenkoblede sensorer/forbrugere. Moderne AMI er i stand til af levere flere og forskellige parametre sammenholdt med traditionelle måleinfrastrukturer, hvor data kun bruges til fakturering. AMI åbner muligheden for at udnytte de tilgængelige informationer til mere effektiv netovervågning, planlægning og kan endda bruges til forudsigelses- og hændelsesdetekteringsformål. Nye datadrevne analysetekniker er derfor nødvendige for at udforske de nye parametre fra AMI, der bringer elforskingsområdet mod digitalisering.

De store og varierede mængder data, der er fremsendt fra AMI, er for nylig blevet omtalt som Big Data inden for både industri og forskingsområder. Denne definition er endvidere forstærket af den moderne kommunikationsinfrastruktur, som gør det muligt at streame data fra AMI med en meget finere granularitet, kendt som realtidsdata.

Formålet med dette ph.d. studie er at undersøge hvordan AMI data kan bidrage til en mere effektiv netdrift for DSO'er ved hjælp af behandling, analyse og visualiseringsteknikker. Den gennemførte forsking er baseret på et realt elnet scenario i samarbejde med en dansk DSO fra Thisted. Der har været fokus på design og implementering af et visualiseringssystem baseret på DSO specifikke behov og krav. Samtidig blev de tilgængelige geografiske og tidsseriedata brugt til at udføre data-nøjagtighedsundersøgelser og til at foreslå en potentiel analyseplatform for DSO'erne.

Brugervenligheds studier (UX) har været en vigtig del af arbejdet, især for at designe et simpelt og effektivt visuelt overblik over lavspændingsnettet. I disse undersøgelser deltog brugerne (DSO'er) i på-stedet interviews og hjalp på den måde med at forme brugergrænsefladen til den prototype på visualisering, som kan anvendes til overvågning, planlægning og forudsigelser. Bidraget består i at højne automatikken for netdrift-operationer på forbrugerniveau ved at designe og implementere et beslutningsstøttesystem. Endvidere bidrog brugen af geografiske informationssystemer (GIS) til rumlig og situationsbevidsthed, især væsentligt i et driftsmiljø afhængigt af menneskelig indgriben.

Derudover har de tilgængelige tidsseriemålinger og GIS-nettopologi været en del af en undersøgelse vedrørende præcisionen af data udvekslet i elnettet. Det har vist sig, at på grund af manglen på et fuldt integreret datasystem er der ofte unøjagtigheder i de data, der udvekles mellem de forskellige parter, hvilet udmønter sig i fejlagtig brug af oplysninger til de forskellige net-operationer. For at bibringe DSO'erne en højnet funktionalitet er forbrugeradfærdsstudier blevet gennemført. På baggrund af deres resultater er en klassificering af lavspændingsnetforbrugerne blevet forslået efter deres energiforbrug. Det blev vist, at de oprettede clusters er nyttige til netplanlægning, selv i tilfælde af manglede oplysninger, samt til at forudsige, hvordan en bestemt kunde måtte opføre sig på baggrund af sin profil.

Endelig indebærer resultatet af dette arbejde optimering af DSO'ernes daglige arbejdsgange ved systemredesign og minimering af OPEX omkostningerne ved at integrere intelligente analysemetoder. Den udførte forsking viser, at selv enkle statistik- og maskinindlæringsmetoder kan bringe intelligens til det nuværende elsystem, som hjælper med automatisk anomalitetsdetektering og med data-nøjagtighedsdiagnoser. I fremtiden vil denne samt anden nuværende samt fremtidig forsking bidrage til udviklingen af de såkaldte Smart Grids.

Contents

Thesis Details

Thesis Title:	Automation of smart grid operations through spatio-temporal data-driven systems
PhD Candidate:	Maria Stefan
Supervisors:	Assoc. Prof. Rasmus Løvenstein Olsen - Aalborg University Assoc. Prof. Jose Manuel Gutierrez Lopez - Aalborg University

This thesis is submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) from Aalborg University, Denmark. The thesis is compiled as a collection of papers resulting in the main part of the thesis being scientific papers published in, or submitted to, peer-reviewed journals and conferences. The work presented in the thesis is the result of three years of research, in the period June 2016 – May 2019, as a PhD fellow in the Section of Wireless Communication Networks (WCN), Department of Electronic Systems, Aalborg University.

The PhD stipend (nr. 8-16026) has been funded as a part of the RemoteGRID project. The ForskEL program under Energinet.dk have together with Aalborg University and industry partners; Thy-Mors Energi and Kamstrup financed this project.

The main body of this thesis consist of the following papers:

A. Maria Stefan, Jose G. Lopez, Morten H. Andreasen and Rasmus L. Olsen, "Visualization Techniques for Electrical Grid Smart Metering Data: A Survey", *IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, 2017

B. Maria Stefan, Jose G. Lopez, Morten H. Andreasen, Ruben Sanchez and Rasmus L. Olsen, "Data Analytics for Low Voltage Electrical Grids", *Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security - Volume 1: IoTBDS, pp.221-228*, 2018

C. Maria Stefan, Jose G. Lopez and Rasmus L. Olsen, "Exploring the Potential of Modern Advanced Metering Infrastructure in Low-Voltage Grid Monitoring Systems", *IEEE International Conference on Big Data*, 2019

D. Maria Stefan, Morten H. Andreasen, Jose G. Lopez, Michael Lyhne and Rasmus L. Olsen, "Automation of smart grid operation tasks via spatio-temporal exploratory visualization", *The journal of Environment and Planning B: Urban Analytics and City Science*, SUBMITTED 2019

E. Maria Stefan, Jose Gutierrez, Pere Barlet, Oriol Gomis and Rasmus L. Olsen, "(Position paper) Characterizing the Behavior of Small Producers in Smart Grids. A data sanity analysis", *Journal of Applied Energy*, SUBMITTED 2019

According to the Ministerial Order no. 1039 of August 27, 2013, regarding the PhD Degree § 12, article 4, statements from each co-author have been provided to the PhD school for approval prior to the submission of this thesis, regarding the PhD student's contribution to the above-listed papers. The co-author statements are also presented to the PhD committee and included as a part of the assessment.

In addition to the listed papers as the main content of this thesis, the following paper is co-authored during the PhD studies. As this paper is not a part of the main body of this thesis it has not been included in print. The reader is therefore kindly asked to refer to the respective publishing channel.

1. Ruben Sanchez, Florin Iov, Mohammed Kemal, Maria Stefan and Rasmus Olsen, "Observability of low voltage grids: Actual DSOs challenges and research questions", *52nd International Universities Power Engineering Conference (UPEC)*, 2017

Preface

I enjoy a challenge and I always make every effort to finish what I have started. However, when I came to Aalborg University in 2013 for my master studies in Wireless Communication Systems I did not expect that I would be pursuing my career towards a PhD researcher. As an engineer with a background in Telecommunications, I always thought that I would keep on shaping my career - either as researcher or as pure engineer, in the field of 5G New Radio communications. However, it so happened that the opportunity arises for me to continue my studies with a PhD in data analysis and visualization in the domain of Smart Grids. The journey was both challenging and exciting, having to refresh my memory about the different computer science and electrical engineering topics that I have covered throughout my previous years of study, as well as becoming up to date to the field of Smart Grids.

Thanks to this opportunity, I got the chance of collaborating closely with my co-supervisor, Jose Gutierrez, who has always been my support both morally and working-wise. Therefore, I would like to extend all my gratitude and respect to Jose Gutierrez, who had a great contribution to the overall work done in the PhD, as well as in my personal development as researcher and as an individual. By the same token, I would like to acknowledge the help and friendship of our colleague, Morten Henius, whose positive attitude always helped me move forward with my work, even in the most difficult times.

Another person who deserves my utmost gratitude is Prof. Josep Solé Pareta from Universitat Politècnica de Catalunya, Spain, who was my adviser during my stay-abroad period. His professional advice and kindness contributed to a significant part of my PhD research, at the same time making me feel like Barcelona is my other home.

I would also like to extend my appreciation towards my closest colleagues, Kaspar Hageman and Thomas Kobber Panum, for the fruitful discussions and for their willingness to listen to my complaints. I do hope that we will get the chance to work together again in the future.

This PhD would not have been possible without the support of my main supervisor, Rasmus Løvenstein Olsen, who introduced me to the field of

Smart Grids and opened up new research possibilities for me. I am grateful to him for all his support and for helping me get through the challenges of managing the PhD studies. Similarly, I give thanks to Michael Lyhne from Thy-Mors Energi for his patience and contribution to this research.

Lastly, I would like to acknowledge the unconditional support of my families - from both the Romanian and the Danish side. I have received a great deal of support from my parents - Radu and Florentina, as well as from my partner, Troels Jessen, who were always there for me even when I have lived far away from home for a very long time. Without their encouragement and positiveness, I would not have found out how far I can get away from my comfort zone, which leads me to think of a quote that boosts my motivation:

There's a better way to do it - find it. - Thomas Edison

Maria Stefan
Aalborg University, May 29, 2019

Part I

Introductory Chapters

Chapter 1 - Introduction

The purpose of this introductory chapter is to bring out the main topics of the PhD thesis. The first two sections present the overall problem definition and motivation for the Danish power system automation. This is followed by the corresponding research challenges and contributions, which aim to give a short overview over this work.

1.1 Electrical grids in Denmark

When H.C. Andersen wrote his adventure stories, few people had knowledge of the value of oil, coal and natural gas. The world was on the dawn of industrial revolution and, not least, oil was its drive. Years after the writer's death, fossil fuels continue to bring welfare for millions of people, however this development has its cost. The unpleasant consequences of a warmer global climate are due to coil-based power stations and oil-based transport. — Jesper Tornbjerg, 2014 [91]

A safe energy supply is the core task of electricity companies all over the world. Danish electricity regulations state that environmentally friendly electricity takes over coal-based power [90]. This means, for example, that the current from wind turbines must be used before the one from power stations. In 2002, 63% of Denmark's electricity was produced by large central power stations, 14% by outlying stations and 23% by wind turbines [53]. The vast majority of plants are combined heating and power (CHP), producing electricity and supply district heating simultaneously. A cold, windy day means more electricity produced and consumed - wind turbines will be generating more, homes will turn up the heat, causing CHP stations to generate more electricity [36] [13]. Such situations can be problematic, especially at night when industrial production and power consumption is lower, as too much current will cause the grid to break down [55].

A Danish electricity customer had power in average 99.9% of the time in 2013 [23]. However, two storms with the wind speed of a hurricane stroke on October 28th and December 5th in 2013, challenging the capabilities of the energy system, as windmills shut down at wind speeds higher that 25 m/s [91] [72]. Automation and monitoring of the electricity grid can ensure that there is current flow in the cables, while the data from the modern smart meters can be utilized to find out what is happening in the electrical network. Data availability opens the possibility for optimized grid planning, such as replacing old installations and fixing errors [15] [96].

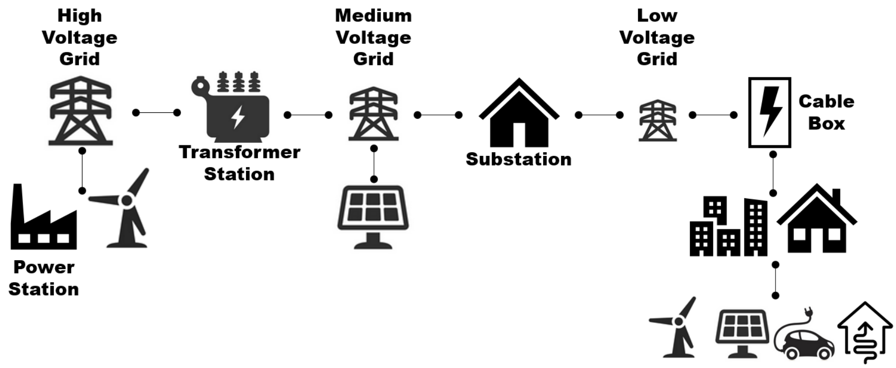


Fig. 1.1: Representation of the modern Danish electrical grid, from transmission to consumers.

The sources of data in the modern Danish low-voltage electrical grid vary from heat pumps, windmills, electrical vehicles to solar panels, which can be depicted in Figure 1.1 [24]. It is expected that the energy consumption will increase with the introduction of more electrical vehicles, CHP and other types of green energy sources. Therefore, applications are required for maximizing the value incoming from distributed energy resources (DER) and efficient energy consumption management [4] [38]. Such an application can be smart control of households, apartment buildings or corporations' energy systems, by obtaining information about energy pricing.

In the next section the background for this PhD research is introduced, given the aforementioned presentation of the low-voltage Danish electrical grids.

1.2 Problem statement

Electricity grid operators prepare for the future as state-of-the-art technologies emerge and as they are implemented to enhance efficiency and business opportunities. The subsequent electricity grid evolution is focused towards the development of smart grids, capable of utilizing complex data analytics correlating different high volume and mixed data sources, also known as the Big Data concept. Daily workflows for grid management can be improved via decision support systems to ensure an affordable, reliable, secure and sustainable electricity supply [21] [92].

The process is further motivated by the national Danish regulations as well as international political climates [58]. As current and future legislation demand not only efficiency via impending requirements, but they are also very much focused on the inclusion of renewable energy sources (RES) as part of a climate centered strategy [99]. New technology systems require utilizing and supporting the enormous influx of smart devices and sensors. The corresponding exponential growth in data originating from these devices reveal anomalies, among which the most common are cable faults or voltage magnitude threshold reached [103]. Also, importantly, the electricity grid was originally designed solely for central operator-controlled electricity production with a one-way flow model. However, driven by commercial and residential energy generation via the integration of RES, primarily photovoltaic and wind turbine generators, electricity grids are shifting from a unidirectional flow topology towards distributed energy generation [70] [74].

Due to the volume and variety of data [69], the Danish Distributed System Operators (DSOs) face operational challenges, since the current system operations rely solely on customers' input to manually report common issues, such as residential power surges and outright power outages [73]. The future proliferation of RES is expected to induce increased instability as a byproduct of the adaptation process towards a decentralized power generation grid architecture [41]. As a consequence, increased stability and reliability in the low-voltage grid for effective grid monitoring and advanced operation becomes harder to maintain for DSOs, in order to allow for preemptive actions as opposed to current reactive workflow patterns.

Future grid management and operations are promising due to utilizing advanced metering infrastructures (AMI) data. AMI units are installed throughout the low-voltage grid, either as natural replacement is required or as direct upgrades [59]. These AMI smart meters are capable of logging and transmitting various detailed information and in much higher resolution than traditional electricity meters [62], with data ranging from electricity consumption to specific phase voltages. Present-day AMI data is utilized for conventional billing purposes [3], without putting to use the full potential of the mass available information and the highly increased data granularity. Also, capabilities that facilitate near-real time monitoring and automated daily operation with instantaneous anomaly detection can be provided through modern AMI [88].

This opens up a new spectrum of possibilities for investigating means to deploy automatic monitoring and planning solutions for the Danish electrical grids, which have been inaccessible up until now.

1.2.1 Hypothesis

In reference to the problem statement presented in Section 1.2, the following hypothesis is formulated in relation to this PhD research:

> *It is hypothesized that efficient data processing, analysis and visualization of smart metering data can help the DSOs in making useful decisions for future grid planning, event predictions and for automatically detecting anomalies in the grid.*

Proceeding from the hypothesis, the main methods and information systems to be investigated are:

- Mapping and visualizing spatio-temporal data using Geographic Information Systems (GIS);
- Use of database architectures that support large amounts of data;
- Data processing and analytics for extracting relevant parameters and knowledge out of the available data;
- Design and implementation of components and interfaces for automatic decision support systems.

1.2.2 Case study - test area at Thy-Mors Energi

This research has been carried out in relation to a real-life test area located in the north-west part of Jutland, Denmark, which is shown in Figure 1.2. The information has been made available by the distribution company in the area, as part of the project - Thy-Mors Energi [6]. The area is relevant for this study due to the presence of renewable resources at the residential level, mostly small wind turbines and PV systems.

Some anomalies have been previously detected in this part of the grid, such as over and undervoltages or imbalances in households' power. While the distribution company is responsible for deriving offline procedures to counteract these issues, an increasing number of reported problems will require more man power and time spent on error debugging procedures, thus implying economical repercussions. Currently, the DSOs from Thy-Mors Energi use various software programs for investigating historical events in the power grid, in the form of visualization and/or parameter calculation tools. For the high and medium voltage parts of the grid, the SCADA system (Supervisory Control and Data Acquisition) is actively used for visualization and anomaly identification. However, the low-voltage information is currently not fully integrated into SCADA, making the various consumer-related data management procedures challenging, as the DSOs have to manually handle different software tools to address errors or other significant events.

With the evolution of AMI and Big Data conceptualization, more effort will be needed towards grid planning and event predictions, rather than the

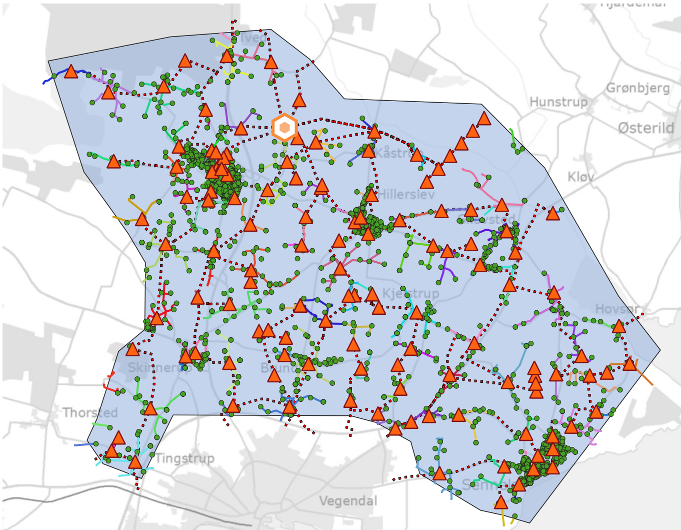


Fig. 1.2: Representation of the test area polygon, including: the primary substation, secondary substations (red triangles), customers (green dots) and their interconnections. The red dots represent the masts in the medium-voltage grid, while the different colors depict how each secondary substation feeds a certain group of users.

current time-consuming manual error debugging. As a consequence, an automatic decision support data-driven system is considered adequate for the DSOs' daily operations.

1.3 Research challenges and contributions

Based on the case study of the Danish DSO presented in Section 1.2.2, an integrated analytical and visual information platform is expected to ease the low-voltage grid interoperability and to increase the DSOs' efficiency in the overall business structure, by minimizing redundant procedures. The work in this PhD study is focused on the following research challenges:

- The choice of database environment to be used and identifying the events involved in the data processing;
- Investigating how to process and convert data to optimize the interaction with the end visualization system;
- Providing the users (DSOs) with adequate information in order to make useful decisions;
- Obtaining an automatic information-based operational system via analytical methods.

The corresponding contributions are made towards resolving the defined research challenges, by designing and implementing a data-driven system suitable for the DSOs' daily operations. This was achieved both from a research and from an enterprise point of view, using the theoretical background to establish the most suitable tools for carrying out the study in both cases. As there is more decision freedom from a research perspective, an enterprise-oriented solution involves more specific knowledge about the DSOs in their working environment, thus adapting the proposed information system accordingly.

To sum up, this PhD study aims to show how a combination of different tools and theoretical knowledge can contribute to developing an automatic decision support system for the Danish DSOs. Particularly, it is shown that the results from research can be applied by distribution companies to optimize the usage of the distribution network resources and to minimize the manual work.

Chapter 2 - Theoretical Background

Traditional electricity grid monitoring and decision support are based on a multitude of different systems, demonstrating a natural additive approach to technology adoption over time [78] [14]. As new capabilities are deemed necessary or advantageous, different systems are introduced , aiming to create a dedicated data-driven technology platform. As mentioned in the introduction, visualizing the low-voltage electrical grid data has the potential to evaluate and to anticipate grid anomalies, and to speed up other corresponding actions regarding grid maintenance and monitoring. Therefore, this chapter will cover the basis of the methods utilized for achieving the proper data presentation in this research, by covering three main topics:

- Geographic Information Systems
- Database Management Systems
- Data processing and analytics techniques

2.1 Geographic Information Systems - GIS

Geographic Information Systems (GIS) are, as the name indicates, a combination of two different disciplines: geography and information systems [61] [18] [17]. *Geography* is the science dealing with the physical, biological and cultural features of the Earth, in other words, data associated to a location. *Information systems* are, generally defined, as a set of components that work together to achieve a common goal, by utilizing the data characteristics/attributes attached to the location. The components involve: data, hardware and software equipment, humans, operational procedures and subsystems for data management, with the main goal of transforming the *data* into valuable *information*, *knowledge* and *wisdom* [7] [33].

The DIKW diagram illustrated in Figure 2.1 shows the structural and functional relationships between data, information, knowledge and wisdom. By undergoing the transition from raw - meaning - context, data brings value to the human interpretation by helping decrease the computational complexity at more advanced stages in the process [26] [66].

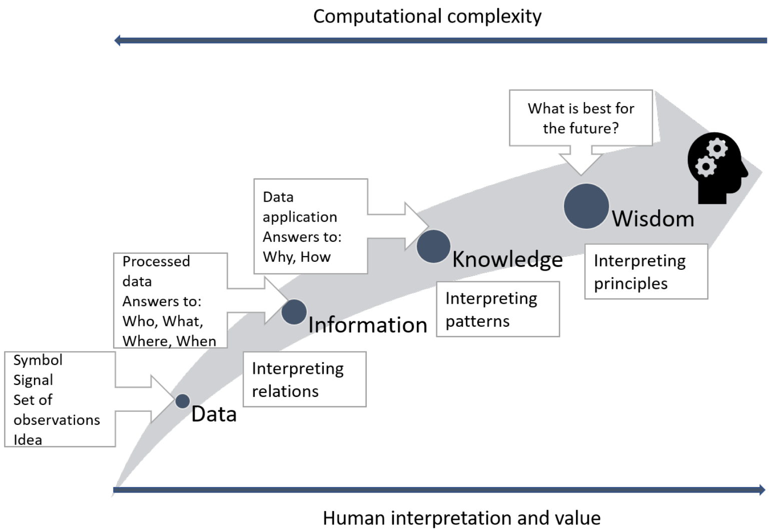


Fig. 2.1: DIKW data transformation diagram [10]

Considering that data is the starting point in GIS, there are three main data types [30]:

- Spatial (vector) data: features represented as points, lines and polygons, as previously shown in Figure 1.2;
- Attribute (tabular) data: qualitative and quantitative characteristics of the spatial entities;
- Raster data: landscape represented as a rectangular matrix of square cells, useful for elevation, terrain, slope and risk analysis, etc.

This PhD research was based on the electrical grid spatial and attribute data in GIS [95], particularly for extracting labels for time series measurements, as it has been done in Paper E. Moreover, the benefits of the spatial features with respect to improving the DSOs' operational procedures were evaluated in the studies performed in Papers C and D.

GIS usually have an integrated database management system [89] [77], where the data model is represented by the different objects in the spatial database and the relationships among them. Each feature on the map can be characterized by attribute data, which is typically manipulated in relational databases by means of queries [48]. Three of the most common types of database systems will be subsequently presented in the following section.

2.2 Database Management Systems

This survey concentrates on the three primary Database Management System (DBMS) categories and will present a high level introduction to these specific database types and highlight the technological characteristics, advantages and inherent shortcomings. Lastly, generic database evaluation criteria are highlighted as a foundation for requirement specification. The survey is based on the following sources: [94] [51] [20] [22] [87] [16].

A database schema is the design blue print of how the DBMS is constructed. The schema defines the basic structures on both the logical and physical level, providing a descriptive detail of the how data is organized, the corresponding relational structures including how every constraint is applied, as well as storage definitions for all database elements. Thus, the schema plays a paramount role in determining application suitability and flexibility as well as both data and transaction integrity parameters. DBMS scalability identifies the abilities of the system to be upgraded and expanded, and hereby determining both present and future performance and capacity capabilities.

The primary DBMS categories chosen for this study are:

1. Relational DBMS: uniquely based on proven mathematical foundation, specifically by Georg Cantor's Set Theory [51] and Relational Theory by Edgar Codd [94], it guarantees a high level of stability and robustness. With a more than 30 years proven track record, the RDBMS is the industry standard, commonly utilized as a comparison baseline, and it is characterized by offering sufficient data storage, protection and access capabilities. Also, its performance is reliable, making it adaptable for business intelligence use cases;

2. NoSQL (Not Only SQL) DBMS: broad descriptor for next generation database systems, typically characterized by being open source, non-relational, distributed and horizontal scaling. Polyglot Persistence [20] is the NoSQL jargon for selection between the different data models of the NoSQL DBMS categories matching use case and application requirements. Therefore, NoSQL databases provide freedom of choice to match a custom architecture specifically to the application and problem set [22], by reducing the DBMS complexity via Polyglot Persistence;

3. In-memory DBMS: An in-memory database (IMDB) is generally a RDBMS using RAM instead of traditional disk storage [22] [87], typically providing full SQL support. It can substitute existing RDBMS with only minor adaptations if not seamlessly. The vast majority of IMDBs are based on a vertically scalable Symmetrical Processing (SMP) architecture, with limiting large-scale scalability and throughput, due to the

inherent inability of SQL join features to operate efficiently in a distributed environment [16]. To take advantage of new middleware in-memory capabilities, IMDBs require keeping current access applications functional via continued SQL processing operations but with significant changes to the existing database. The cost model for IMDBs is hence primarily dependent on single server architecture pricing.

The current Thy-Mors data model is implemented as a relational DBMS. Considering the current implementation is based on Microsoft SQL Server 2014, which introduced Memory-Optimized Tables and Natively-Compiled Stored Procedures In-memory features [5] [97], a continuation with relational DBMS was considered favorable. Enhancement contrary to a complete technology switch allows integration with the current data model implementations as well as encourages taking advantage of the already existing in-house expertise and knowledge base. The In-memory Online Transaction Processing (OLTP) [16] [5] support of MS SQL Server enables adequate performance enhancements and legacy capabilities for both near-real time (dynamic) and historical (static) data types. The workload areas which benefit the most from In-memory OLTP technology [22] include high data rate insert rate with smart metering as the primary example, read performance and scale, computer heavy data processing and low latency workload categories.

At the reseach level, an implementation of the relational DBMS based on MS SQL Server meets the main requirement for managing large historical data sets and is considered suitable for this work. Additionally, given the aim of visualizing spatio-temporal data, the PostgreSQL DBMS is the most suitable choice. This is integrated with the PostGIS spatial database extender which provides support for different spatial capabilities to the existing database, such as geometrical data types and geocoding in-built functions.

For experimental purposes, MSSQL's in-memory data storage was used in Paper C to simulate data streams. Due to the requirements for spatio-temporal data visualization, the research conducted with real-life measurements and grid topology in Papers D and E was carried out using PostgreSQL.

2.2.1 Database Privacy Features

Electricity meters track energy use which can violate the right to privacy and protection of personal data [44]. The customers' energy consumption can reveal information about the number of people in a household, daily routine and usage of appliances. Sometimes, the data can reveal particularly sensitive information, such as criminal offenses. Protection of personal data is therefore regulated in detail by the General Data Privacy Regulation (GDPR) [9]. This project is carried out taking into consideration the privacy regulations regarding personal data. In this context, there are two types of datasets related to individuals: GIS data (location of the meters - addresses, as X and Y coordinates) and metering data (value of any relevant parameter measured by the meters).

For the research purposes, it is essential to keep these two datasets separated and uncorrelated in order to preserve the privacy rights of the Thy-Mors customers in the test area. The concept is illustrated in Figure 2.2. The

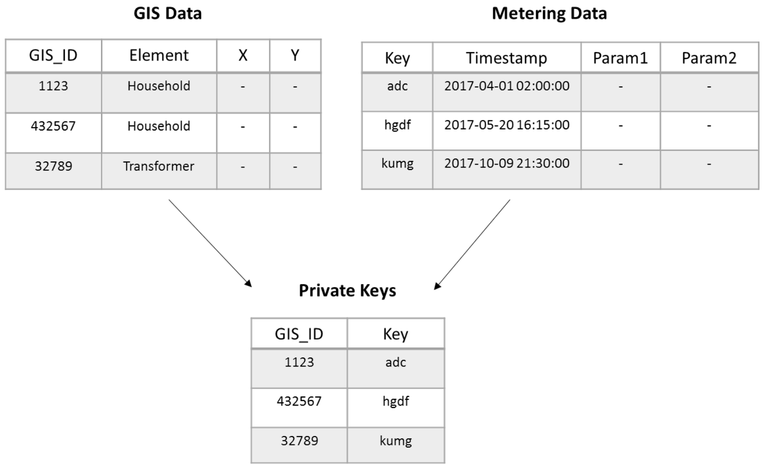


Fig. 2.2: Separation of data tables and elements to secure privacy and anonymity of data collected

data is anonymized by removing all references to physical meters such as ID and giving each data point a *key*, each of the meters within the test area having their own unique key. The keys can only be decoded using the Private Keys table, necessary to relate a key to a specific address. Two copies of the table are at:

- The meter distributor (Kamstrup) for data anonymization;
- Thy-Mors to obtain the GIS_ID, which is useful for the visualization in the GIS environment.

By legal authorities, this table is only used when strictly necessary, in accordance with applicable data protection and privacy regulations:

- for preventing or respond to cyberthreats or cybersecurity incidents [67];
- for preventing other criminal actions or for preventing actions by the consumers which may cause a risk to the functionality of the grid.

In reference to the requirements of data processing, the next section will focus on different processing and analytical techniques that build upon the prerequisite of obtaining valuable knowledge from the data.

2.3 Data processing and analytics techniques

Data processing is commonly categorized as either batch or stream processing. Where batch processing constitutes classic periodic data processing, and by inverting the paradigm, stream processing implements persistent data flows, queries, analytics and application logic. Batch data sets and workloads are characterized by having a finite data source, representing static at-rest information. Inversely, stream data sets and workloads have a theoretically infinite data structure, frequently described as event time series in-motion information.

2.3.1 Data processing

The following sections present the main characteristics of the two common processing paradigms - batch and stream.

2.3.1.1 Batch Processing

This section is based on the following sources: [100] [101] [32] [1] [68].

Batch processing (Figure 2.3a) represents the common case processing, with management and computations over all or most of a data set. The processing is run off-line on persistent data blocks frequently according to predefined periodic schedules. Traditionally, batch processing is focused on throughput and complexity performance, designed to manage large data volumes while executing computational intensive algorithms. Latency is considered a secondary objective, typically measured in minutes to hours.

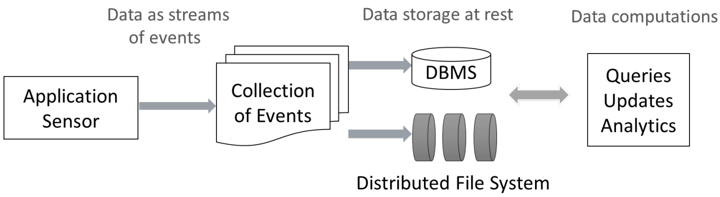
The main characteristics of batch processing are:

- Static finite data at-rest data sets[32];
- Volume-centric processing [32];
- High throughput performance [101];
- Scheduled offline processing;
- Periodic recalculation over all/most data;
- Data persistence [101];
- Data scope in the range of minutes to years.

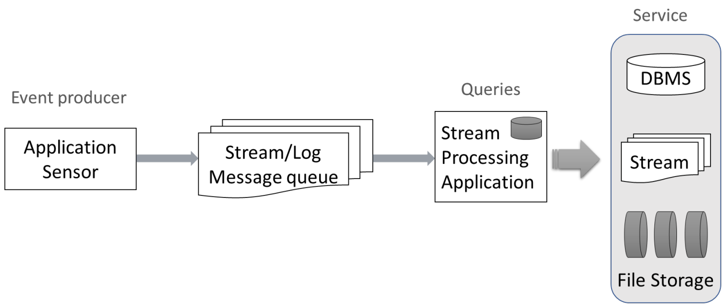
2.3.1.2 Stream Processing

The technical content of this section is based on the following sources: [54] [86] [100] [19] [25].

The stream computing data processing paradigm (Figure 2.3b) enables management of continuously generated data, falling in the category of Complex Event Processing (CEP) as an element of Big Data technology. The primary purpose of stream processing is to provide exceptionally low latency



(a) Representation of batch processing [19]



(b) Representation of stream processing [19]

Fig. 2.3: Data flow in batch and stream processing techniques.

velocities independent of high volume mass storage. Stream processing is a technology introducing real-time or near-real-time, which, depending on the environment, is defined as microseconds to several days. Processing performance ensures uninterrupted information streams as well as direct interaction with non persistent data before any potential storage procedures. Stream processing unifies analytics and applications in a single common architecture, introducing direct analysis result integration into applications for automatic and instantaneous action. Providing continuous query capability is essential for sensor applications, web events, machine and application logs, social data.

The main characteristics of stream processing are:

- Low latency processing [86];
- Instantaneous application and analytics reaction to input events;
- Management of individual records or micro batches;
- Continuous and unbound event driven data sets;
- Periodic recalculation over all/most data [100];
- Decentralization and decoupling of infrastructure [100].

2.3.2 Data Analytics

In this section the main data analytics concepts are introduced. These concepts serve the purpose of converting processed data into relevant information, which subsequently is prepared for either additional analysis or directly organized for presentation and visualization objectives [12] [35].

Data analytics are organized into two main categories:

1. Historical: analysis based on the past; data-at-rest corresponds to batch data processing [42]. Historical data analytics provide insight by uncovering data patterns and trends, allowing for a concise presentation of large data sets. By utilizing different algorithms for reducing complex data sets [56] [8], event forecasting is also possible [76]. The drawbacks of historical analytics are related to their limited reactions to past events and update intervals resulting from the batch processing.

2. Real-time/streaming: analysis based on the present; data-in-motion equals stream data processing [65] [46]. Real-time data analytics make it possible to enhance the reaction time for decision makers via clarity on current unfolding events. As a result, correlations between multiple and diverse data sources can be detected [98] [10], at the same time opening the possibility for predicting imminent events or failures (i.e. fraud detection). Despite the explicit advantages, real-time analytics are very much platform and hardware dependent [34], causing potential incorrect analysis and decisions. At organizational level, adapting to new work patterns to take advantage of the continuous flow of information is also challenging.

Thus, the choice of suitable analysis techniques is based both on the application requirements and on the implementation flexibility.

In the process of knowledge and information discovery, different types of analytics can provide various levels of in-depth knowledge of the data, depending on the available data set and the application requirements. This is achieved through the four traditional types of analytics - *descriptive*, *diagnostic*, *predictive* and *prescriptive* [35] [52] [31], as shown in Figure 2.4. The trade-off between the obtained value and the implementation difficulty increases for predictive and prescriptive analytics, as they open the possibility for process optimization, as well as for better understanding and exploring the value extracted from the data.

In this PhD research the focus in mainly on characterizing the behavior of low-voltage grid consumers using basic statistical studies (descriptive and diagnostic). Furthermore, the study in Paper E is solely based on the available static data, in the form of GIS grid topology and historical time-series energy consumption measurements. A step forward is taken in the analysis with

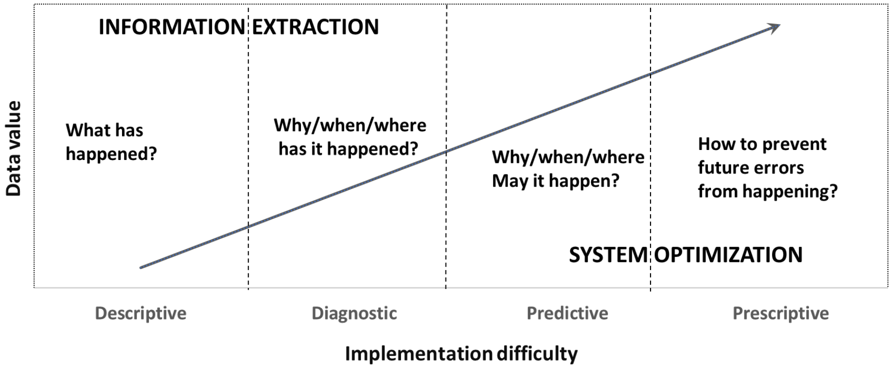


Fig. 2.4: Representation of descriptive, diagnostic, predictive and prescriptive analytics types and the corresponding questions to which they answer [35].

respect to predictive analysis by applying forecasting models to the available data set and by prescribing recommendations for human assessment.

2.4 Decisions

The three topics covered in this chapter - GIS, DBMS, data processing and analysis, were meant to give an overview of the main tools used in this PhD research. The study was conducted both from a practical and from a scientific point of view, leading to two main research tracks:

- User experience (UX) studies
  The DSOs were the central part of the UX studies. The purpose was to evaluate the DSOs' daily working procedures and to identify some scenarios where they are constrained from operating the grid efficiently due to manual error debugging. The DSOs' feedback was useful for deciding which analytical techniques are most suitable for designing and implementing an automated information system for monitoring, planning and prediction (Paper D).

- Developing an information system for the low-voltage electrical grid
  This track was oriented towards the information system development, based on the previous UX studies. The system comprises of relational DBMS for data storage and feature extraction (pre-processing) techniques, such as filtering and selection. The features can be used for statistical analysis, data mining and forecasting, as it was done in Paper E. In terms of visualization tools, WebGIS was used for data presentation in Papers C and D, while QGIS was employed for both data presentation and analysis (Paper E).

Chapter 3 - Research tracks

The topic of this chapter is related to this work's research tracks, which are depicted on the roadmap in Figure 3.1. Considering that the end user of a data visualization and analysis system are the DSOs, their requirements and needs come first when designing a dedicated application. User experience studies focused on the DSOs workflows are first presented. Secondly, the prior UX knowledge contributes to deriving a data-driven solution for the current electrical grid system.

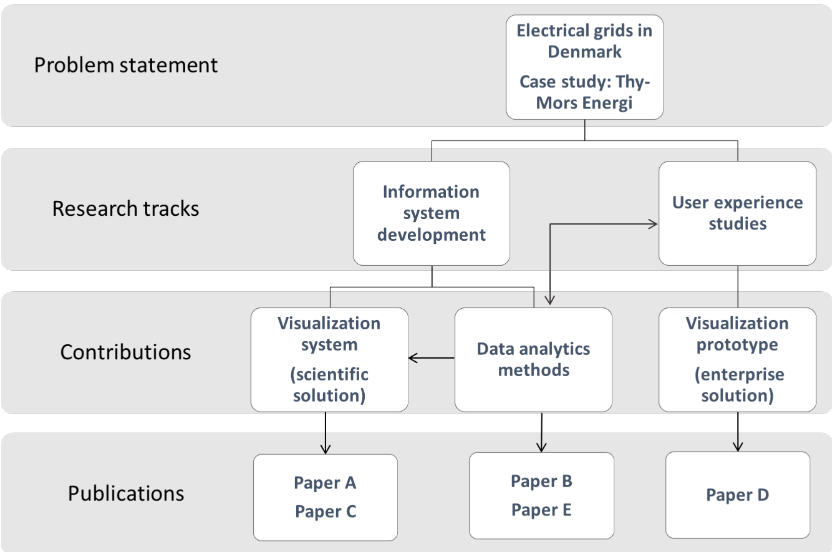


Fig. 3.1: Roadmap of the PhD research.

3.1 User experience studies (UX) - Distributed System Operators (DSOs)

In this section, the approach for performing UX studies is covered, where the users are the DSOs from Thy-Mors Energi (case study in Section 1.2.2). The UX study was performed based on on-site interviews. The users' daily work routine is analyzed via the 'Day-in-the-Life' model and user profiles are created for the involved DSOs.

3.1.1 'Day-in-the-Life' Model

The purpose of utilizing the Day-in-the-Life Model [40] [39] in this work is for identifying where DSOs operations can be improved time-wise. A general user profile, as represented in Figure 3.2, can help in understanding a DSO's routine in a normal working day. The Day-in-the-Life model brings together the overall structure of how work fits into the user's day and how this is supported by different mobile and stationary devices. The focus of this model is on the different places, timings and platforms that together contribute to activities getting done.

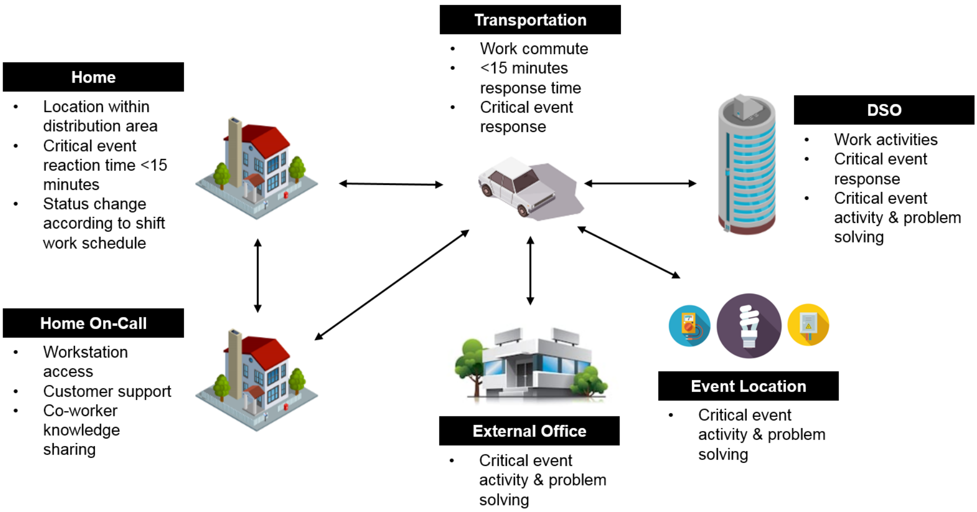


Fig. 3.2: Day-in-the-Life Model applied to the Danish DSOs, showing scenarios for an operator on a shift at home, during transportation, at the TME headquarters and smaller offices.

Three main activities and spatial contexts are identified during the DSOs' day in their weekly shift: at the work place, at home doing everyday activities (on call) and at home during an ongoing event. An on call DSO has to live inside the distribution area and be able to act upon an event within 15 minutes from its signaling. This limitation in the DSO's daily life also implies interrupting everyday activities at any time of the day. The DSO might also need to call another colleague for knowledge sharing and advice, meaning that this limitation is general among the DSOs. At the work place, the DSO needs to be able to interrupt a routine activity and prioritize an important event. The DSOs communicate with both internal and external actors - customers (private households and companies), technicians, contractors, other DSOs and departments, through an *error messaging system*. One artifact here is an SMS message to the customer informing them about a possible power outage. Having smaller sub-offices at several transformer stations in the distribution area makes it possible for the DSOs to drive to the nearest office in case they are more than 15 minutes away from the official work place.

The user analysis during a working day can help establishing which of the work processes are most time consuming, so that they can be automated for an optimum operation and planning of the electrical grid.

For example, acting on a calling customer can involve less time if the exact event and location (address) of the customer are signaled through visual alarms. The demand of being able to work on the error within 15 minutes, as well as to being alert at all times, impact the timeliness of these alarms.

In order to achieve the automatic fault detection in the electrical grid, user experience studies have been made in order to establish what data and how the end user (DSO) wants to visualize it. The steps of these studies are:

- Establishing different user profiles from the control center – electrician, electrical engineer;
- Establishing different scenarios where errors are reported;
- Defining sequence models for the chosen scenarios and the current procedures to address these errors;
- Identifying the main themes to be addressed as part of the final visualization prototype;
- Designing a prototype that would bring an automated solution to overcome some of the challenges reported by the evaluated users;
- Validate the solution by involving the different actors in the value chain (vendors and DSOs).

3.1.2 User profiles

User profiles are seen as part of the consolidation in relation to contextual design models, in which the focus is directed on the DSOs in their working environment, therefore they are the central part of the design process. The purpose of these profiles is to ensure that the system design will benefit the users' workflow and, as a consequence, it will be more directed to the DSOs. For this study, the following user profiles have been identified in Table 3.1.

Table 3.1: Distributed System Operator profiles based on on-site interviews.

	Education	Job title	Time in the company	Area of responsibility	Competencies
1	Electrician	System operator	3.5 years	Planning, developing, building and maintaining transformer stations	Knowledge about building transformer stations
2	Electrical engineer	System operator	6 months	Reviewing the 10/0.4 kV distribution stations and reporting of errors	Engineering background
3	Electrician	System operator	32 years	Filing reports to the energy consumption agency, when the electricians have solved an error; Measuring the electricity grid	Experienced in the field and knowledgeable of the company working structure

3.2 Information system for the low-voltage electrical grid

The second research track is focused on designing a strategic data-driven information system for low-voltage electrical grids, aiming to combine knowledge from both the UX studies and the processing of measurement data.

Measurement data from AMI in Denmark is nowadays typically done every 15 minutes and only for billing purposes. The aim of this research is to provide ways to process and analyze the incoming smart meter measurements with various types of readings containing different electrical parameters. The end goal is to provide a meaningful data display to the end users (DSOs), in order to obtain an overview over the current and the historical processes that take place in the power grid.

An overview of the proposed data system is shown in Figure 3.3. The two research tracks are illustrated as main inputs to the structure of the whole information system. The UX input is used as starting point for designing the data visualization platform. The incoming metering data from the AMI network is passed on to the DBMS for storage and processing, while the processing output can be utilized both for statistical analysis and prediction, and for feature extraction in relation to the visualization. Furthermore, depending on the result, feature extraction has an impact on what is most relevant for the application design.

Historical data is relevant for understanding different consumers' behavior by evaluating their consumption patterns, through classification/clustering

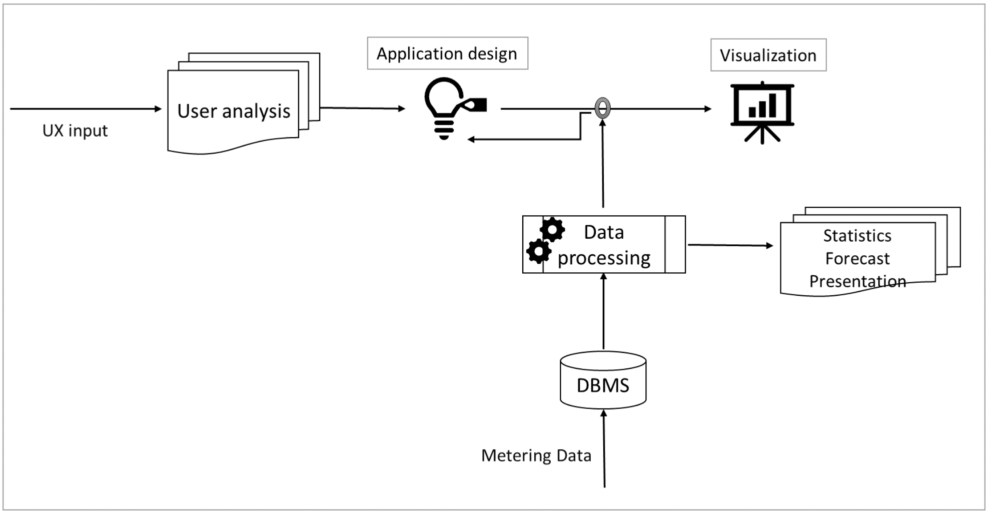


Fig. 3.3: Data flow for the proposed information system for low-voltage electrical grids.

methods, for example, consumers with or without installed RES. Consumption patterns also depend on environmental factors such as time of day (morning, afternoon, evening) or season. Consequently, the DSOs can use this information for optimizing or planning ahead future updates in the smart grid infrastructure, as well as for fraud and outage detection.

A meaningful information display is based on the DSOs requirements and needs, thus impacting which type of processing techniques should be utilized to extract certain features of the data. The aim is to discover which information is helpful for the DSOs to monitor the grid status and how to present this in an integrated data system.

3.2.1 Main visualization themes as identified by the DSOs

The following visualization themes are meant to contribute to the design of the visualization system, according to the user analysis and the proposed data flow.

1. *Geographical low-voltage grid map*

   The purpose is to be able to see the values from a smart meter directly on the map, when clicking on a particular data point of interest. Eventually, the system should lead its user throughout the troubleshooting process by identifying all the interconnections between the high, medium and low parts of the power grid. Therefore, map interactivity and sorting mechanisms are important in the visual design of the map, as overloading of measurement data points becomes an issue with the low-voltage nodes density. In this case, feature extraction refers to filtering and/or selecting of the significant areas.

2. *Historical data visualization*

   Historical data display comes as an extension of the map interactivity function and it is mostly useful for planning grid reinforcements in the areas which are prone to problems. Currently the ID of the meters which issued anomalies is manually found by the DSOs [88] and their corresponding nodal measurements are also manually extracted from the DBMS. This activity is redundant, time consuming and can easily result in errors, such as incorrect or incomplete measured values. Due to this, data labeling can contribute to easily match nodal measurements with their corresponding GIS location and to easily extract historical measurements belonging to a certain consumer.

3. *Alarm visualization*

   Alarm display is a subcase of the raw metering data visualization. Any kind of anomaly in the grid (over/undervoltage, flickering) should be also visible among the incoming raw data. Map layer filtering and the interactivity function make it possible to detect the extent of the anomalies in the grid, whether they are related to a specific substation or dispersed across a large area.

The design of the new software solution has to fit the users' life and their different activities throughout the day, as shown in Figure 3.2. The purpose is to facilitate the DSOs' job in solving different tasks and debugging errors when the required functionalities are integrated in one system. Different types of visualization techniques are supposed to be alternative ways to the current manual searches of customer information and to allow for the possibility of cross-integrating multiple software tools, which are currently utilized at the enterprise level.

The system assessment is done in terms of Capital and Operating Expenditures - CapEx and OpEx, as depicted in Figure 3.4. Although no market research was done in relation to this study, the plot conceptually illustrates the economic feasibility of developing and implementing such an information-based system. The research done in this thesis indicates that there is a clear benefit (OpEx) reduction when implementing this model, by reducing the time and human resources required to solve incidents in the electrical grid. In each specific concrete case (different DSOs), the relation between OpEx reduction and the cost of developing and implementing the proposed system would indicate the economic feasibility of such migration in terms of Return of Investment (RoI). Normally, the acceptable RoI for software-based solutions is between one and three years [64] [45]. However, considering that

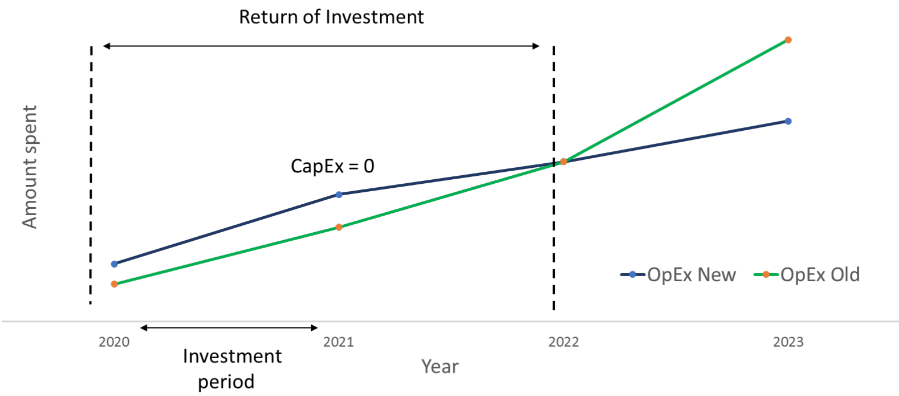


Fig. 3.4: Accumulated OpEx analysis for automation system integration at enterprise level.

electricity meters have a lifetime of up to 20 years [47] and that the proposed solution is fully cloud-based, the DSOs may consider longer a RoI as acceptable.

The graph in Figure 3.4 is related to the automatic decision-based visualization system presented in Paper D.

The two research tracks depicted in this chapter serve as a link towards the contributions of the PhD study, which is the topic of the next chapter. As the contributions are in the form of a collection of papers, the scientific and practical contributions will be detailed by referring to the specific articles.

Chapter 4 - Contributions

The combination of scientific and practical contributions is the strong point of this PhD research, presented in this chapter. The main body of the thesis comprises the papers A to E, which follow mainly the techniques for designing and implementing a strategic information system for the domain of electrical grids.

Following the description of the research tracks in Chapter 3, the papers will be categorized according to their scientific (research-oriented) and practical (enterprise-oriented) nature.

4.1 Scientific contributions

The main scientific contribution is to **investigate and experiment on the possibilities for designing and implementing a data-driven solution (information system) for the current electrical grid infrastructure**. By introducing intelligence in the current power system where planning and monitoring operations are broadly manual, the contribution is made towards the development of the so-called "smart grids".

4.1.1 Visualization systems

Paper A is an initial survey of visualization techniques for electrical grid smart metering data. The study focuses on the different techniques that contribute to obtaining a visualization system for monitoring and planning purposes, taking into consideration the real-time and historical data types as main use cases. As a part of the system design, the uses of the Common Information Model (CIM) in both research and industry fields are presented, given that CIM is currently utilized for defining the standard data model in electrical grids (ENTSO-E, Statnett). Due to the modern advances in the AMI networks and the variety in data issued by the small producers in the low-voltage grids, the paradigm of Big Data is introduced to further emphasize the need for efficient data analytics and visualization. Lastly, a survey of different visualization desktop tools justifies the choice of using Quantum GIS as data display tool.

The survey was made based on the initial project requirements for real-time and large volume (Big Data) historical data visualization. One concluding remark is related to the benefits and drawbacks of the different visualization software tools in the study, which is presented in Table 4.1. It has been

Table 4.1: Comparison of different desktop GIS tools.

Desktop GIS Software	**Advantages**	**Disadvantages**
ArcGIS	Receive real-time data from a wide variety of sources (GeoEvent Processor extension)	Proprietary (expensive license)
Quantum GIS (QGIS)	Open source, integration with other open source tools (GRASS, gvSIG), fast processing speed	Difficult to export files and to insert map elements
MapInfo	Track frequently updated data using (animation layer add in)	MapBasic implementation (not an accessible language)
GRASS	Modules for data management, spatial modelling and visualization	Inconvenient user interface, slow processing speed
gvSIG	User-friendly GUI, fast loading of large data volumes	Limited compatibility with open source tools (GRASS)
Maptitude	Good for basic GIS mapping purposes	Little support for advanced GIS processing

decided that QGIS fits best with the requirement for historical data management, which is why it has been utilized for modeling and evaluating the electrical grid topology, as well as for extracting labels related to the consumption measurements.

A Web-based GIS solution was chosen as proof of concept, which is presented in **Paper C**. This paper proposes an implementation of a near-real-time monitoring system for the DSOs. The study was partly done in collaboration with the DSOs from Thy-Mors Energi who helped identifying the case of household power outage. The working procedures for debugging this case have been defined based on interviews with the involved DSOs, who also demonstrated the use of different software tools within the company. The research-oriented solution in Paper C is done by emulating a real-life AMI (Virtual AMI) and by creating real-time voltage measurements through a replay functionality. Thus, some of the theoretical knowledge about stream processing was included as part of the scientific research. The implementation of a visualization-based monitoring system clearly reveals the advantages of spatial and situation awareness, leading to faster recovery and improved service to end consumers/prosumers. Also, the time-consuming operations are reduced, thus eliminating the risk of incorrect decisions during debugging.

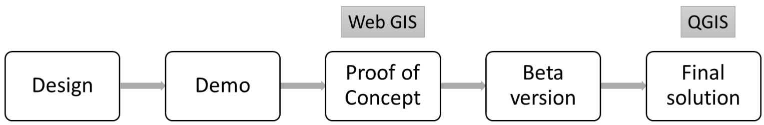


Fig. 4.1: Stages of development for the visualization system.

Figure 4.1 depicts the stages of development for the proposed visualization system. The aim for the final solution is to integrate the visual elements with their corresponding data analytics. Therefore, the appropriate choice for a data-driven operational system is QGIS, which was used for the analytical purpose in Paper E.

4.1.2 Data analytics methods

A more in-depth study of analytical methods is presented in **Paper B**, departing from the theoretical basis described in Section 2.3.2. The main motivation for this study was to identify the potential processes in an information system for low-voltage electrical grids and their suitable analytic techniques. Therefore, the use case of low-voltage grid observability is chosen. In this case, observability refers to making the knowledge available for the DSOs, either through explicit visualization or through mathematical or statistical models.

At the time Paper B was published, the design of the information system was still under definition, therefore the aim of this work is to propose a system design for the grid observability case. The proposal takes into consideration both streaming and historical data types, by evaluating the pros and cons offered by the different analytics. The concept of distributed system state estimation (DSSE [60]) is introduced as an analytic method for predicting values of missing or incorrect data. The flow of data in automatic and interactive events aims to justify the timing uncertainty in such a system, thus introducing the notion of near-real-time data, and the potential need for parallel in-memory and disk processing techniques.

Significant influence in the data analysis domain was brought by having access to real-life electrical grid data, in the form of GIS grid topology and time-series measurements, as explained in Section 3.2. The analysis of this data is presented in **Paper E**. The motivation for the study is closely related to the work done in Papers C and D, from which it was concluded that the behavior of small produces has a large impact on the DSOs operational workflows.

Based on this knowledge, Paper E depicts the study of the electrical grid small producers' behavior, with the focus on data accuracy analysis. The relevance of this study is strongly related to the real-life data flow, by determining unaccounted for information links. As the purpose is to evaluate what can be extracted from the available data, clustering methods have been applied to classify the different consumers by their energy consumption. The clusters were then used to test the accuracy of an ARIMA predictor.

The study in Paper E aims to evaluate the different contributions brought to the DSOs by the data analysis with respect to anomaly detection, power balancing, planning and monitoring operations. The future potential concerns for data inaccuracy originate from data scalability with the increasing development of AMIs, electrical vehicle mobility and the requirements for maintaining customers' privacy. Therefore, the work can be seen as a baseline for future development of information systems for power grids, as it will be further explained in Section 4.3.

4.1.3 Overall scientific contribution

The overall scientific contribution of the PhD study was brought in terms of visualization and data analytics systems for low-voltage electrical grids. Concurrently, these two support the design and implementation of a decision-based information system, which aims to take over the redundant grid operation tasks by the DSOs. In this way, more time can be dedicated towards planning, reinforcing and forecasting potential events that may induce instability in the power grid, thereby affecting the power quality at the customer end.

An initial version of such an information system is presented in the next section as part of the practical/on-site contributions towards the involved distribution company.

4.2 Practical contributions

Departing from the study made in Paper C, **Paper D** is an extension of that study. The focus of Paper D has been to perform a more in-depth analysis of the DSOs in their working environment through UX studies, as previously presented in Section 3.1. Besides the power outage case in Paper C, two more sequence models have been defined for the cases: light flickering at household level and customer notification of potential errors in the medium and high voltage levels.

Therefore, the implications of the study in Paper D are practical, oriented towards the needs and requirements of the distribution company.

4.2.1 Design of the visualization prototype

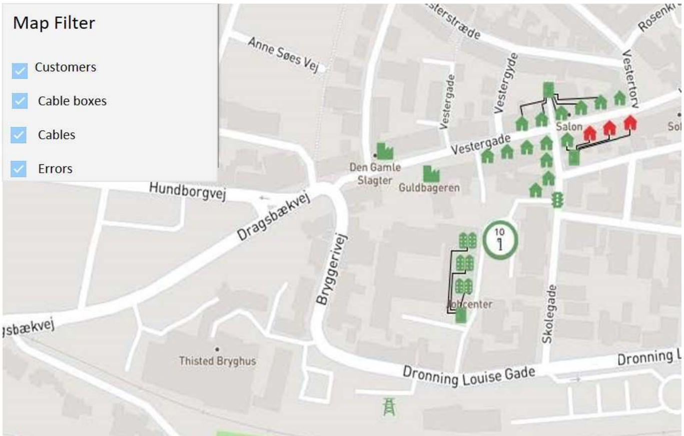


Fig. 4.2: Web-application map tab, including filters, customers, cable boxes and their interconnections

The design process is based on the user stories/sequence models identified by the DSOs and presented in Papers C and D. Spatial awareness is the key in designing this prototype, as the main visualization themes identified in Section 3.2.1 are all interlinked via a GIS map. Figure 4.3 shows the main items contained by the designed prototype: displaying all metering data points, statistics about households, cable boxes and substations, listing of the current reported errors, warnings and other kinds of alerts, and the low-voltage grid overview.

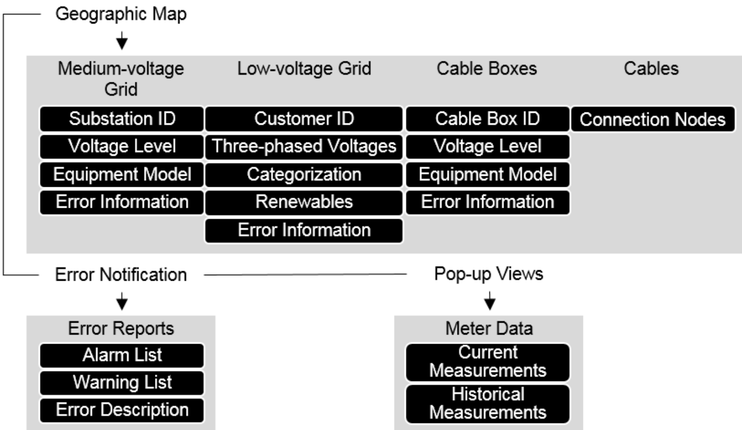


Fig. 4.3: Focus areas in the overall structure of the visualization system.

The challenge lies in achieving the most relevant knowledge from the visual perception, thus the focus is on the front-end design. It is expected that the large number of elements to be displayed leads to an overcrowding of the GIS map, therefore a visual event detection will not be trivial.

The contribution comes from the advantages of GIS: organizing data into layers and having the possibility to select/deselect a layer of interest. This feature was implemented as a filtering function, as well as the errors are sorted out and displayed based on their severity. The interactive map allows to choose the different components (household, cable box, transformer station, cable, renewable energy or customer type), as depicted by Figure 4.2.

In terms of workflow analysis, similarly to Paper C, it is shown that the different debugging processes are time-wise minimized and optimized so that more knowledge can be acquired from the existing system, due to the integration of a data-driven approach. Moreover, prediction methods are made possible through data integration, bringing the traditional data processing and analysis methods towards a prescriptive information system.

At the business level, the advantages materialize in the form of Capital Expenditures (CapEx) and Operating Expenses (OpEx), as it was previously shown in Figure 3.4.

4.3 Baseline for future development

This PhD's contributions in terms of data-driven analysis and visualization systems for low-voltage electrical grids opens up possibilities for future development. Particularly scalable solutions are of interest in the domain of smart grids, since it is expected that the volume and variety will increase with the advancements in the different utilized technologies. Some of the potential development tracks are identified as follows:

1. Hybrid processing techniques

   The data processing domain is evolving towards an intermix of stream and batch paradigms, purposely designing frameworks fully integrating both forms of processing introducing hybrid architectures and genuine hybrid processing engines [2] [63]. The development of hybrid processing engines is primarily fueled by stream technology maturing to a level capable of high performance as well as computational complexity [29] [79] [28]. As stream processors continue to evolve capabilities such as complete fault tolerance, fault recovery and producing accurate results, while delivering high performance computing, there is no longer an incentive to make a choice between fast or accurate results [75].

As the data system described in Section 3.2 is expected to evolve towards both stream and batch data types, hybrid processing techniques are architectures should be considered. Apache Spark and Apache Flink are two hybrid data processing frameworks [27] [84] [11]. Apache Spark [85] is the next generation framework for batch processing that also includes streaming capabilities. Speed (due to in-memory computation) and versatility (standalone cluster) are the pros of Apache Spark, while its limitations are due to high latency when processing large data streams and the high cost of running it in RAM. The Apache Flink open-source framework [37] is oriented towards distributed stream processing. It has advantages in terms of accuracy (delayed data), stateful and high throughput performance when scaling to thousands of nodes [80] [43]. However, its drawback is that Flink is not so widely deployed yet. An enterprise-oriented deployment of Flink for large scale networks would definitely bring a contribution towards stream processing in smart grids.

2. Focus on system optimization - predict and prescribe

   With the evolution of the data processing methods, it is expected that there will be higher requirements in terms of their corresponding real-time and historical analytics. The current power system is mostly based on the information extraction. However, due to scalability, data granularity and velocity challenges, the future analytical techniques will be more focused towards system optimization, previously referred to as predictions and prescriptions in Figure 2.4.

   Overall processing and analysis performance is a question of data availability and velocity, typically ranging from near instantly available to once or twice a day. The question is all about when data is processed and analyzed, which is determined by data and service time-degradation dependencies as real-time data expires continuously. Two of the most frequently referenced architecture frameworks implemented to support a combination of both batch and streaming workloads are Lambda and Kappa. Lambda [57] [71] [68] [83] is the target recommendation for the real-time data processing architecture, where batching is used as the primary processing method and streams are used to supplement early but unrefined results. Kappa [102] [81] may be considered for experimental and research purposes, in line with the Apache Flink framework [93]. The Kappa architecture contains only one stream processing layer making it easier to maintain due to its lower implementation complexity, as opposed to Lambda having separate processing layers for stream and batch.

3. Topology analysis using Graph Neural Networks

   Graph Neural Networks [82] is a concept which is currently used mostly in the domains of biology, chemistry and computer vision [104] [49] [50]. One possible research direction is to investigate their potential in the domain of electrical grids. By making use of the available medium and low-voltage grid topology, graph neural networks can help identifying how different topologies affect the losses in the power grid lines - the difference between measured powers at medium and low-voltage levels gives an indication of losses in the lines. Therefore, the starting point would be to first analyze the different topologies at medium-voltage and then extend the research to the low-voltage level.

   Case studies can be done by evaluating the secondary substations with the highest number of connected consumers and meshed connections. The contribution of this study consists of proposing different ways of organizing the power flow in the system and thus performing an analysis of the power system reliability.

Chapter 5 - Conclusion

The overall theme of this PhD study was **automation of Smart Grid operations through spatio-temporal data-driven systems**. As the Danish climate regulations aim to introduce 100% green energy by year 2050, the increasing number of small producers in the low-voltage electrical grid challenges the DSOs daily operations for delivering a reliable electricity supply. Therefore, the research in this PhD was focused on investigating how to **design and develop an automatic decision-support system for the DSOs via efficient data processing, analysis and visualization of smart metering data**. To achieve this, the scope of this work was focused on exploring means to develop a suitable information system for the low-voltage electrical grid, based on user experience studies. The main contribution was realized with the **design and implementation of a data-driven system for the DSOs' daily tasks**, using the existing operational system as baseline for the research.

The outcome has been evaluated both from a scientific and from a practical implementation point of view. On the **scientific** level, the performed experiments regarding system interfacing and and data analysis were examined from both near-real-time and historical data perspective. These demonstrate how efficient data analytics as part of the integrated system opens up for a wider spectrum of opportunities for the DSOs, than with the existing system. Analytical techniques such as statistics, forecasting and visualization have shown to bring deep insight into the consumption behavior of the small producers, enabling the DSOs to make documented decisions about the future grid changes. In other words, smartness in the current electrical grid operation is brought by evolving towards prescriptive types of actions.

An important part in this research has been **data visualization**, designed as the final front-end solution. Thanks to input provided by the DSOs from Thy-Mors Energi regarding their daily operations, it was possible to perform user experience studies. These studies have shown that manual error troubleshooting restraints the DSOs from committing to more advanced operational tasks. Thereby, the user studies have had a big impact in the design and implementation of the visualization solution, which proved to minimize the current error debugging time. At the same time, the advantages of integrating data analytics into one system aid to improve grid monitoring, planning and prediction of events. Detecting and predicting errors automatically beforehand thereby upgrades the current manual debugging process.

The proof-of-concept analysis the scientific level provided means to experiment interfacing between the different chosen tools - GIS, DBMS and analytics, in order to develop the desired data-driven system. A step forward was taken with respect to the **practical contribution** of the work, by creating a **visualization prototype** dedicated for the DSOs involved in the study. From this, it was concluded that the spatial and situation awareness provided by the GIS capabilities can help the DSOs cut down on some of their repetitive working procedures when addressing different kinds of errors.

Considering the current climate regulations and the impact of renewables on the energy supply's reliability, this PhD work has overall contributed to demonstrating the **advantages of automatic spatio-temporal information system integration** for the low-voltage grid operation. It is shown that even though the initial cost of operating is higher than operating the current system, more functionality help the DSOs to prepare for the future advances in the electrical grid. In the long run, the return of investment will be collectively acquired from time saving operations, increased operation efficiency due to the new system's intelligence and a flatter trend in the OpEx.

By focusing on the future smart grid capabilities, there will be less investments spent on adapting the already-existing operational system to future challenging use cases. Simultaneously, the DSOs can take over exciting tasks which involve new business areas, potentially increasing productivity and bringing value towards general grid operation, management and to the society.

References

[1] Benchmarking big data architectures for social networks data processing using public cloud platforms. *Future Generation Computer Systems*, 89:98 – 109, 2018.

[2] Zaia Alimazigui Abdeslam Mehenni1 and Mohamed Ahmed-Nacer. Decreasing data analytics time: Hybrid architecture mapreduce-massive parallel processing for a smart grid. *Journal of Electrical Systems*, pages 74 – 85, 2017.

[3] Fadi Al-Turjman and Mohammad Abujubbeh. Iot-enabled smart grid via sm: An overview. *Future Generation Computer Systems*, 2019.

[4] Dieter Gantenbein Anders Holm Foosnæs, Claus Amtrup Andersen; Linda Christensen and Ningling Rao. *EDISON – Study on optimal grid integration of electric vehicles*. DTU Library, 2011.

[5] Artemakis Artemiou. Introducing sql server in-memory oltp, `https://www.red-gate.com/simple-talk/sql/learn-sql-server/introducing-sql-server-in-memory-oltp/`, 2015.

[6] Thy-Mors Energi Elnet A/S. Beskrivelse af forskningsprojekter i elnettet, `https://www.tme-elnet.dk/om-os/forskningsprojekter`.

[7] Shivanand Balram, Balram Shivanand, Suzana Dragiâceviâc, and Suzana Dragicevic. *Collaborative Geographic Information Systems*. 2006.

[8] Mininath Bendre and Ramchandra Manthalkar. Time series decomposition and predictive analytics using mapreduce framework. *Expert Systems with Applications*, 116:108 – 120, 2019.

[9] Paul De Hert; Vagelis Papakonstantinou; Gianclaudio Malgieri; Laurent Beslay and Ignacio Sanchez. The right to data portability in the gdpr: Towards user-centric interoperability of digital services. *Computer Law and Security Review*, 34(2):193 – 203, 2018.

[10] Simon Elias Bibri. *Smart Sustainable Cities of the Future*. 2018.

[11] Ying Zhou Bilal Akil and Uwe Röhm. Technical report: On the usability of hadoop mapreduce, apache spark and apache flink for data science. Technical report, Mar 28 2018.

[12] M. Boehm, A. Kumar, J. Yang, and H. V. Jagadish. *Data Management in Machine Learning Systems*. Morgan and Claypool, 2019.

[13] Henrik Breum. The Danish Energy Model. Innovative, efficient and sustainable. Technical report, The Danish Energy Agency.

[14] Wayne Cascio and Ramiro Montealegre. How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 3:349–375, 03 2016.

[15] Richard Charnah. What exactly is a smart grid? a smart electrical grid has the potential to play a key role in the effort to lower energy costs for consumers, reduce energy dependency and cut greenhouse gas emissions. how can all of this be realized and what can governments do to ensure such a grid can be mode a reality? dr. richard charnah, technology director of areva t&d uk, explains.(smart grids). 17:60, 2009.

[16] René J. Chevance. 8 - dbms and server architectures. In René J. Chevance, editor, *Server Architectures*, pages 549 – 584. Digital Press, Burlington, 2005.

[17] Timothy C Coburn. Geographic information systems: an introduction: 2nd ed. tor bernhardsen. *Computers and Geosciences*, 26(7):853 – 854, 2000.

[18] Peter Dale. Information technology in geography and planning: Principles of geographical information systems: Ian bracken and christopher webster routledge, london, uk, 1990, 444 pp. *Cities*, 8(1):73, 1991. Special Feature: GIS and urban management.

[19] dataArtisans. da platform stream processing for real-time business powered by apache flink. Technical report, 2018.

[20] Ganesh Chandra Deka. Chapter nine - nosql polyglot persistence. In Pethuru Raj and Ganesh Chandra Deka, editors, *A Deep Dive into NoSQL Databases: The Use Cases and Applications*, volume 109 of *Advances in Computers*, pages 357 – 390. Elsevier, 2018.

[21] Werner Dubitzky. *Data mining in grid computing environments*. 2008.

[22] Siddhartha Duggirala. Chapter two - newsql databases and scalable in-memory analytics. In Pethuru Raj and Ganesh Chandra Deka, editors, *A Deep Dive into NoSQL Databases: The Use Cases and Applications*, volume 109 of *Advances in Computers*, pages 49 – 76. Elsevier, 2018.

[23] Energinet. Electricity Security of Supply Report. Technical report, 2017.

[24] Energinet. Energinet's analysis assumptions. Technical report, 2017.

[25] Hueske Fabian and Ewen Stephan. Stateful stream processing with apache flink, `https://www.infoworld.com/article/3267966/big-data/stateful-stream-processing-with-apache-flink.html`, 2018.

[26] Zbigniew J. Gackowski. The helix of human cognition: knowledge management according to dikw, e2e, and the proposed view. 15:93–, 2012.

[27] García-Gil, Diego García Gil, Sergio Ramírez Gallego, Salvador García, and Francisco Herrera. A comparison on scalability for batch big data processing on apache spark and apache flink. 2:1–11, 2017.

[28] Jana Giceva and Mohammad Sadoghi. *Hybrid OLTP and OLAP*, pages 1–8. 01 2018.

[29] Corinna Giebler, Christoph Stach, Holger Schwarz, and Bernhard Mitschang. Braid: A hybrid processing architecture for big data. In *Proceedings of the 7th*

International Conference on Data Science, Technology and Applications, DATA'18, pages 294–301. SCITEPRESS, July 2018.

[30] Daniel Goldberg, Miriam Olivares, Zhongxia Li, and Andrew G. Klein. Maps&gis data libraries in the era of big data and cloud computing. 10:100–122, 2014.

[31] Venkat N. Gudivada. Chapter 2 - data analytics: Fundamentals. In Mashrur Chowdhury, Amy Apon, and Kakan Dey, editors, *Data Analytics for Intelligent Transportation Systems*, pages 31 – 67. Elsevier, 2017.

[32] Iris Reinhartz-Berger; Jens Gulden; Selmin Nurcan; Wided Guédria and Palash Bera. *Enterprise, Business-Process and Information Systems Modeling*. Springer, Cham, 2017.

[33] Angappa Gunasekaran, Gunasekaran, and Maqsood Sandhu. *Handbook On Business Information Systems*. 2010.

[34] Gurusamy, Vairaprakash Gurusamy, S. Kannan, and K. Nandhini. The real time big data processing framework advantages and limitations. 5:305–312, 2017.

[35] John Hagerty. 2017 planning guide for data and analytics. Technical report, Gartner, Inc., 2016.

[36] Benoît Bizet Hans Peter Slente, Peter Alexandersen and Torsten Hasforth. Wind Energy Moving Ahead. How wind energy has changed the Danish energy system. Technical report, Think Denmark, July 2017.

[37] Paris Carbone; Asterios Katsifodimos; Stephan Ewen; Volker Markl; Seif Haridi and Kostas Tzoumas. Apache flink™: Stream and batch processing in a single engine. *IEEE Data Eng. Bull.*, 38:28–38, 2015.

[38] Richard Heinberg and David Fridley. *Our Renewable Future. Laying the Path for One Hundred Percent Clean Energy*. June 2016.

[39] Karen Holtzblatt and Hugh Beyer. 10 - the bridge from data to design: The wall walk. In Karen Holtzblatt and Hugh Beyer, editors, *Contextual Design (Second Edition)*, Interactive Technologies, pages 253 – 275. Morgan Kaufmann, Boston, second edition edition, 2017.

[40] Karen Holtzblatt and Hugh Beyer. 7 - building experience models. In Karen Holtzblatt and Hugh Beyer, editors, *Contextual Design (Second Edition)*, Interactive Technologies, pages 147 – 206. Morgan Kaufmann, Boston, second edition edition, 2017.

[41] M.S. Hossain, N.A. Madlool, N.A. Rahim, J. Selvaraj, A.K. Pandey, and Abdul Faheem Khan. Role of smart grid in renewable energy: An overview. *Renewable and Sustainable Energy Reviews*, 60:1168 – 1184, 2016.

[42] H. Hu, Y. Wen, T. Chua, and X. Li. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2:652–687, 2014.

[43] Fabian Hueske. How to build stateful streaming applications with apache flink. 2018.

[44] Abdulaziz Almehmadi James Curzon and Khalil El-Khatib. A survey of privacy enhancing technologies for smart cities. *Pervasive and Mobile Computing*, 55:76 – 95, 2019.

[45] Mark Jeffery. Return on investment analysis for e-business projects. 2003.

[46] Shibily Joseph, Jasmin E.A., and Soumya Chandran. Stream computing: Opportunities and challenges in smart grid. *Procedia Technology*, 21:49 – 53, 2015. SMART GRID TECHNOLOGIES.

[47] Kamstrup. Electricity meter production, `https://www.kamstrup.com/en-en/electricity-solutions/smart-electricity-meters/electricity-meter-production`.

[48] Michael D. Kennedy. *Introducing Geographic Information Systems with ArcGIS : A Workbook Approach to Learning GIS*. 2013.

[49] Jure Leskovec Keyulu Xu, Weihua Hu and Stefanie Jegelka. How powerful are graph neural networks? *CoRR*, abs/1810.00826, 2018.

[50] Nicholas Choma; Federico Monti; Lisa Gerhardt; Tomasz Palczewski; Zahra Ronaghi; Prabhat; Wahid Bhimji; Michael M. Bronstein; Spencer R. Klein and Joan Bruna. Graph neural networks for icecube signal classification. *CoRR*, abs/1809.06166, 2018.

[51] Dariusz Król, Ngoc Thanh Nguyen, and Kiyoaki Shirai. *Advanced Topics in Intelligent Information and Database Systems*, volume 710. 2017.

[52] Katerina Lepenioti, Alexandros Bousdekis, Dimitris Apostolou, and Gregoris Mentzas. Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50:57 – 70, 2020.

[53] Clinton J. Levitt and Anders Sørensen. *The Cost of Producing Electricity in Denmark*. The Rockwool Foundation, November 2014.

[54] X. Liu, A.V. Dastjerdi, and R. Buyya. Chapter 8 - stream processing in iot: foundations, state-of-the-art, and future directions. In Rajkumar Buyya and Amir Vahid Dastjerdi, editors, *Internet of Things*, pages 145 – 161. Morgan Kaufmann, 2016.

[55] Zhenya Liu. Chapter 6 - innovation in global energy interconnection technologies. In Zhenya Liu, editor, *Global Energy Interconnection*, pages 239 – 272. Academic Press, Boston, 2015.

[56] Zhihui Lu, Nini Wang, Jie Wu, and Meikang Qiu. Iotdem: An iot big data-oriented mapreduce performance prediction extended model in multiple edge clouds. *Journal of Parallel and Distributed Computing*, 118:316 – 327, 2018. Special Issue on Advanced Algorithms and Applications for IoT Cloud Computing Convergence.

[57] M. Iacono M. Gribaudo and M. Kiran. A performance modeling framework for lambda architecture based applications. *Future Generation Computer Systems*, 86:1032 – 1041, 2018.

[58] Z. Ma, S. Sommer, and B. N. Jørgensen. The smart grid impact on the danish dsos' business model. In *2016 IEEE Electrical Power and Energy Conference (EPEC)*, pages 1–5, Oct 2016.

[59] S. T. Mak and E. So. Integration of pmu, scada, ami to accomplish expanded functional capabilities of smart grid. In *29th Conference on Precision Electromagnetic Measurements (CPEM 2014)*, pages 68–69, Aug 2014.

[60] Ruben Sánchez Martin-Loeches, Florin Iov, Mohammed Seifu Kemal, Maria Stefan, and Rasmus Løvenstein Olsen. Observability of low voltage grids: actual dsos challenges and research questions. In *Proceedings of the 2017 52nd International Universities' Power Engineering Conference (UPEC)*. IEEE Press, 8 2017.

[61] Patrick McGlamery. *Library and Information Science Research*, 27(3):418 – 419, 2005.

[62] S. E. McHann. Grid analytics: How much data do you really need? In *2013 IEEE Rural Electric Power Conference (REPC)*, pages C3–1–C3–4, April 2013.

[63] Abdeslam Mehenni, Zaia Alimazighi, Tarek Bouktir, and Mohamed Ahmed-Nacer. An optimal big data processing for smart grid based on hybrid mdm/r architecture to strengthening re integration and ee in datacenter. 2018.

[64] Express Metrix. Quantifying ROI: Building the Business Case for IT and Software Asset Management. Technical report.

[65] Z. Milosevic, W. Chen, A. Berry, and F.A. Rabhi. Chapter 2 - real-time analytics. In Rajkumar Buyya, Rodrigo N. Calheiros, and Amir Vahid Dastjerdi, editors, *Big Data*, pages 39 – 61. Morgan Kaufmann, 2016.

[66] Amit Kumar Mishra. A dikw architecture for cognitive engineering. *Procedia Computer Science*, 123:285 – 289, 2018. 8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017 (Eighth Annual Meeting of the BICA Society), held August 1-6, 2017 in Moscow, Russia.

[67] Thomas Brandstetter Morta Vitunskaite, Ying He and Helge Janicke. Smart cities and cyber security: Are we there yet?a comparative study on the role of standards, third party risk management and security ownership. *Computers and Security*, 83:313 – 331, 2019.

[68] A. A. Munshi and Y. A. I. Mohamed. Data lake lambda architecture for smart grids big data analytics. *IEEE Access*, 6:40463–40471, 2018.

[69] Amr A. Munshi and Yasser A.-R. I. Mohamed. Big data framework for analytics in smart grids. *Electric Power Systems Research*, 151:369 – 380, 2017.

[70] A. Mutule, A. Obushevs, A. Lvov, H. Segerberg, N. Shao, S. You, R. Bacher, H. Sauvain, J. Remund, U. Schröder, and A. J. Engen. Efficient identification of opportunities for distributed generation based on smart grid technology. In *IEEE PES ISGT Europe 2013*, pages 1–5, Oct 2013.

[71] Thanisa Numnonda. A real-time recommendation engine using lambda architecture. *Artificial Life and Robotics*, 23(2):249–254, Jun 2018.

[72] CPH Post Online. Hurricane-force storm crossing denmark, `http://cphpost.dk/news/national/hurricane-force-storm-crossing-denmark.html`.

[73] M. Ourahou, W. Ayrir, B. EL Hassouni, and A. Haddi. Review on smart grid control and reliability in presence of renewable energies: Challenges and prospects. *Mathematics and Computers in Simulation*, 2018.

[74] Trine Pallesen and Peter Holm Jacobsen. Solving infrastructural concerns through a market reorganization: A case study of a danish smart grid demonstration. *Energy Research and Social Science*, 41:80 – 88, 2018. Energy Infrastructure and the Fate of the Nation.

[75] Mrutyunjaya Panda. Intelligent data analysis for sustainable smart grids using hybrid classification by genetic algorithm based discretization. *Intelligent Decision Technologies*, 11, 02 2017.

[76] S. Perera. Large scale data processing in real world: From analytics to predictions. In *2014 14th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 8–8, Dec 2014.

[77] Pries, Kim H. Pries, and Robert Dunnigan. *Big Data Analytics*. 2015.

[78] Alexander Maria Prostejovsky. *Increased Observability in Electric Distribution Grids. Emerging Applications Promoting the Active Role of Distribution System Operators*. PhD thesis, Technical University of Denmark, 2017.

[79] Xiangnan Ren and Olivier Curé. Strider: A hybrid adaptive distributed rdf stream processing engine, May 16 2017.

[80] Paris Carbone; Stephan Ewen; Gyula Fora; Seif Haridi; Stefan Richter and Kostas Tzoumas. State management in apache flink? 10:1718–1729, 2017.

[81] Iman Samizadeh. A brief introduction to two data processing architectures - lambda and kappa for big data, March `https://towardsdatascience.com/`, 2018.

[82] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, Jan 2009.

[83] Amazon Web Services. Lambda architecture for batch and stream processing. Technical report, October 2018.

[84] Ashansa Perera Shelan Perera and Kamal Hakimzadeh. Reproducible experiments for comparing apache flink and apache spark on public clouds. *CoRR*, abs/1610.04493, 2016.

[85] Matei Zaharia; Reynold S. Xin; Patrick Wendell; Tathagata Das; Michael Armbrust; Ankur Dave; Xiangrui Meng; Josh Rosen; Shivaram Venkataraman; Michael Franklin; Ali Ghodsi; Joseph Gonzalez; Scott Shenker and Ion Stoica. Apache spark. 59:56–65, 2016.

[86] Anshu Shukla, Shilpa Chaturvedi, and Yogesh Simmhan. Riotbench: An iot benchmark for distributed stream processing systems. 29:e4257–, 2017.

[87] Krishnarajanagar G. Srinivasa and Srinidhi Hiriyannaiah. Chapter five - comparative study of different in-memory (no/new) sql databases. In Pethuru Raj and Ganesh Chandra Deka, editors, *A Deep Dive into NoSQL Databases: The Use Cases and Applications*, volume 109 of *Advances in Computers*, pages 133 – 156. Elsevier, 2018.

[88] Maria Stefan, Jose G. Lopez, and Rasmus 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)*, pages 3543–3548, 2018.

[89] Sheila Lakshmi Steinberg and Steven J. Steinberg, Sheila Lakshmi; Steinberg. *GIS Research Methods*. 2015.

[90] Ida Strand. 2050. Something's Green in the State of Denmark. Scenarios for a sustainable economy. Technical report, 2012.

[91] Jesper Tornbjerg. *Green Energy for All*. KLS Grafisk Hus A/S, November 2014.

[92] Lamberto Tronchin, Massimiliano Manfren, and Benedetto Nastasi. Energy efficiency, demand side management and energy storage technologies – a critical analysis of possible paths of integration in the built environment. *Renewable and Sustainable Energy Reviews*, 95:341 – 353, 2018.

[93] Kostas Tzoumas and Stephan Ewen. Real-time stream processing: The next step for apache flink, May `https://www.confluent.io/blog/real-time-stream-processing-the-next-step-for-apache-flink/`, 2015.

[94] K. R. Ishwarya V. Vidhya, G. Jeyaram. *Database management systems*. 2016.

[95] Shaohua Wang, Yang Zhong, and Erqi Wang. An integrated gis platform architecture for spatiotemporal big data. *Future Generation Computer Systems*, 94:160 – 172, 2019.

[96] Chris Webb. Will the european grid smarten up? 15:20–21,23–24, 2007.

[97] The Microsoft website. Sql server 2016, `https://www.microsoft.com/en-us/sql-server/sql-server-2016/`.

[98] Wingerath, Wolfram Wingerath, Norbert Ritter, and Felix Gessert. *Real-Time & Stream Data Management*. 2019.

[99] Q. Wu, E. Larsen, K. Heussen, H. Binder, and P. Douglass. Remote off-grid solutions for greenland and denmark: Using smart-grid technologies to ensure secure, reliable energy for island power systems. *IEEE Electrification Magazine*, 5(2):64–73, June 2017.

[100] Fatos Xhafa. *Modeling and Processing for Next-Generation Big-Data Technologies*, volume 4. 2015.

[101] Vinit Yadav. *Processing Big Data with Azure HDInsight*. 2017.

[102] Jared Zagelbaum. A different way to process data: Kappa architecture, January `https://www.blue-granite.com/blog/a-different-way-to-process-data-kappa-architecture`, 2018.

[103] Kenneth K. Zame, Christopher A. Brehm, Alex T. Nitica, Christopher L. Richard, and Gordon D. Schweitzer III. Smart grid and energy storage: Policy recommendations. *Renewable and Sustainable Energy Reviews*, 82:1646 – 1654, 2018.

[104] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Yang Cheng, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications, Mar 07 2019.

Part II

Papers

Paper A

Visualization Techniques for Electrical Grid Smart Metering Data: A Survey

Maria Stefan, Jose G. Lopez, Morten H. Andreasen
and Rasmus L. Olsen

The paper has been published at the:
2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)

Abstract

One of the considerable initiatives towards creating a smart society could be the guarantee of a smart, resilient and reliable power grid. As an attempt to improve the electricity supply service, it would be meaningful for the distributed system operators (DSOs) to be able to monitor the current status of the grid. The prediction of future possible critical situations would then be feasible using the available information, whereas, based on historical data, further grid expansion and reinforcement may be planned. A proper presentation and visualization of the near-real time metering data may constitute the baseline for bringing improvements to the power grid. This paper presents an approach to build an efficient visualization system so that the extracted smart meters information can be used in a meaningful manner. An overview of the use cases related to the visualization features is first presented, as a motivation for the choice of the relevant state of the art research. In relation to the knowledge provided by the metering data, a definition of the big data concept will be further introduced, according to the requirements established by the project definition. Geographic Information System (GIS) tools are useful to help visualize the collected big data in near-real time. For this reason, a survey of existing GIS software will be made so that the choice of the most suitable tool can be justified. Also, the integration of GIS technologies into the Common Information Model (CIM) aims to improve the visualization efficiency. As a consequence, investigating methods for adapting CIM standards to the GIS platform are also important.

I Introduction

In the process of the development of a more efficient electrical grid, the concept of the so-called "smart grids" has emerged. The purpose is to create an affordable, reliable and sustainable electricity supply. As a consequence of the development of smart grids, the distributed system operators (DSOs) in the Danish electricity distribution grid are facing operational challenges due to a large number of new smart electronic devices. These devices load the producer utilities with a high amount of data, reporting issues such as cable and converter faults, voltage magnitude outside standard limits and network congestion [13]. In order to address these challenges, intelligent features are required so that the DSOs can obtain an overview of their low voltage grid. This would allow the execution of *near real-time* daily operations in the grid, as well as long term grid management and planning. The efficiency of the electrical grid could be improved through the collection, processing and analysis of data and the outcome would have people as the main beneficiary. Eventually, the grid's efficiency would be characterized by user satisfaction, economic implications, population reach etc. In the long run, the pursuit of progress in public and private sectors constitutes an initiative to create a smart society. In the current paper different methods are investigated towards building such a visualization system (GIS tools, CIM modeling, implementation languages). The uniqueness of this research consists of experiments that correlate different methods to achieve the desired visualization.

Current electricity grids are the baseline for future grids, which could account for the changes in innovative technologies, customer needs, environmental issues and increasing network congestion [10] [20]. The future distribution grids' architecture is evolving from a one-direction power flow towards a bi-directional flow between suppliers and consumers, according to the European Commission's view on electricity systems. The aim is to create a customer-oriented electricity system that will be flexible, accessible, reliable and economic [10].

Traditional power systems are developing towards *digitalization*, with the emerging ICT (Information and Communication Technology). In [27] digitalization is defined as a key element in the further development of the power system, being able to provide an efficient time-critical monitoring of the grid state, where issues would be signaled through the display of alarms.

In the energy sector, large amounts of data are accumulated daily. The main source of data in a smart grid is the *adaptive metering infrastructure* (AMI), where a large number of smart meters are deployed at the user end side [32]. Mechanisms for collecting data from the smart meters are addressed in [18] for the real-time state estimation of the grid. They are advantageous because they can facilitate the power flow control and identify exceeding limits of current and voltages. In other words, data collection mechanisms offer some degree of certainty of data quality. Correctness, completeness and timely data are some of the main attributes that can describe the data quality and it is important that they are ensured prior to its processing and viewing.

Three main steps need to be accomplished in order to obtain the data visualization platform, as shown in Figure A.1: from storing data in a database to the "human eyes". Establishing efficient ways of transforming smart meters raw data into meaningful information can contribute to the operation, management and planning of the power grid.

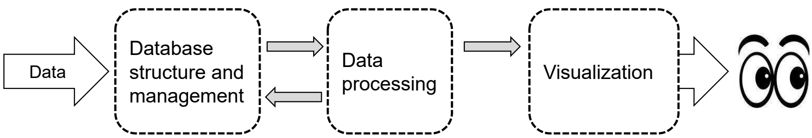


Fig. A.1: Data flow perspective.

A systematic storage of smart grid data is possible using *database structures*. As it is presented in [19], database architectures and methods also allow the processing and analysis of large amounts of data, continuously aggregated by time flow.

A more accurate representation of real-time data is closely related to the delays involved in data transmission from the smart meters to the visualization platform. The delays involved in the delimitation of the "real-time" definition can be defined as the transmission and the processing delays, represented in Figure A.2. The transmission delay (smart meter - database) is bound to the adaptive data collection mechanisms presented in [18]. The data processing delay (database - visualization) is the aspect to be analyzed in this research and it will be used for defining the use case related to the real-time display of information.

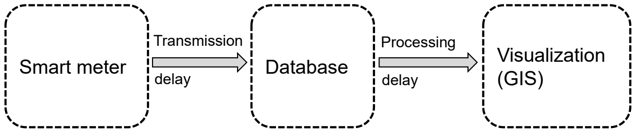


Fig. A.2: Delays corresponding to grid state estimation and data processing.

The dynamics of information impacts the visualization in the sense that information that changes rarely over time can be treated differently. More processing time can be spend on such information compared to very volatile information elements that often change values, meaning that it is highly dependent on the information granularity.

The outlined challenges lead therefore to investigating scalable and secure IT infrastructures for real-time events management in the smart grid. Extracting significant information is important for the decision-making process and this is intended to be made through *big data analytics*. The challenge of the big data research is to include a set of technologies that would ensure users' privacy, but still extract the valuable information needed for real-time grid operations and long term scale planning.

A possible implementation of data analytics can be done via the open standard Common Information Model (CIM). This model facilitates exchange of power system network data between companies and allows data exchange between applications within a company; thus, easier implementation of data analytics. Its purpose is to increase reliability and reduce expenses in smart grid infrastructures, as shown in [8].

The paper is organized as follows: Section II introduces the use cases concerning the real-time and historical features of the visualization platform. In Section III the motivation for choosing the CIM standard is presented, along with its newest uses and challenges both in research and industry. Section IV describes the related work regarding: big data use cases, visualization using GIS software and their relation to big data analytics. Furthermore, a definition of big data will be explained here in relation to the previously-mentioned use cases. In Section V different GIS tools are evaluated to motivate the choice of the most suitable one for carrying out the project.

II Data Visualization Use Cases

Data collected from the smart meters is a raw set of facts, without any particular meaning or usability. The process of how knowledge is obtained out of raw data is presented in the DIKW diagram shown in Figure A.3.

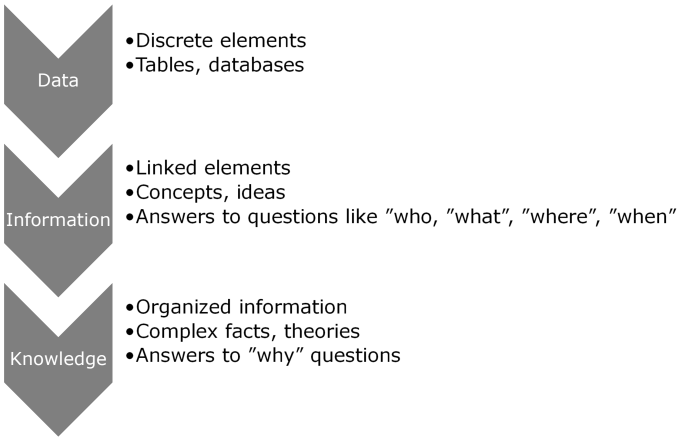


Fig. A.3: DIKW Diagram.

In order to obtain valuable information out of raw data, some management tools have to be established. First of all, database systems are a fundamental tool for storing any kind of data, for later processing and analysis. The processing of data aims to provide the proper and timely input to the future visualization platform, moving the geospatial data to near-real-time access [12].

Information is the outcome of data processing, whose patterns and facts can be analysed to determine what is needed in the development of the visualization system. It becomes *knowledge* when it is applied in a particular situation to answer questions such as "why" and "how".

In order to establish the relevant state of the art survey, some use cases have to be defined. The visualization features can be divided into: time-critical (grid monitoring) and non-time critical (grid history).

A. Grid Monitoring

The real-time visualization features of the system are highly dependent on how often data is received and stored in the database and the events involved in the real-time data processing. Hence, one has to define what data needs to be processed for each visualization feature, to achieve the minimum processing delays in this context. At the same time, the visualization features need to be scalable according to the grid size and the volume of data stored in the database.

The delay in data processing is time-critical for the monitoring system. It has to be minimized taking into account: the amount of data required to make qualified decisions, performance cost of data monitoring and the triggered events, and the amount of time required for the necessary data to be received.

Example: Voltage drop/rise

With the current technology it is possible to display the voltage magnitude of the system. The voltage drop is an issue that has to be adjusted and it can be done by monitoring the voltage stability margin, as explained in [20].

The overall quality of the distribution power grid can also be affected by voltage unbalance (phase inbalance) [7]. Furthermore, voltage rise (due to large amount of distributed generation power output) is presented in [3] as one of the main challenges for the DSOs to regulate the voltage levels in the grid.

Another useful feature that enables operators to quickly identify the fault location is be the ability to display real-time alarms in the visualization system [20]. This could be achieved using dedicated visualization features of GIS tools, as presented in Table A.1. For example, the ArcGIS GeoEvent Processor makes it possible to track dynamic data, which changes location frequently. Likewise, MapInfo's Animation Layer add in is utilized for applications where data features update constantly.

The difference between the time stamp where the first suspicious event is received and how big is the delay to the actual alarm visualization should be investigated as well.

B. Grid History

The non-time-critical features are related to the electrical grid planning, to create models for future state estimates. Therefore, it should be established what parameters to keep track of, which data should be passed on to the visualization system and how to process data so that the database visualization system can be optimized. Data analytics and data mining techniques can help in discovering data patterns, that can be later useful in future grid planning in a quicker and more efficient manner.

Example: Power balancing

An example of how tracking historical data helps providing a better understanding of the grid state is presented in [1], for forecasting electrical loads. Statistical analysis can be created and further used to evaluate grid operating conditions and thus, failure assessment.

The collected historical data contains information about grid events, for example frequent oscillations in the electrical load power. Important information related to power failures can be extracted using historical data, while data mining techniques can provide an in-depth explanation for the failure cause. Parameters such as temperature, weather conditions and electrical load may be considered in the data analysis. Making use of historical data creates the foundation for a more efficient future grid planning, creating the opportunity to avoid power failures.

III CIM Modeling

The initial motivation for the development of the CIM was driven by the increasing requirements for Energy Management Systems (EMS) to enhance upgradability, scalability and interoperability. Ambitious European energy and climate goals dictate the future and will change the very nature of the power system, promoting cooperation between European utility enterprises (DSOs) in order to support the implementation of the EU energy policy. Special focus is on the integration of Renewable Energy Sources (RES), which expands distributed power generation, and the Internal Energy Market (IEM) to meet the EU's energy policy objectives of affordability, sustainability and security of supply. This will invariably increase the data exchange necessities both internal and between the European utility enterprises and hence the relevance and need of a common semantic framework like CIM [31].

The use of ICT and advances in the design of power systems result in an increase and variation of the data. Its analysis and recognition require advanced data mining techniques, which are however heavily restricted by insufficient description of the data models. These issues lead the development from a vendor hard- and software specific API to a focus on common semantic and syntax models for data exchange between the EMS database and applications, which advanced into the CIM architecture comprising several standards under the International Electrotechnical Commission (IEC) Technical Committee 57 (TC57) and associated workgroups (WG) [16].

The integration of CIM has been adopted by Statnett, Norway's national main grid owner and operator. The future Statnett vision for CIM integration is centered on the perceived benefits in conjunction with predicted necessities for next generation smart grid development. CIM is envisioned to provide

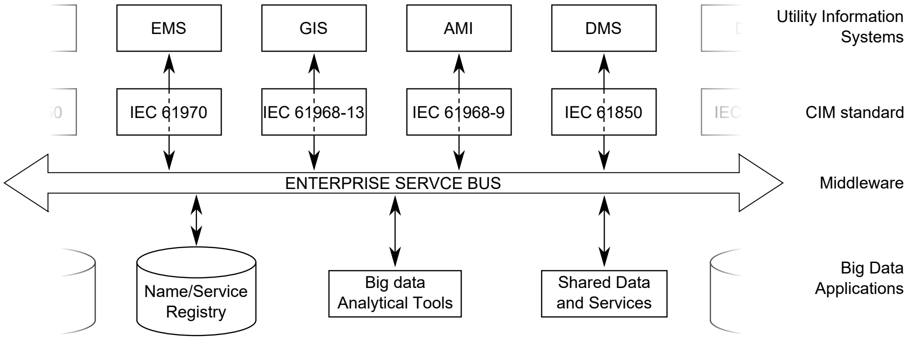


Fig. A.4: CIM-based utility big data integration [33].

One common Power System Model (PSM), delivering a complete enterprise service oriented integration that is adaptable to future requirements; introducing a standard for data exchange via a common Enterprise Service Bus (ESB) (CIM EAI message bus - Figure A.4), capable of handling all data governance and data management services. As part of the development of the future power system, Statnett is undertaking pilot projects analyzing the potential and performance of smart grid technologies and communication systems. One such project includes Demand Side Response (DSR) load control via AMI, where CIM was utilized as the standard for data exchange between the distribution management system (DMS) and the AMI front end, utilizing CIM XML messaging (DSO - AMI front end; IEC 61968-9) [30].

The European Network of Transmission System Operators (ENTSO-E) is a collaboration between 42 European TSO's representing 35 countries, including Norwegian Statnett and Energinet.dk of Denmark. By the same token, smart meter penetration of European consumers, i.e. households and Small and Medium Enterprises (SMIs), is currently on the rise and is expected as a continuous trend for the near future. ENTSO-E utilize CIM IEC standards to provide common data exchange formats to ensure compatibility for the various information sharing between transmission system operators (TSOs), third parties and service providers alike. Direct cooperation is made with the IEC workgroups responsible for CIM for transmission (IEC TC57 / WG13) and CIM for energy markets (IEC TC57 / WG16) securing TSO influence and compliance, and supporting the continuous development of CIM [14].

The uses and challenges of CIM in the research area are addressed in [16], which deals with data driven interactive visualization of power systems. CIM based model visualization is done via data manipulation algorithms based on empirical or mathematical derived utility data. The main benefit is to provide smart grid operation and analytics decision support tools, enabling electric system operators and analysisists to perpetually monitor big data information and events by images, diagrams, animations promoting communication and interpretation with enhanced pattern recognition. Furthermore, emerging business requirements of the electric power industry gain from the power systems data visualization.

Complete GIS and Supervisory Control and Data Acquisition (SCADA) integrated for monitoring of the power distribution network are presented in [29], including a common graphical user interface and shared network model. CIM is utilized for modeling the distribution network and its standard compliance facilitates data exchange. This allows for the aggregation of power equipment information and spatial GIS data with real time operational status information, enhancing decision support and abnormal alert response. Application development costs are reduced with multi-platform support superior to commercial GIS, avoiding duplication of data while enhancing data validity and reducing human error.

The integration of smart substations in smart grid architecture supporting intelligence aggregation in utility operation and management is addressed in [8]. Complying with IEC CIM standards supports substation analytics and system integration with enhanced value adding information exchange. As a result, smart connectivity is provided for Intelligent Electronic Devices (IEDs), promoting interoperability at all utility system levels and enabling operating and functional information exchange suited for individual IEDs, utility and decision support systems.

In [33], an electric utility company utilizes big data via analytics and the proposed software framework is based on CIM IEC standards, in order to convert utility big data into operational decision support, promote efficiency and save costs. Figure A.4 illustrates how utility big data applications can interact with each other using a CIM-based integration architecture, through a common enterprise service bus.

Given the above-mentioned examples of how CIM is applied both in industry and in research, it can be concluded that its implementation is beneficial for an effective visualization of the electrical smart metering data due to the following reasons:

- Shared common services through a message bus interface: mapping of the CIM class structure to an application's external interface;

- Facilitation of a real time environment for dynamic data messaging;

- Enhanced data analytics;

- Identifying and resolving issues in order to ensure the quality of data;

- Platform independence.

CIM modelling is feasible via multiple formats [23]:

- UML-CIM: standard defined in UML using classes, attributes and relationships;
- XML-CIM: class structure mapping and data encapsulation format. Encoding of plain text enables human-machine interaction;
- RDF (Resource Document Framework): an XML schema that defines relationships between XML nodes (outside parent/child class relationships). Nodes are assigned unique RFD IDs and resource attributes. In addition, an *RFD schema* is needed to provide the vocabulary for describing an object oriented type of system for RDF. The combination of RDF and RDF schema supports a class hierarchy structure XML schema through inter-class properties;
- CIM, XML and RDF: their combination is used to model the entire CIM power system. Provides a readable format for both humans and machines, due to the platform independent plain text format;
- XML Messaging: data exchange is provided through the XML data structure (CIM messages), associated with an XML schema that defines classes and attributes interpretation.

Therefore, competitive, privacy or security concerns prohibiting open exchange of complete model data can be alleviated by layered data exchange via CIM with restrictions ensuring only the required data is shared.

IV Big Data and Data Visualization

In the energy sector, the progressive penetration of Distributed Generation (multiple sources of small scale power generation) brings deep changes in the design of the grid [24]. At the same time, the penetration of a large number of power electronic devices (PVs, heat pumps, smart meters) has brought major changes in the volume of collected data.

Big data analytics can bring new opportunities in the management of a smart grid, in terms of data storage, analysis and mining, as mentioned in [32]. The work also states the importance of GIS as a traditional and complex source of big data, characterized by spatial attributes.

In various parts of the existing literature big data is often referred to as data of very high volume, variety and massive continuous flow. Variety, volume and velocity, also known as the "3Vs" are some of the most common big data characteristics. Thus, the notion of real-time is closely related to the speed required for processing and analyzing the data [32] [33] [21].

But these parameters cannot provide meaningful decision support to ensure a smart, resilient and reliable power grid, unless valuable knowledge is extracted from big data. In [32] the core of smart grid big data management is presented: data mining techniques and knowledge representation/visualization.

Another study [33] proposes additional solutions to enhance utility big data, apart from the "3Vs". Utilization and analytics are recommended for implementing a user-centered application framework, while a presentation of the collected big data is useful for the visualization framework.

The design of a data platform supporting enterprise level Big Data Integration (BDI) is addressed in [21]. The methodology proposes data integration and big data analytics schemes into a common data repository platform featuring scalability, real time data and security. The objective is to overcome conventional solution challenges and to create user friendly and powerful data query visualization and analytics tools.

Some examples of big data use cases are presented in [26]. The big data reference architectures of social networking services, such as Facebook, Twitter, LinkedIn and Netflix, are shown. The implementation of such use cases requires a great variety of technologies regarding data integration, storage, analysis and visualization, with the purpose of a better understanding of consumers' needs.

The necessity of being able to display in near real-time the acquired information from different sources has led many researchers to develop GIS-based systems. [22] and [4] address the issue of real time data integration by using open source desktop GIS software, such as QGIS integrated with Grass. Other studies approach the combination of different types of GIS tools to display information: ArcGIS and QGIS [5], QGIS and Pmapper [22], QGIS, GRASS and MapServer [4]. An overview of GIS software tools utilized for visualization purposes is presented in [2] and [17], the latter concluding that QGIS is a better choice for data visualization and spatial analysis.

Given the aforementioned big data related work and the use cases presented in Section II, it can be concluded that the "3Vs" may be a relevant big data definition. However, the knowledge acquired from big data has a heavy impact in a decision making process. Therefore, a more accurate definition should include techniques on how and what big data can be actually used for, in the direction of developing a reliable, secure and effective power grid: data mining, visualization and data analytics.

Table A.1: Survey of GIS Tools

Desktop GIS Software	**License**	**Implementation Language**	**Platforms**	**Visualization features (plug-ins/extensions/add ins)**	**User Interface and Applications**
ArcGIS	Proprietary	Python, C++, R	Windows OS	GeoEvent Processor, Tracking Analyst (extension)	ArcMap, ArcCatalog
QGIS	Free	Python, C++	Linux, Unix, Mac OSX, Windows OS, Android	Open Layer (plugin)	QGIS Browser
MapInfo	Proprietary	MapBasic	Windows OS	Animation Layer (add in)	Color stretch tool
GRASS	Free	Python, C++	Linux, Unix, Mac OSX, Windows OS	Temporal Framework (extension)	Vector Digitizer, 2.5D/3D visualization with wxNviz
gvSIG	Free	Jython (Python implementation in Java)	Linux, Mac OSX, Windows OS	3D Animation Manager (extension)	3D Interface
Maptitude	Proprietary	GISDK (Geographic Information System Developer's Kit)	Linux, Unix, Mac OSX, Windows OS	Drive-Time Rings, Drive-Time Territories (geographic analysis tools)	MapWizard (thematic maps)

V Visualization Survey Analysis

Based on the studies in Section IV, some of the most popular desktop GIS tools are summarized in Table A.1, according to a few key characteristics that explain the motivation for choosing the appropriate tool for the purpose of this research: license type, supported platforms and plugins/extensions/add ins that could make possible a near-real time implementation.

Various tasks carried out by different kinds of GIS software are presented in Table 1.1 in [34]. Due to privacy concerns and the possibility for local manipulation of data, desktop GIS type has been chosen for the development of this project.

ArcGIS proprietary software is popular for its analytical functions, scripting tools and the possibility for user developed functions in multiple programming languages. The Tracking Analyst extension makes it possible to reveal and analyse data patterns [25], while GeoEvent Processor is able to process time critical events [15] [25].

In MapInfo, the first desktop GIS product, additional tools can be implemented through its dedicated MapBasic programming language, such as the animation layer add in, which is used for tracking frequently updated data, as in the case of real-time applications [28] [11].

Many of the common functionalities of a desktop GIS can also be found within the Maptitude commercial software. It does not provide any real-time related analytic capabilities, but these can be customized using the GISDK application development platform [9].

GRASS and gvSIG are open source GIS software that come in handy for storing and managing spatio-temporal data and solving planning issues [2]. 3D visualization of data and animations are achieved through their user interfaces and the customized extensions.

Open source QGIS software runs on various operating systems and supports data formats from both ArcGIS and MapInfo [34]. Its browser interface makes it possible to access, organize, and visualize data within the supported spatial layers [6]. Similar to MapInfo and Maptitude, its functionalities may be extended by creating additional plugins using Python or C++. Therefore, it is possible to integrate features for real-time display of the data, as it has been done in [19].

VI Conclusion and Future Work

This paper addressed the motivation and challenges for building an accessible and effective system to display real time and historical visualizations based on data acquired from smart meters. Due to the advantages of CIM, it will be a part of the future work to model the components of an electrical grid, in order to manage smart metering data and to represent GIS data. It is expected that the implementation of CIM will result in enhanced possibilities for data analysis techniques, which constitutes one of the main steps that have to be defined towards building the visualization platform. The choice of a suitable GIS tool was done according to the requirements in the defined use cases, especially the challenge of integrating real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported by multiple platforms, and its big variety of plugins, that can be implemented using commonly known programming languages.

The next step is to establish the requirements and specifications for data storage using database architectures. It includes the choice of a suitable implementation language, database structure and the description of the electrical network structure, aiming to create a platform scalable with the integration of CIM and GIS.

Acknowledgment

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 related to paper A

[1] *Use of real-time/historical database in Smart Grid*, 2011.

[2] *Software Tools Required to Develop GIS Applications: An Overview*, 2012.

[3] *Voltage stability improvement through centralized reactive power management on the Smart Grid*, 2012.

[4] *UAV/UGV cooperation for surveying operations in humanitarian demining*, 2013.

[5] *GIS as a tool for enhancing the optimization of demand side management in residential microgrid*, 2015.

[6] Packt Books. Adding real-time weather data from openweathermap: Qgis python programming cookbook, May `https://www.gislounge.com/`, 2015.

[7] Stuart Borlase. *Smart Grids*. 2012.

[8] Jose L. P. Brittes, Paula S. D. Kayano Luiz C. Magrini, Osvaldo Rein Junior Ferdinando Crispino, Wagner S. Hokama Jose A. Jardini, and Luis G. Fernandez Silva. Substation smartizing: An iec based approach for utility smart analytics development. *International Journal of Electronics and Electrical Engineering*, 4:284–289, 2016.

[9] The Caliper website, `http://www.caliper.com/`.

[10] The European Commission. Electricity network codes and guidelines, november `https://ec.europa.eu/energy/en/topics/wholesale-market/electricity-network-codes/`, 2016.

[11] Larry Daniel, Paula Loree, and Angela Whitener. *Inside MapInfo Professional*. OnWord Press, New York; NY, 3. edition, 2002.

[12] Arup Dasgupta. The continuum: Big data, cloud and internet of things, august `https://www.geospatialworld.net/article/the-continuum-big-data-cloud-and-internet-of-things/`, 2016.

[13] The Energinet.dk website, `https://www.energinet.dk/`.

[14] The European Network of Transmission System Operators for Electricity website, `https://www.entsoe.eu/`.

[15] ESRI. Arcgis enables real-time gis, `http://www.esri.com/esri-news/arcnews/spring13articles/arcgis-enables-real-time-gis`.

[16] E. Jun Zhu, Jun Zhu, Eric Zhuang, Chavdar Ivanov, and Ziwen Yao. A data-driven approach to interactive visualization of power systems. 26:2539–2546, 2011.

[17] R. R. N. Kanapaka and R. K. Neelisetti. A survey of tools for visualizing geo spatial data. In *2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 22–27, Dec 2015.

[18] Mohammed S. Kemal and Rasmus L. Olsen. Adaptive data collection mechanisms for smart monitoring of distribution grids. *CoRR*, abs/1608.06510, 2016.

[19] Sung Ah Kim, Dongyoun Shin, Yoon Choe, Thomas Seibert, and Steffen P. Walz. Integrated energy monitoring and visualization system for smart green city development. *Elsevier*, 22:51–59, 2012.

[20] Fangxing Li, Wei Qiao, Hongbin Sun, Hui Wan, Jianhui Wang, Yan Xia, Zhao Xu, and Pei Zhang. Smart transmission grid: Vision and framework. 1:168–177, 2010.

[21] Hesen Liu, Jiahui Guo, Wenpeng Yu, Lin Zhu, Yilu Liu, Tao Xia, Rui Sun, and R. M. Gardner. The design and implementation of the enterprise level data platform and big data driven applications and analytics. In *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, pages 1–5, May 2016.

[22] M. Mangiameli and G. Mussumeci. Real time integration of field data into a gis platform for the management of hydrological emergencies. XL-5/W3:153–158, 2013.

[23] Dr Alan W. McMorran. An introduction to iec 61970-301 and 61968-11: The common information model. Technical report, Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde, january 2007.

[24] Rosario Miceli, Salvatore Favuzza, and Fabio Genduso. A perspective on the future of distribution: Smart grids, state of the art, benefits and research plans. 5:36–42, 2013.

[25] ESRI Developer Network. What is arcgis desktop?, http://edndoc.esri.com/arcobjects/.

[26] Pekka Pääkkönen and Daniel Pakkala. Reference architecture and classification of technologies, products and services for big data systems. *Elsevier*, 2:166–186, 2015.

[27] Maher Chebbo Pieter Vingerhoets and Nikos Hatziargyriou. The digital energy system 4.0. Technical report, Smart Grids European Technology Platform, 2016.

[28] Pitney Bowes Software Inc. *MapInfo Pro Version 12.5.1 User Guide*, 2014.

[29] L. Qi, C. Wang, W. Zhou, and Z. Yang. Design of distribution scada system based on open source gis. In *Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2011 4th International Conference on*, pages 523–526, July 2011.

[30] The Statnett website, http://www.statnett.no/en/.

[31] Bruce Wollenberg, Jay Britton, Ed Dobrowolski, Robin Podmore, Jim Resek, John Scheidt, Jerry Russell, Terry Saxton, and Chavdar Ivanov. A brief history: The common information model, february http://sites.ieee.org/pes-enews/2015/12/10/a-brief-history-the-common-information-model/, 2016.

[32] Kaile Zhou, Chao Fu, and Shanlin Yang. Big data driven smart energy management: From big data to big insights. 56:215–225, 2016.

[33] J. Zhu, E. Zhuang, J. Fu, J. Baranowski, A. Ford, and J. Shen. A framework-based approach to utility big data analytics. *IEEE Transactions on Power Systems*, 31(3):2455–2462, May 2016.

[34] Xuan Zhu. *GIS for Environmental Applications: A practical approach*. Routledge, New York; NY, 2016.

Paper B

Data Analytics for Low Voltage Electrical Grids

Maria Stefan, Jose G. Lopez, Morten H. Andreasen,
Ruben Sanchez and Rasmus L. Olsen

The paper has been published in the:
Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security - Volume 1: IoTBDS, pp.221-228, 2018.

Abstract

At the consumer level in the electrical grid, the increase in distributed power generation from renewable energy resources creates operational challenges for the DSOs. Nowadays, grid data is only used for billing purposes. Intelligent management tools can facilitate enhanced control of the power system, where the first step is the ability to monitor the grid state in near-real-time. Therefore, the concepts of smart grids and Internet of Things can enable future enhancements via the application of smart analytics. This paper introduces a use case for low voltage grid observability. The proposal involves a state estimation algorithm (DSSE) that aims to eliminate errors in the received meter data and provide an estimate of the actual grid state by replacing missing or insufficient data for the DSSE by pseudo-measurements acquired from historical data. A state of the art of historical and near-real-time analytics techniques is further presented. Based on the proposed study model and the survey, the team near-real-time is defined. The proposal concludes with an evaluation of the different analytical methods and a subsequent set of recommendations best suited for low voltage grid observability.

I Introduction

At the beginning of the 21st century, a massive improvement of Information and Communications Technology (ICT) gave an opportunity for solving some existing limitations of the electrical grid, while also reducing the operational costs [23]. This sparked people involved in the development of the future energy market to think of new concepts. Of these ideas, *smart meters* and *smart grid* were the most popular, by adding ICT intelligence to the system, wherever useful.

These ideas led many countries to support various research programs in the smart grid domain. Denmark, already having a long tradition in the green electricity market, published a set of recommendations for implementing these concepts in the report called *Smart Grid in Denmark*. One Danish financed research program is ForskEL [12], meant to support the development and integration of environmentally friendly power generation technologies and grid connection.

One goal of ForskEL is to help the Distributed System Operators (DSOs) in making sensible decisions regarding future power grid planning and fault diagnosis in near-real-time. This calls for the utilization of intelligent methods for grid data visualization, as presented in [27].

The new challenge for the Danish DSOs arises as more distributed power generation is introduced at the low voltage grid level. This affects their ability to monitor the state of the power grid without encountering operational constraints. One of the DSOs' primary tools are to obtain full observability of the low voltage grid, by making use of scalable data analytics, as intended with the Danish RemoteGRID project [21]. Hence, high-performance data processing and analytical methods are fundamental for efficiently managing distribution grid data.

Two relevant data types are considered in relation to the power grid:

- Geographic data: electrical network structure (cables, transformers, substations, meters) and their geographical coordinates;
- Measurement data: three-phased generic grid measurements from each load or connection point containing multiple loads (voltage, current, consumption).

This paper introduces a study of analytical methods suitable for obtaining low voltage grid observability. The paper is organized as follows: Section II presents the flow of data in a smart grid application. The proposed study is presented in Section III and it underlines the advantages of pseudo-measurements and state estimator for the smart grid scenario. In Section IV, both generic and state of the art analytic methods will be presented. Given the chosen case study and background, the most suitable analytics will be emphasized in Section V, along with the definitions of bulk and stream data types. Section 5 will summarize the aforementioned study requirements with future action plans for testing the above concepts.

II Study background

The underlying application structure is defined based on the requirements imposed by the analytical methods suitable for the state estimation algorithm introduced in Section III. In this study, the application structure is proposed as a client-server application, based on the IEC 61868-100 standard [8]. The data flow is depicted in Figure B.1. The IEC standard is meant to provide guidelines regarding message exchange and interface specifications for utility enterprise distribution systems. Consequently, the key terms are clarified as follows:

- Advanced Metering Infrastructure (AMI) [32]: main data source in a smart grid, characterized by a large number of nodes (meters) located at customer premises;
- Meter Data Management (MDM) [17]: software entity that involves the storage and management of the AMI data. This includes the Database Management System (DBMS);
- Enterprise Service Bus (ESB) [24]: software-based integration layer specifying a standardized communication interface facilitating services (rout-

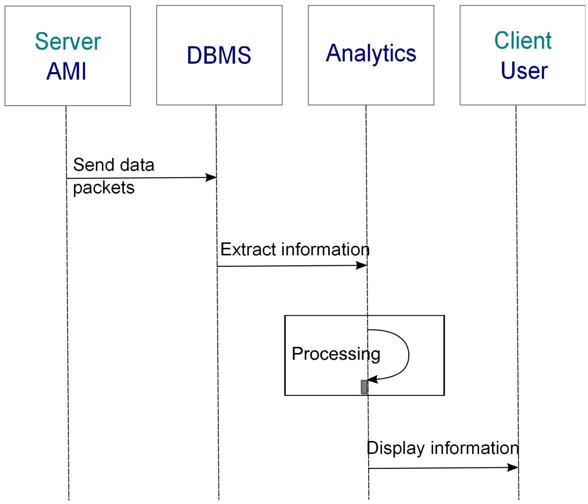


Fig. B.1: Data flow and exchange of automatic events according to IEC 61968-100.

ing, mediation, recording of data etc.) via standard event-driven messaging. The ESB middelware works as an adapter between different data formats and protocols in a Service Oriented Architecture (SOA).

Data is generated at the AMI (*server entity*) as an encoded packet, which is then decoded at the MDM level and sent to a database management system (DBMS) for storage via XML messaging [22]. In this back-end architecture [31], the DBMS is defined as an integration feature, which provides the *ESB* middleware with raw data to be sent to the analytics module for processing. A processing unit in the analytics module extracts the desired information to be displayed for the user (*client entity*). In the smart grid context, the user is usually located in the DSO control center. The event progression of data is storing - information extraction - information display. These events take place in a cyclic manner and thus, they are referred to as *automatic events*.

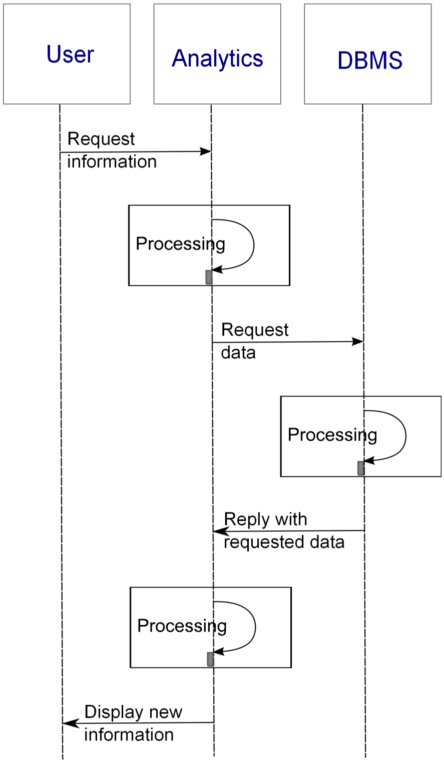


Fig. B.2: Data exchange of interactive events based on client request.

The data exchange sequences are not solely one-way. In case the client (user) detects unusual patterns or missing information in a certain geographical area, additional data from that specific area (or specific meters) can be requested for enhanced monitoring purposes. If so, the data flow is based on so-called *interactive events*, as shown in Figure B.2. The client's request for more detailed information is transmitted to the DBMS via the ESB, to search if there is a match for the requested data in the database. If a match is found, a reply is sent to the client for display and visualization. If not, the data request may be forwarded to the AMI, which will configure the meters to send the required data. Timing is crucial in the DSOs decision making process and is notably affected by delays in the transmissions from data collection to data display. Requesting certain information all the way from the AMI will result in additional delays due to the increased number of messaging sequences between entities.

As a part of the analytics module, the next section will introduce the *Distributed System State Estimator (DSSE)*.

III Study outline: Low voltage grid observability

Low voltage grids are undergoing a transformation from a passive to a more active role in the electrical network. Traditionally, conventional large gas or coal power plants, among others, are the source of electrical power generation [30]. After being transmitted at a high voltage level, the energy is distributed to supply the loads in the system. Lately, the penetration of distributed generation, especially from Renewable Energy Resources (RES), at the low voltage level has increased. It creates operational challenges for the DSOs since the low voltage grid was not designed to operate under such conditions. For example, generation peaks from RES do not necessarily match peaks of consumption, introducing power flows from the low to the high voltage level.

In order to address operational concerns, the DSOs require advanced management tools. Grid monitoring is the first step towards a more reliable operational approach [3]. In fact, nowadays, the low voltage grid electrical parameters are not monitored in the DSOs control centers. Monitoring the system allows DSOs to determine whether or not the system is operating under normal conditions. A system is considered to operate under normal conditions if all the loads can be supplied without violating any operational constrains [3].

A. Low voltage grid state estimation - LV DSSE

Grid observability depends on where the measurement points are placed along the electrical grid. In the case of low voltage grids, these measurements are provided by the smart meters. However, the information extracted from the meter's data contains errors due to various factors, such as communication issues or measurement deviations in the devices. Thus, as a first step in control centers, efficient data analytics are required to properly determine the state of the electrical grid. The state is defined as "known" if the voltages and phase angles with respect to a certain voltage and angle reference are known at every node (point where two or more circuit elements meet) [3].The process in charge of eliminating errors and providing the best estimate of the system state in distribution systems is the so-called *distribution system state estimation (DSSE)* [5].

Figure B.3 shows the block diagram of the observability analysis performed based on the raw measured data. This analysis determines if the system state can be estimated based on the set of acquired near-real-time readings. For example, few or non-existing measurements are sometimes provided from a specific geographical area of the system. This implies that the available data is insufficient to successfully estimate the state of the system. In that case, other data analytics methods are needed, where the unavail-

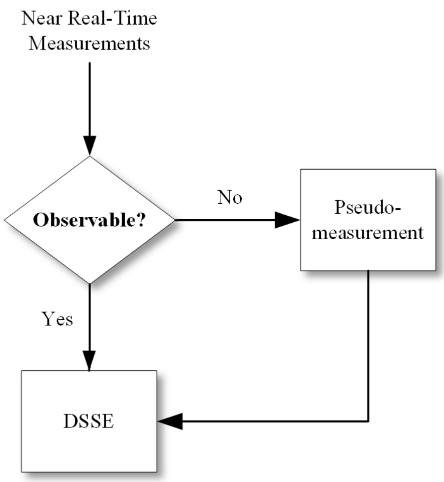


Fig. B.3: Evaluating observability based on the field near-real-time measurements.

able near-real-time measurements are substituted by the so-called pseudo-measurements obtained from historical data [18].

B. Pseudo-measurements

Traditionally, pseudo-measurements have been obtained from standardized daily load and generation profiles (DLP-DGP). Those are created for different customer classes based on socio-demographic factors [19]. However, other approaches seeking more precise accuracy have been developed in the literature. Artificial Neuronal Networks (ANN) are used in [11]. Besides, different clustering techniques were utilized as it is the case of k-means [6], principle component analysis [1], spectral clustering method [4] or finite mixture model [28], among others.

New solutions are to be studied in order to provide robust pseudo-measurements for low voltage grid applications based on the utilization of AMI data. Unpredictable behavior from RES is a challenge where efficiency in terms of the amount of stored data needs to be considered given the large number of nodes at the low voltage level.

IV Analytic Methods

AMI data is by definition part of the Internet of Things (IoT) umbrella, in the sense that smart meters act as sensors in the electrical grid infrastructure. IoT data analytics is characterized by autonomous or semi-autonomous examination of data, employing sophisticated techniques and tools, typically beyond those of traditional Business Intelligence (BI). These techniques help to reduce complex data sets into actionable insights, enhance and empower BI decision support systems. By this token, some traditional analytics and algorithms include data mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis [13].

Analytics are classified by two main categories: historical and near-real-time analytics.

- Historical analytics: based on the past data values. Data-at-rest corresponds to batch data processing;
- Near-real-time analytics: based on the present. Data-in-motion equals stream data processing.

A. Historical Analytics

Four traditional types of historical analysis are presented in the following subsections. They are a trade-off between the provided information value and the implementation difficulty. This is illustrated in [7].

A..1 Descriptive

This type of data analysis is used to provide insight into past events, by identifying overall themes and patterns. Descriptive analytics is commonly classified as BI and is the de facto standard analytics methodology. Typical outputs include dashboards, reports and status emails stating historical observations by summarizing raw data for human interpretation. These are mainly obtained through methods such as data mining and data aggregation.

An example of descriptive analytics can be found in [20], where daily profile of consumption trends are obtained by means of data aggregation. This helps to understand daily habits of consumers and, at the same time, to insure the privacy of end users through data anonymization [10].

A..2 Diagnostic

Diagnostic analysis helps answering questions like "Why was a certain event triggered", by providing a deep understanding of a limited problem space via in-depth data analysis, discovering the root causes and characteristics of an event. Advancing from aggregate and summary information to detailed data, based on specific focus attribute(s), is done via selection and querying of data sets. Data granularity defines the limit for the analytic level of detail. The resulting output is typically an analytic dashboard.

Correlation methods are part of obtaining a diagnosis analysis. The review in [25] proposes a method for characterizing power system loads by the correlation between load demand and weather variables.

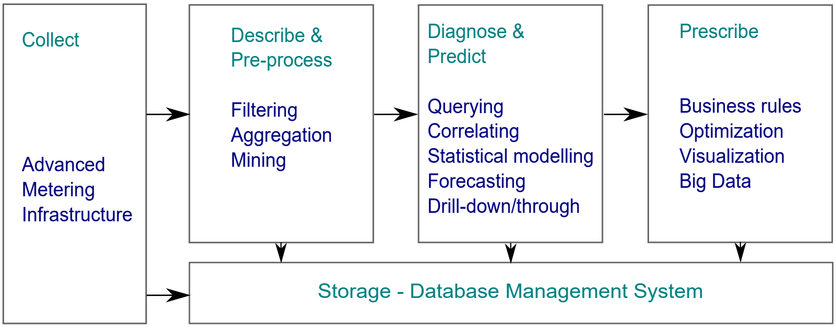


Fig. B.4: Proposal of streaming analytics architecture for low-voltage electrical grids [33].

A..3 Predictive

Predictive analytics is about foreseeing the future based on historical data patterns. Future predictions and scenarios come from data mining, machine learning and statistical modeling of raw data. Thus, actionable insights are obtained via plausible estimates of future outcomes. Typical deliverables are in the form of predictive forecasts based on probabilistic and correlation analysis.

Load forecasting is a common use case of predictive analysis in smart grids [10]. The study made in [2] approaches a forecast method which is based on a combination of ANNs and time series data models. Load forecasting can be achieved using not only correlations, but also through machine learning solutions, such as the MapReduce processing model [26] [14]. MapReduce allows for massive scalability across a cluster of computers, for large data sets (in the range of Terabytes), which is a suitable solution in case of AMI infrastructures.

A..4 Prescriptive

The primary focus of prescriptive analytics is to provide real-world recommendations. Datasets are evaluated via analytical models and the preferred cause of action for each specific event is selected. Then the result, in the form of explicit actionable information, is presented for human interaction, typically making the final decision on acceptance or rejection. Hence, prescriptive analytics takes a step further than predictive analytics by reducing complex data and algorithms to non-technical descriptors for immediately recognizable advice on predicted future outcomes. The analysis aids the decision-making process, having the potential to both maximize positive outcomes as well as prevent undesirable events [16].

Simultaneous utilization of multi-source datasets includes historical and real-time data, transactional and big data analytics, that affect marketing strategies [9]. For example, one significant tool to help utility companies navigate towards a smart grid platform is the Vitria IoT Analytics Platform, reported in [33]. This white paper states that a combination of prescriptive analytics and smart decisions provide the highest throughput in the analytics value chain.

B. Near-Real-Time Analytics

The resilience of the power grid is part of the future requirements for evolving towards intelligent grids. The main motivation for near-real-time analytics lies in the lack of limited grid functionality to timely detect and prevent failures. This extends to the discovery of natural disasters or criminal actions that might have caused the failures. Therefore, these can be prevented by making use of real-time intelligence [33].

B..1 Streaming Analytics

Near-real-time decision support can be provided via data-in-motion pre-database processing, inspection, correlation and analysis. It enables instantaneous management, monitoring, and continuous statistical analysis of data. Introducing real-time KPI overview, immediate access to metrics, and reporting, improves reaction time and accelerates decision-making.

Streaming analytics provide value from the data in a similar manner as traditional historical analytics. The value of streaming data decreases non-linearly over time, meaning that events should be reacted upon quickly, in near-real-time. The progression from historical methods comes as analytics are no longer performed ”at-rest”. Instead, data is processed before it is stored and therefore the decision-making process becomes timely and more efficient [15] [29]. A summary of the modules involved in the data streaming based on the surveyed analytics types is shown in Figure B.4. This figure shows that the same principles as in historical analytics can be applied to streaming data.

V Main findings and discussion

The study presented above emphasized the importance of introducing analytical methods to monitor the status of low voltage electrical grids and to plan future grid reinforcements. Historical data is used to create pseudo-measurements, aiming to fill in missing or erroneous data received from a smart grid infrastructure.

Given the back-end client-server architecture presented in Section II, the automatic ingestion of data can be defined as a ”stream of data”:

Table B.1: Advantages and disadvantages of using historical and near-real-time analytics for providing data to the DSSE.

Analytics	Pros	Cons
Historical (context awareness)	• provide insight by uncovering data patterns and trends	• accuracy and realiability dependent on time
	• quickly accessible and detailed (available and verified data)	• most machine learning algorithms do not deal with temporal effects
	• clarity by presentation of reduced complex data sets - thorough presentation of large data sets	• reliance on batch processing and consequently limited by the resulting update intervals
Near-real-time (situation awareness)	• detects gross errors - accuracy	• highly dependent on the delays in the communication network
	• avoid latency from filtering disk data	• difficult to adapt to platform and hardware requirements
	• detect emerging correlations between multiple data sets	• risk of incorrect analysis via implementation dependency
	• immediate pre-database data availability	

Near-real-time measurements are characterized as a continuous, fast changing and voluminous data flow, commonly known as stream.

To support the above-mentioned definition, the notion of *near-real-time data* can be given in the context of the data flow architecture in Section II and the use case presented in Section III:
Assuming that the data packets sent from the low voltage grid arrive consecutively with a fixed period of time, then a near-real-time data stream can be defined as: *a data packet characterized by the arrival granularity and received in a timely manner at the user side. Timing is then relative to the types of events involved in the data flow: automatic or interactive.*

The analytical methods involved in the DSSE algorithm are based on both historical and near-real-time data. Due to their timely nature, the near-real-time measurements are more reliable and accurate than the historical ones. Therefore, the DSSE needs near-real-time data, that should be preprocessed in order for the estimator to "understand" it, equivalently to the streaming analytics procedures shown in Figure B.4. There are typically not enough near-real-time measurements available to successfully perform the DSSE. Therefore, there is not enough data to provide full grid observability. In order to fill in the gaps of missing information, pseudo-measurements can be created by requesting raw data that has been previously stored in a database. The requested information can therefore be extracted by means of filtering, mining or querying, making it comprehensible for the DSSE. In this case, the most suitable analytics are descriptive.

A summary of pros and cons of the aforementioned analytics for the DSSE is presented in Table B.1. The novelty of this study is based on the integration of traditional analytics into the energy-related field, which consists of the DSSE algorithm. As historical based analytics are useful to build periodic reports for strategic and long-term decisions, they are also limited by the temporal effects. Historical data may not give a true pattern of a data trend, if this has changed with time. While near-real-time analytical tools can address the temporal dependency, they are also platform sensitive.

VI Conclusion

This study addresses the challenges for choosing suitable data analytics methods in the domain of low voltage smart grids. DSSE is an analytical method for providing a reliable source of information related to the state of the grid, by filtering the raw data and detecting gross errors. Ideally, DSSE makes use of near-real-time data to provide a successful estimation. In many cases, this data is insufficient or non-available, so pseudo-measurements generated from historical data will fill in for the lack of information. Traditional historic analytics can build predictive outputs useful for the DSSE, but there is a higher error probability in the pseudo-measurements.

By this token, the data analytics module should be built on a platform that can accommodate for both historical and near-real-time analysis. The next step in this research is to test the functionality of a DSSE algorithm and analyze the capabilites of processing large amounts of historical batch data. At the same time, the test aims to characterize the performance and bottlenecks of parallel processing of both stream and batch data types, taking into account parameters such as memory usage, processing time and in-memory processing behavior.

Acknowledgment

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 related to paper B

[1] Joana M Abreu, Francisco Câmara Pereira, and Paulo Ferrão. Using pattern recognition to identify habitual behavior in residential electricity consumption. *Energy and buildings*, 49:479–487, 2012.

[2] M. A. Abu-El-Magd and R. D. Findlay. A new approach using artificial neural network and time series models for short term load forecasting. 3:1723,1724,1725,1726, 2003.

[3] Ali Abur and Antonio Gomez Exposito. *Power system state estimation: theory and implementation*. CRC press, 2004.

[4] Adrian Albert and Ram Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, 28(4):4019–4030, 2013.

[5] Arash Alimardani, Francis Therrien, Djordje Atanackovic, Juri Jatskevich, and Ebrahim Vaahedi. Distribution system state estimation based on nonsynchronized smart meters. *IEEE Transactions on Smart Grid*, 6(6):2919–2928, 2015.

[6] Ignacio Benítez, Alfredo Quijano, José-Luis Díez, and Ignacio Delgado. Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers. *International Journal of Electrical Power & Energy Systems*, 55:437–448, 2014.

[7] The four types of analytics, `http://www.ciandt.com/card/four-types-of-analytics-and-cognition`, 2014.

[8] International Electrotechnical Commission. Application Integration at Electric Utilities - System Interfaces for Distribution Management - Part 100: Implementation Profiles. Technical Report 61968-100, International Electrotechnical Commission, July 2013.

[9] Houda Daki, Asmaa El Hannani, Abdelhak Aqqal, Abdelfattah Haidine, and Aziz Dahbi. Big data management in smart grid: concepts, requirements and implementation. *Journal of Big Data*, 4(1):13, Apr 2017.

[10] Panagiotis D. Diamantoulakis, Vasileios M. Kapinas, and George K. Karagiannidis. Big data analytics for dynamic energy management in smart grids. 2:94–101, 2015.

[11] MB Do Coutto Filho, JCS Souza, RSG Matos, and M Th Schilling. Preserving data redundancy in state estimation through a predictive database. In *Electric Power Engineering, 1999. PowerTech Budapest 99. International Conference on*, page 271. IEEE, 1999.

[12] The Energinet.dk website, `https://www.energinet.dk/`.

[13] Gartner it glossary - advanced analytics, `https://www.gartner.com/it-glossary/advanced-analytics`, 2017.

[14] GJSissons. Adaptive mapreduce: Part 1, `https://developer.ibm.com/hadoop/2014/09/28/adaptive-mapreduce-part-i/`, 2014.

[15] Daniel D. Gutierrez. InsideBIGDATA - Guide to Streaming Analytics. Technical report, Impetus Stream Analytix, 2016.

[16] Descriptive, predictive, and prescriptive analytics explained, `https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/`, 2017.

[17] IEC/TC. Application Integration at Electric Utilities - System Interfaces for Distribution Management - Part 9: Interfaces for meter reading and control. Technical Report 61968-9, International Electrotechnical Commission, October 2013.

[18] Amin Khodabakhshian, Rahmatollah Hooshmand, and Yaser Raisee-Gahrooyi. A new pseudo load profile determination approach in low voltage distribution networks. *IEEE Transactions on Power Systems*, 2017.

[19] Vladan Krsman, Bojan Tesanovic, and Jovan Dojic. Pre-processing of pseudo measurements based on ami data for distribution system state estimation. 2016.

[20] Xiufeng Liu, Lukasz Golab, Wojciech Golab, Ihab F. Ilyas, and Shichao Jin. Smart meter data analytics. 42:1–39, 2016.

[21] Ruben Sánchez Martin-Loeches, Florin Iov, Mohammed Seifu Kemal, Maria Stefan, and Rasmus Løvenstein Olsen. Observability of low voltage grids: actual dsos challenges and research questions. In *Proceedings of the 2017 52nd International Universities' Power Engineering Conference (UPEC)*. IEEE Press, 8 2017.

[22] Dr Alan W. McMorran. An introduction to iec 61970-301 and 61968-11: The common information model. 2007.

[23] Rosario Miceli, Salvatore Favuzza, and Fabio Genduso. A perspective on the future of distribution: Smart grids, state of the art, benefits and research plans. 5:36–42, 2013.

[24] Scott A. Neumann and Terry D. Nielsen. Cim interoperability challenges. pages 1–5, 2010.

[25] Muhammad Qamar Raza and Abbas Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. 50:1352–1372, 2015.

[26] Nikzad Babaii Rizvandi, Javid Taheri, and Albert Y. Zomaya. A study on using uncertain time series matching algorithms in map-reduce applications. *CoRR*, abs/1112.5505, 2011.

[27] M. Stefan, J. G. Lopez, M. H. Andreasen, and R. L. Olsen. Visualization techniques for electrical grid smart metering data: A survey. In *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 165–171, April 2017.

[28] Bruce Stephen, Antti J Mutanen, Stuart Galloway, Graeme Burt, and Pertti Järventausta. Enhanced load profiling for residential network customers. *IEEE Transactions on Power Delivery*, 29(1):88–96, 2014.

[29] Weighing the pros and cons of real-time big data analytics, `https://www.techopedia.com/2/31245/technology-trends/big-data/weighing-the-pros-and-cons-of-real-time-big-data-analytics`, 2017.

[30] David Trebolle, Per Hallberg, Gunnar Lorenz, Pavla Mandatova, and Jorge Tello Guijarro. Active distribution system management. In *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on*, pages 1–4. IET, 2013.

[31] A beginner's guide to back-end development, `https://www.upwork.com/hiring/development/a-beginners-guide-to-back-end-development/`, 2017.

[32] Noelia Uribe-Pérez, Luis Hernandez, David de la Vega, and Itziar Angulo. State of the art and trends review of smart metering in electricity grids. 6:68, 2016.

[33] Vitria. Advanced Analytics for Energy Utilities: The Fast Path to a Smart Grid. Technical report, Vitria Technology, 2015.

Paper C

Exploring the Potential of Modern Advanced Metering Infrastructure in Low-Voltage Grid Monitoring Systems

Maria Stefan, Jose G. Lopez and Rasmus L. Olsen

The paper has been published in the:
IEEE International Conference on Big Data, 2019.

Abstract

Energy systems are evolving towards 100% green energy production. The share of green energy in electrical distribution systems is progressively increasing, implying also an increment on the number of renewable energy units in the low-voltage grid. Following this trend, thousands of consumers connected to the power grid in a decentralized manner become small producers, changing the traditional paradigm of energy distribution from top to bottom. Currently, the modern Advanced Metering Infrastructure (AMI) enables the possibility of collecting several types of status data from the end-users which Distribution System Operators' (DSOs) can use to their advantage to optimize management and planning operations. As a part of this optimization, having a spatial overview over the low-voltage grid can speed up the monitoring processes and allows to obtain a real-time insight on what is happening in the grid, compared to the traditionally used analysis methods. Many business structures for smart grid cyber physical systems are looking into how to integrate advanced data management models. Such models should provide the means for obtaining meaningful data visualization where only the relevant data is timely processed, filtered and visualized for the operators to efficiently react to grid anomalies in real-time.

The purpose of this paper is to investigate how to efficiently design a monitoring/visualization system for low-voltage electrical grids based on the DSOs' needs and feedback. The proposed system implementation stands on emulating an existing geographic scenario by a virtual AMI integration. The efficiency of the prototype is evaluated versus the traditional monitoring operations derived from user experience studies, such as a reduction in time to perform a specific anomaly detection operation. Furthermore, the advantages of spatial awareness are meant to further strengthen the motivation for integrating measurements into a Geographic Information System (GIS) environment.

I Introduction

The gradual replacement of traditional energy sources with green/renewable energy, especially in countries with a long tradition in the green electricity market, such as Denmark, has sparked people to think about new concepts. Many business models are nowadays exploring integration methods that can accommodate for the interaction between wide topics of interest, such as: intelligent power grids, big data, data analytics and visualization. At the enterprise level, it is challenging to scale up with these rapidly advancing technologies, without modifying too much the already working business structure. The research community can help in this sense with designing and testing modular systems that can be easily integrated within companies.

In Denmark, the electricity transmission network is built around an alternating current network (AC) of 132 kV, 150 kV, 220 kV and 400 kV. The Western Danish Power System is a regular transit area with large interconnections to Norway (HVDC), Sweden (HVDC) and Germany (AC) [2]. The system can already operate with 20% wind and 50% CHP [3]. However, the energy plan in Denmark is to introduce 100% renewable energy by 2050, as a part of the smart grid strategy [4]. As new renewable energy units are connected (local CHP plants, wind turbines), the grid becomes more active because the consumers can also act as small energy suppliers (prosumers) and the grid evolves towards a decentralized architecture [7]. This will influence the way the traditional power grid operates.

Currently, the incoming measurement data from the advanced metering infrastructure (AMI) is used only for billing purposes. However, the already deployed metering infrastructure allows collecting many more grid parameters. The availability of these parameters offers the possibility of implementing automated monitoring solutions for real-time anomaly detection in the grid, which have been out of reach in the past. Nowadays, the most common issues reported by the consumers to DSOs are: power outage at the consumer end and flickering of lights.

It is expected that with the high penetration of renewables there will be an increase in reported issues by the consumers. Measurement data can be utilized to characterize the consumption patterns and understand the behavior of the users. In this case, the need for a more efficient monitoring and operation of the grid becomes essential.

The Danish RemoteGRID project [5] [18] is a research initiative that aims to support the Danish DSOs in gaining a visual overview over their low-voltage electrical grid, using GIS-based systems. As a starting point, it is important to identify some of the DSOs' challenges in a typical work day, based on some concrete examples. In this article, a near-real-time monitoring system is proposed and tested in comparison with the examples from user experience studies. The aim is to show that a GIS-based monitoring solution helps the human operators in identifying measurement anomalies faster and more effectively than this is currently done, thereby improving the grid monitoring procedures.

II Related Work

Given the case of small producers described in Section 1 and the DSOs' need for adequate grid monitoring, there is a clear motivation for investigating data management and visualization techniques in power systems.

Monitoring systems. Electrical networks are getting increasingly more complex with the implementation of Information and Communication Technologies (ICT). These so-called Smart Grids collect data about power consumption, which is used to diagnose problems in the network and prevent outages. The resulting large amount of data makes it more difficult to monitor the electrical grid.

Some available monitoring systems of real-time operational data use GIS [6] for data modeling (ArcGIS) and visualization (Web GIS). Apart from getting an overview of the grid status, the role of monitoring may also imply fault location. The article in [10] proposes such a system that is able to localize faults and their cause in a medium-voltage power grid, by analyzing only the electric current. The alarms are presented for human interaction by means of a SCADA (Supervisory Control and Data Acquisition) system [9].

Common challenges in the existing monitoring systems developed for smart grids relate to the cost of the required hardware and software components, as well as scalability issues that arise due to finer data granularity and the amount of smart meters deployed.

Real-time data visualization and GIS. Traditionally, real-time data visualization and geographical information systems (GIS) in smart grids are decoupled [22]. Visualization is usually achieved through dashboards, while GIS is the tool to represent the electrical grid topology. The AMI generates different types of measured values: three-phased voltages, currents, active and reactive powers. The key benefit consists in assigning these values to their corresponding measuring points on a GIS map and thus, providing an integrated solution for data monitoring.

A cloud-based framework for visualizing time-series energy readings from an electrical grid infrastructure is proposed in [20]. The application was focused on creating status dashboards using Tableau software in connection with the Apache Hadoop platform. Using timestamped measurements for fault/event detection in the electrical power system for a frequency monitoring network (FNET) is described in [11].

The paper [8] highlights the importance of spatial awareness on a visualization platform by demonstrating the spatio-temporal features of QuantumGIS (QGIS). An animation that used Time Manager plug-in made it possible to display real-time values in QGIS.

Big data management in smart grids. The variety in measured data values comes from deploying renewable energy resources (RES) in the low-voltage grid, interfering with its usual operation. This translates into the need for understanding the extent and behavior of the small producers in order to keep track of them, by processing and analyzing the available measurement data. Stream computing [12] is one of the main approaches for data analysis in real-time, given the amount of data generated per day in a smart grid (approximately 1 terabyte according to [24]). This is done through some dedicated distributed stream computing platforms, such as Apache Storm

[24] [13], StreamBase [25], Apache Spark [23], etc.

Other research initiatives have involved cloud computing frameworks. Nephele has been used in [17] to simulate data incoming from 1000000 smart meters in order to provide pricing updates every 10 seconds. Pricing strategies are also proposed in [14], using multiple cloud compatibility for stream data computing, and in [17] through Infrastructure-as-a-Service (IaaS) clouds for parallel stream processing. These solutions point out the importance of scalability in a big data management system.

Different data storage possibilities are supported within distributed computing, which are meant to ease the real-time data analysis process. A smart meter data analysis system was developed in [15] by means of PostgreSQL database and MADLib in-database machine learning library. This work is followed by a similar article presented in [16], where the processing of data is done using in-memory tables.

III Model and Simulation

Based on the aforementioned literature review, this paper will focus on designing and testing the efficiency of a monitoring system based on human needs and feedback. Two sequence models were identified to represent the intents and steps of activities regarding the low-voltage grid. The sequences describe the troubleshooting process when a consumer calls with no power, or how to identify the geographical area affected by an error.

The designed system is currently costly and challenging to deploy in real life, where it is not certain that the clocks (and thereby measurements) are properly synchronized. This makes simulation the proper tool for assessment of the near-real-time functionality, where results can be compared in a synchronized setting. Evaluating and comparing the sequence models before and after the simulation will characterize the impact of GIS on the monitoring system.

A. Case description - household power outage

The most frequent reports received by the DSOs are customers calling in to report the loss of power in their households. This user story has been identified by means of interviews with one of the Danish DSOs.

In this scenario, the DSO's task is to figure out whether they are responsible for the outage, by checking the three-phased voltage values in the corresponding meter assigned to a specific address. The DSO is able to demand information regarding 45 different values recorded by the meter. For this scenario, the DSO is merely interested in the three-phased voltages. Selecting

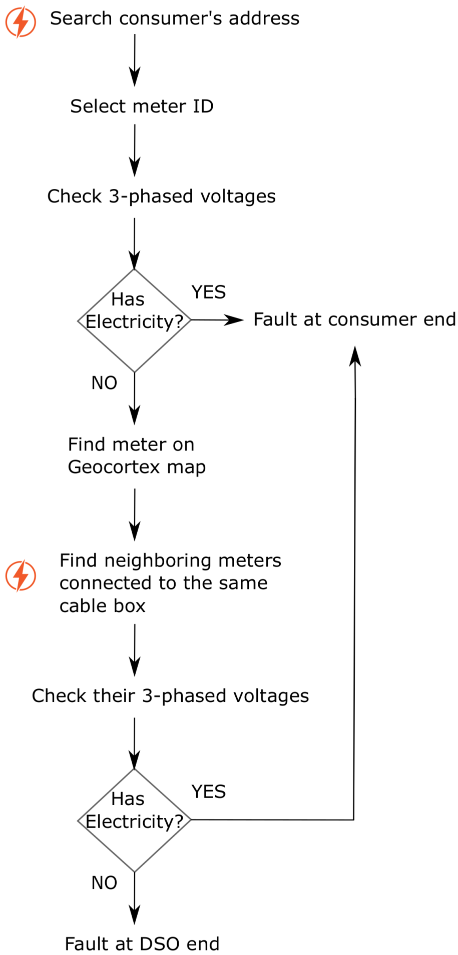


Fig. C.1: Sequence model for identifying if the outage at the consumer level is the DSO's responsability or the consumer's.

these values and sending a demand to the smart meter takes around 45 - 60 seconds due to a slow radio frequency connection between the meters. If the customer is found without power, the DSO tries to establish whether his neighbors have the same problem.

The meter is located on a Geocortex map, together with the neighboring meters connected to the same physical cable box. The same procedure of checking the three-phased voltage values is applied to the identified neighbors. If any of the neighboring addresses is also without electricity, the outage is the responsibility of the DSO.

The sequence model for evaluating the power outage in a household is

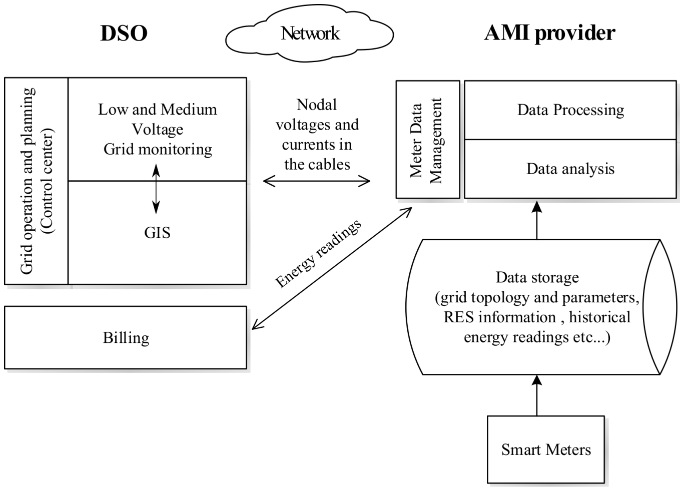


Fig. C.2: Data flow and information exchange between the meter provider and the DSO.

presented in Figure C.1. The lightnings represent alerts for the operations that are the most time consuming, due to the risk of choosing a wrong address or finding the relevant voltage measurements from the meters.

Throughout the interview, it was found that different software programs are used by the DSOs to assess different aspects of the low voltage grid, depending on specific problems. The two most commonly used programs are MidtVest Teknikerportal and Geocortex. Together these two were used to locate the errors when customers called, claiming they have no power. MidtVest Teknikerportal is a technical program used to explore whether there are any fluctuations in the three phases on the given address. Geocortex is then used to see whether any of the neighboring houses have any fluctuations as well and determine if there are any problems with the cable box.

The effects of added renewable energy such as personal windmills or solar panels are also calculated using SonWin, to determine whether an outage at a consumer is the DSO's responsibility or the consumer's.

The next step is to simulate a near-real-time system, which can monitor the received voltage measurements by means of a GIS map.

B. Data flow

Figure C.2 shows the flow of data in the power grid. The readings from the smart meters are transferred from the AMI provider to the DSO over a radio network (it can be mesh, power line communication, fiber). All types of radio networks have the common characteristics they were developed to support only the billing aspect, meaning that in most cases a low capacity network has

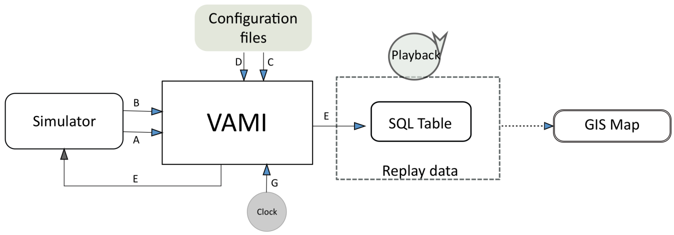


Fig. C.3: VAMI interface overview diagram.

been used. The AMI provider has information about the grid topology and parameters, customers (with or without RES), type of customers, historical energy readings and other value types. This data is filtered and analyzed according to the needs and requirements of the DSO: billing, planning or monitoring.

An efficient monitoring system usually requires additional measurements, such as three-phased voltages in the nodes and cable currents. The challenge lies in sending a large amount of varied information over the radio network, which overloads the network and slows down the monitoring process in near-real-time. At the same time, acquiring a large amount of mixed measurements overloads the map intended for visualizing this data.

It is important to analyze the work flow of the DSOs' routine operations in order to establish how much information should be displayed in the monitoring system and which of the steps in the work flow are most time consuming. It is hypothesized that spatial awareness leads to an effective monitoring by minimizing subsidiary procedures.

C. VAMI

The performance of a real life infrastructure of smart meters (as illustrated in Figure C.2) is emulated by a generalized software component called *Virtual Advanced Metering Infrastructure (VAMI)*. The main functionality of VAMI is to intake and delay messages/signals from an output block in the used Simulator (Simulink, in this case), as shown in Figure C.3.

The working principle of VAMI is based on its ability to handle a large amount of incoming data on many different ports. In order to satisfy this requirement, Java Socket Selectors from Java NIO are used [1] [21]. The interfaces between the different components are listed in Table C.1.

In order to evaluate the performance of the monitoring data system, the near-real-time functionality has been represented as a *playback* of already recorded historical data measurements provided by this virtual provider. Incoming data is outputted to a SQL table, where the playback functionality

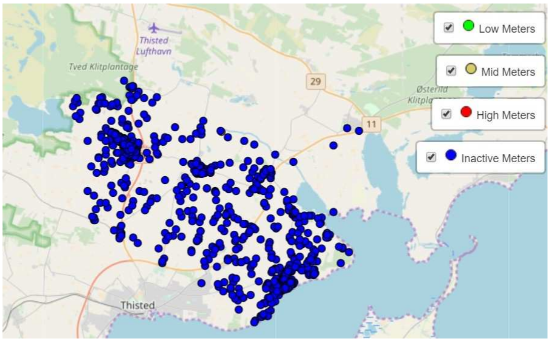
Table C.1: Main interfaces in the VAMI architecture

Interface	Type	Comment
A - Smart meter data input	UDP - any vector of data elements without separation characters, limited to one packet per vector element	Configurable port ranges
B - Simulator time synch signal	UDP - double value of time in simulator at sending time	Configurable port - optional to use this synchronization method via VAMI configuration file
C - CDF for random numbers	Time series separated via comma	Currently no implemented limit to the time series - limited by Java's internal capacity to store data in a vector
D - Configuration file	XML - proprietary	Set via first parameter when starting VAMI
E - Adapted output	Output signals adapted to a given sink - interfaces dependent on output sink. Default is UDP, raw: unmodified data received on Interface A	Configurable in VAMI configuration file
G - Real time synch	System clock	Alternatively to using simulation synchronization, VAMI can be time synchronized via the system's real/time clock if simulator or any other component is also time synchronized this way

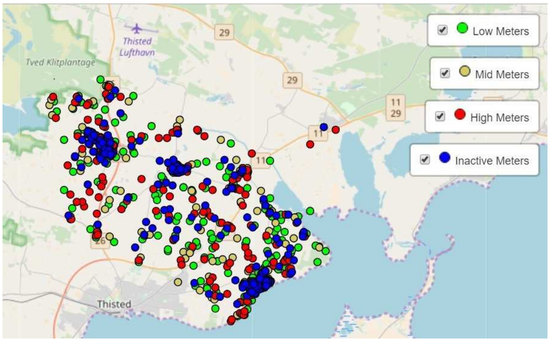
accesses the stored historical data to create a dynamic display on the WebGIS map. The data set is gradually overwritten with the runtime rate of VAMI and, as a result, it is assumed that the raw incoming data is near-real-time. Timing plays a vital role in a monitoring system and it is relative to the interfaces between different entities and the processes that take place in each entity module. This can be seen in a real life scenario, as shown in Figure C.2.

D. WebGIS

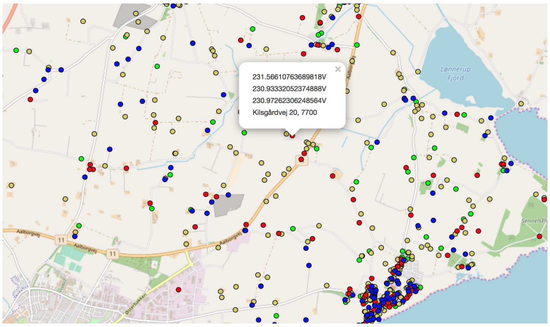
Smart grid data for approximately 1000 households/smart meters is simulated and displayed in a GIS environment. The purpose of the displaying information on a WebGIS map is to achieve a visual perspective over the low-voltage electrical grid. This means that a DSO is able to understand the consumer's behavior, as well as detect if a certain area of the grid presents



(a) Initial state inactive meters.



(b) Near-real-time GIS map update.



(c) Pop-up example containing three-phased voltage values and customer address.

Fig. C.4: Update and application of the WebGIS map in the monitoring system

anomalies. For example, supply voltage variations in households should not exceed ±10 % of the nominal voltage value, which represents the normal operating conditions in the electrical grid [19].

Initially, the data set contains only 0 voltage values, meaning that all the

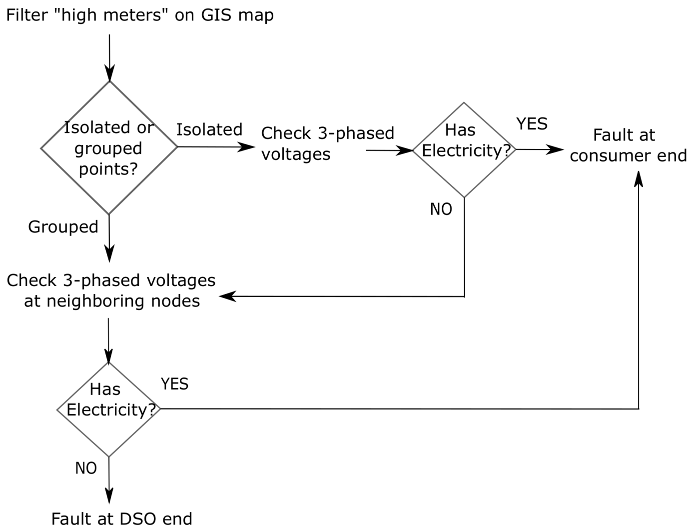


Fig. C.5: Sequence model for identifying if the outage at the consumer level is the DSO's responsability or the consumer's using the GIS map.

meters are in a state called *inactive* (Figure C.4a). As the data is gradually recorded, the voltage values will randomly be updated and displayed on the map. Three subsets are defined based on the minimum and the maximum voltage values registered: low, medium and high voltage. This classification gives the visual overview over the measured values by placing them in one of the three categories.

From a user experience point of view, having this viewpoint over the electrical grid can help to quickly establish the areas that require most attention. For example, the high voltage measurement points denoted "high meters" are represented in red in Figure C.4b. This layer may be of particular interest for the DSOs, while a pop-up is displayed if the user would want to check a particular point on the map. This pop-up in Figure C.4c shows the voltage values on the three phases and the location of the smart meter.

The advantage of integrating nodal measurements with GIS lies in the possibility for data management from a visual point of view. Filtering the layer of interest on the map decreases the amount of data displayed considerably and enhances the general overview when the map is zoomed out. Some of the sub-areas also constitute an interest since they are rich in RES. GIS gives the possibility of zooming in and selecting only the meters from that specific area in order to analyze the behavior of those customers.

IV Results and Discussion

The visualization system for the low-voltage grid monitoring has been achieved through a theoretical approach. The focus is on the human perception in relation to different ways of visualizing a map, such as noticeable differences regarding the number of data points or map refreshing time. This approach does not require testing on system operators, but focuses on a broader range of people targeting the general human perception.

According to the designed GIS-based monitoring system, a new sequence model was identified in Figure C.5. From the process efficiency point of view for the specific case described in Subsection A., it can be noticed that the number of operations has been reduced from 6 to 2 due to GIS integration.

Compared to the initial sequence model from Section III, Figure C.1, the new model does not contain alerts regarding time consuming operations. The monitoring performance is improved due to linking measurement data and the GIS location of the consumers, which makes it easier for the DSOs to spot anomalies, as well as their geographical location and address. Filtering only the points of interest on the map minimizes the amount of steps required to determine whether a power outage report is due to an anomaly in the grid. Thus, the time to act on a certain report is also shortened.

The voltage data is only pushed to the visualization system whenever it fluctuates more than 10% of the display value and therefore, in normal conditions, there are no great requirements in terms of memory or data transactions. Moreover, it does not make sense to display a huge area at once for a human user, and therefore, the data is only displayed when zooming in to substation level. The maximum number of meters in a substation is around 2000, needing an initial data pull for 2000 entities. The rest of the time a substation is being displayed, the requirements would be even lower than that initial case.

V Conclusion

From the data flow perspective, taking advantage of the unused new features of the current AMI in combination with intelligent data filtering and GIS visualization, allows the operators to understand the consumption patterns of the customers and to identify issues or anomalies that may arise from the large number of RES.

On that account, the monitoring system helps to speed up the decision making process under these special conditions. With the support of spatial awareness, the efficiency of operating the power grid improves, and thereby the overall human working experience and competences.

The next step is to improve the design of the presented GIS monitoring system by focusing on the system operators in their working environment. Taking this contextual approach would therefore include the DSO as a central part of the design process and would be more directly useful for electricity companies.

Acknowledgment

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. The authors would like to thank the two student groups at Product Design Psychology studies at Aalborg University who conducted the interviews with the DSOs in the spring of 2018.

References related to paper C

[1] The rox java nio tutorial, `http://rox-xmlrpc.sourceforge.net/niotut/`.

[2] Kriegers flak - combined grid solution, `https://en.energinet.dk/Infrastructure-Projects/Projektliste/KriegersFlakCGS`, 2017.

[3] Danish energy agency, February `https://ens.dk/service/statistik-data-noegletal-og-kort`, 2018.

[4] Smart grid strategy. the intelligent energy system of the future, May `https://ens.dk/sites/ens.dk/files/Globalcooperation/smart_grid_strategy_eng.pdf`, 2013.

[5] Projektbeskrivelse, `http://www.remotegrid.dk/projektbeskrivelse/`.

[6] Atefeh Dehghani Ashkezari, Nasser Hosseinzadeh, Ayoub Chebli, and Mahammed Albadi. Development of an enterprise geographic information system (gis) integrated with smart grid. *Sustainable Energy, Grids and Networks*, 14:25 – 34, 2018.

[7] Danish Energy Association. Smart grid in denmark 2.0, `https://www.usef.energy/app/uploads/2016/12/Smart-Grid-in-Denmark-2.0-2.pdf`, 2016.

[8] C. Xuen D. Nga O. See, D. Quang and L. Chee. Visualization techniques in smart grid. *Smart Grid and Renewable Energy*, 3:175–185, 2012.

[9] Houda Daki, Asmaa El Hannani, Abdelhak Aqqal, Abdelfattah Haidine, and Aziz Dahbi. Big data management in smart grid: concepts, requirements and implementation. *Journal of Big Data*, 4(1):13, Apr 2017.

[10] Eduardo F. Ferreira and J. Dionísio Barros. Faults monitoring system in the electric power grid of medium voltage. *Procedia Computer Science*, 130:696 – 703, 2018. The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.

[11] H.A. Gabbar, A. Zidan, and M. Xiaoli. Chapter 16 - data centers for smart energy grids. In Hossam A. Gabbar, editor, *Smart Energy Grid Engineering*, pages 433 – 452. Academic Press, 2017.

[12] Joseph, Shibily Joseph, E. A. Jasmin, and Soumya Chandran. Stream computing: Opportunities and challenges in smart grid. 21:49–53, 2015.

[13] S. Joseph and E. A. Jasmin. Stream computing framework for outage detection in smart grid. In *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, pages 1–5, Dec 2015.

[14] H. Li, M. Dong, K. Ota, and M. Guo. Pricing and repurchasing for big data processing in multi-clouds. *IEEE Transactions on Emerging Topics in Computing*, 4(2):266–277, April 2016.

[15] X. Liu, L. Golab, and I. F. Ilyas. Smas: A smart meter data analytics system. In *2015 IEEE 31st International Conference on Data Engineering*, pages 1476–1479, April 2015.

[16] Xiufeng Liu and Per Sieverts Nielsen. *Streamlining Smart Meter Data Analytics*. International Centre for Sustainable Development of Energy, Water and Environment Systems, 2015.

[17] B. Lohrmann and O. Kao. Processing smart meter data streams in the cloud. In *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, pages 1–8, Dec 2011.

[18] Ruben Sánchez Martin-Loeches, Florin Iov, Mohammed Seifu Kemal, Maria Stefan, and Rasmus Løvenstein Olsen. Observability of low voltage grids: actual dsos challenges and research questions. In *Proceedings of the 2017 52nd International Universities' Power Engineering Conference (UPEC)*. IEEE Press, 8 2017.

[19] D. Mezera. Voltage quality in the low voltage distribution grids with the high penetration of distributed energy sources. In *2015 16th International Scientific Conference on Electric Power Engineering (EPE)*, pages 292–295, May 2015.

[20] A. A. Munshi and Y. A. I. Mohamed. Cloud-based visual analytics for smart grids big data. In *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, Sept 2016.

[21] Nonblocking system architecture, `http://www.onjava.com/pub/a/onjava/2002/09/04/nio.html?page=2`, 2017.

[22] P. A. Parikh and T. D. Nielsen. Transforming traditional geographic information system to support smart distribution systems. In *2009 IEEE/PES Power Systems Conference and Exposition*, pages 1–4, March 2009.

[23] R. Shyam, Bharathi Ganesh H B, Sachin Kumar S, Prabaharan Poornachandran, and K. P. Soman. Apache spark a big data analytics platform for smart grid. 21:171–178, 2015.

[24] The right big data technology for smart grid – distributed stream computing, `https://www.accenture.com/us-en/`, 2014.

[25] Wenlu Yang, A. Da Silva, and M. L. Picard. Computing data quality indicators on big data streams using a cep. In *2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, pages 1–5, Oct 2015.

Paper D

Automation of smart grid operation tasks via spatio-temporal exploratory visualization

Maria Stefan, Morten H. Andreasen, Jose G. Lopez,
Michael Lyhne and Rasmus L. Olsen

The paper has been submitted to:
The Journal of Environment and Planning B: Urban Analytics and City Science, 2019

Abstract

The evolution of electricity production is moving towards the use of 100% green energy resources, influenced to a large extent, by worldwide policies supporting the climate mitigation agenda. Consumers are evolving into energy producing prosumers, shaping the future of low-voltage grids into a distributed architecture.

State-of-the-art Advanced Metering Infrastructure (AMI) empower Distribution System Operators (DSOs) to collect various end-point status and condition data, in order to improve the grid operation, management and planning. In addition, the increase in data availability and granularity opens up the possibility for investigating the new possibilities and benefits of integrating data-driven models into the operation and management processes for electrical grids. The purpose of this article is to explore opportunities for designing a competent state-of-the-art monitoring and visualization system for low-voltage electricity grids. Primarily, the objective is to highlight the benefits of moving from traditional operation support systems in the field, to modern smart grid cyber-physical systems, allowing among other features, immediate grid monitoring awareness and preemptive actions.

The research and proof-of-concept in this work has been developed in close collaboration with a DSO in Denmark, stating the real-world requirements such a system must fulfill, in order to provide some benefits compared to the already existing system. These benefits are evaluated and quantified by migrating the traditional monitoring operations and workflows (extracted from on-site surveys) to the proposed reference models, possible due to the massive deployment of AMI the DSO has done in the past few years. The results indicate that a high level of automation is possible when combining the already existing AMI infrastructure with data-driven processes and visualization in the evaluated cases. This automation implies a significant reduction of required time and human resources to resolve the investigated conflicts, decreasing the total operational costs (OpEx) of the grid.

I Introduction

In consideration of the rapid evolution of electricity grids, this study proposes a software solution for electrical grid monitoring and planning, by data analysis and feature presentation [15]. The solution aims to provide the Distributed System Operators (DSO) with an efficient decision support system which can accommodate for a large number of prosumers (low-voltage customers with both demand and generation), variety in the incoming metering data (Advanced Metering Infrastructure - AMI readings) and automatic error detection. The achieved solution comes in the form of a visualization prototype, dedicated for users (DSOs) with different education and professional backgrounds [16]. Data visualization is important as it enhances the decision making processes due to the large implication of the human factor in the current grid management operations [7].

The work has been carried out departing from on-site interviews with employees from an energy company in Denmark. Their feedback is valuable for shaping different user profiles and for understanding their needs. Current daily workflows in error debugging are analyzed to identify the most time consuming operations regarding the low-voltage electrical grid. The preliminary user studies are then used as input to sketch the design and implementation of an appropriate visualization prototype. Findings from the assessment of the implemented visualization underline the importance of data presentation in practical applications and the benefits offered by geographic information systems (GIS) in terms of spatial awareness.

II Related work

A. Smart grids

Electricity grids undergoing the evolution towards Smart Grids [1] and the subsequent myriad data generators, naturally brings grid monitoring and control systems into the Big Data era. The significant increase in data variety and volume unlocks the possibility to exploit this data resource, to develop new functionalities and to improve performance via integrated systems, business intelligence, geographic information systems, big data processing and analytics technologies, etc [23] [13].

The inclusion of consumer power generators, primarily photovoltaic and wind generators, impose challenges for an electricity grid conceptually designed for one-way electricity flow. Thus, modern grid interoperability necessitates both enhanced monitoring, tighter control and regulation capabilities [9]. The smartness in grid monitoring operations is introduced in this work by integrating the low-voltage electrical grid topology with time-series metering measurements, both historical and near-real-time [17], and with visual information provided by Earth API.

B. Data management and system design

Continuous data generation requires that it is presented into comprehensible and perceptible formats by humans.

Originating from a data generator, such as an AMI customer, the data goes through a chain of procedures before it is either presented and or stored. Common data manipulation procedures include: sorting, feature extraction, processing, analytics and visualization [20] [19] [10] [5].

Where traditionally there is a tradeoff between processing speed and information precision, current data processing architectures theoretically display performance and capabilities where no compromise is necessary [3] [14].

However the actual performance of an application is highly implementation dependable and it is taken into consideration in this study, by evaluating the improvements in the DSOs' workflows, brought by the proposed visualization system.

C. Data visualization and GIS

Typically, visualization of selective grid parameters is presented in dashboard style or by individual system interfaces with a disperse representation by the different subsystem components [15] [12] [22], requiring switching and exchanging data between systems for either continued work flow or comparison.

Geographic Information Systems (GIS) make it possible for utilities to smartly operate their electricity grid, by interlocking data management and analysis, situation awareness, grid planning and workforce automation [6]. Situation and spatial awareness in particular bring smartness to the existing grid, enabling the utilities to detect and solve a power outage before customers call in [2] [4]. Aside from the data analytics platform, GIS also offers the possibility to outline possible places for renewable energy resources, thus it is also a tool for green energy planning.

Moreover, workforce automation can be obtained through WebGIS development and sharing of information via maps from office to field, due to its compatibility with both desktop and mobile devices [21], which is why it has been utilized for implementation in this work.

III Sequence models and user stories

Thee following section presents two of the most common scenarios for failure debugging. The two cases are evaluated both from a mathematical and sequence modeling point of view, with the purpose of extracting wokflow patterns for the DSOs.

A. Flickering of lights

One of the most frequent issues reported by the consumers to the DSOs is the flickering of lights. This use case is interesting for studying due to the DSOs different steps and approaches for evaluating and eventually solving the error. A general sequence model diagram for the case of lights flickering is presented in Figure D.1.

The scenario depicts the case of a customer calling in with an undervoltage error, claiming that their lights are flickering. In this scenario, the DSO follows the procedure of looking up the phase voltages in a *simplified user experience (UX) AMI interface*. Aside from this, two other programs are used: *a* *billing management software* and an Excel spreadsheet as *calculation software*. To examine the cause of the undervoltage, the DSO uses a *GIS tool* to find other consumers connected to the same cable box as the calling consumer. At the cable box level it is possible to evaluate if this is a multiple issue across the low-voltage grid consumers, in which case the problem is labeled as internal and has to be passed on to the network planning department for further investigation.

If the reported issue is local with respect to the single cable box found on the GIS map, the DSO will proceed to the next step: calculating the grid distance from the consumer who reported the error to the nearest transformer station, in order to determine the theoretical values which should have been measured instead of the erroneous ones. To find these, the DSO changes back and forth between the AMI interface and the map to evaluate distances, as well as values from the individual smart meters. This calculation is repeated until the point of reading the next transformer station (action labeled as "Loop" in Figure D.1). Furthermore, the billing management software provides the calculation of the amount of power generated and consumed from personal renewable energy sources.

The next step of the procedure is to set up a measuring box in the cable box related to the calling consumer. This is done in order to collect three-phased voltage measurement for one week period. This data is analyzed and the DSOs are then able to see whether they do or do not meet the requirements of normal voltage boundaries: ± 10 % of the nominal voltage value (230 V) for 95% of the time. If these requirements are not met, the previously made calculations are useful in deciding what kind of electrical grid reinforcements should be applied: installing bigger cables or transformers. These actions are also depending on the type of user (household or company) and operator. If the voltage is within boundaries, it is the customer's responsibility to contact their own electricity company.

It can be seen from the DSOs approach and work process that the above-mentioned case of light flickering at the consumer end is not a trivial one. The first step of checking the three-phased voltage measurements of the calling consumer is based on the latest identified values and does not take into account any historical recordings of data. Even though the calculation loop seems time consuming, in the rural areas there are typically 1-2 connected customers per cable box, for which these calculations can be done fairly quick.

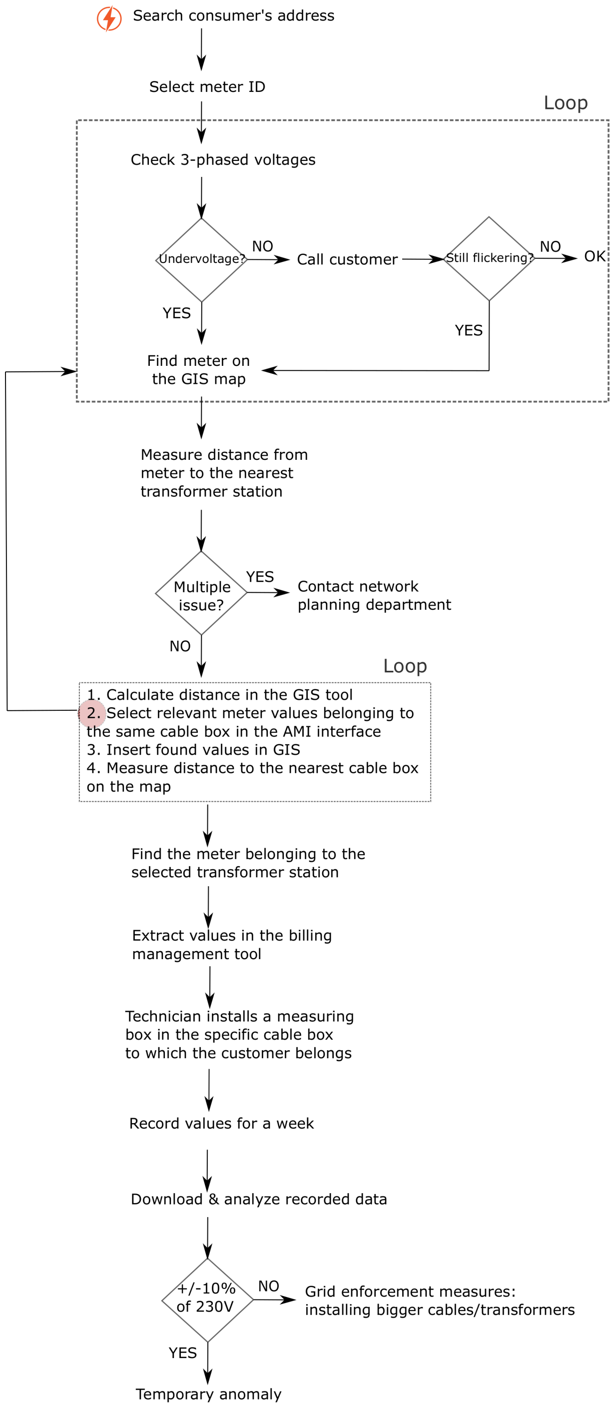


Fig. D.1: Sequence model representing the DSOs' actions in case a household reports constant light flickering.

However, the complexity increases with the number of connected customers like in the case of dense areas (i.e. cities), where the number of cable box connected areas multiply.

However, depending on the severity of the issue, a continuous flickering of lights could potentially turn into a power outage case, which was introduced in [18]. The DSOs would have to identify the affected area and inform all the corresponding customers about it, which leads to examining the sequence model of this use case.

The total time spent on debugging the case of light flickering (T_{flick}) can be modeled as in Equation D.1:

$$T_{flick} = t_s + \sum_{i_m=1}^{m} t_{i_m} \left(t_{call} + t_{GIS}\right) + t_d \times \left(m \times d_m \sum_{i_m=1}^{m} (m-1) t_{i_m} \left(t_{call} + t_{GIS}\right) + t_{map} + t_{d_{cab}} \right) + t_{SW} + t_{rec} \times t_{install} \quad \text{(D.1)}$$

Which can be expressed as Equation D.2:

$$T_{flick} = t_s + \left(t_{call} + t_{GIS}\right) \sum_{i_m=1}^{m} t_{i_m} + t_d \times \left(m \left(t_{map} + t_{d_{cab}}\right) d_m + (m-1) \left(t_{call} + t_{GIS}\right) \sum_{i_m=1}^{m} t_{i_m} \right) + t_{SW} + t_{rec} \times t_{install} \quad \text{(D.2)}$$

Where the variables denote:

- m: number of meters;
- i_m: meter IDs;
- t_s: time to search a consumer's address;
- t_{i_m}: time to select the meter ID and to check the three-phased voltage values;
- t_{call}: time to call the customer;
- t_{GIS}: time to find the meter on the GIS map;
- t_d: time to measure the distance to the nearest transformer station;
- d_m: time to calculate the theoretical distance on the map;
- t_{map}: time to insert meter values in GIS;
- $t_{d_{cab}}$: time to measure the distance to the nearest cable box on the map;
- t_{SW}: time to find the meter in the selected transformer station and to extract the values in the billing management software;
- t_{rec}: time to install, record and analyze one week worth of measurements in the specific cable box.
- $t_{install}$: time to perform grid reinforcement measures.

The human factor impacts all the aforementioned timings and it is therefore not possible to assign exact numbers to them. However, it can be drawn from

Equation D.1 that the complexity of the working flow is highly dependent of the number of meters (m) and on the meter IDs (t_{i_m}). A large m results in searching across multiple t_{i_m} for possible anomalies and thus creating redundancy between the two loops.

B. Identifying the area in the electrical grid that was affected by an error and inform the corresponding customers

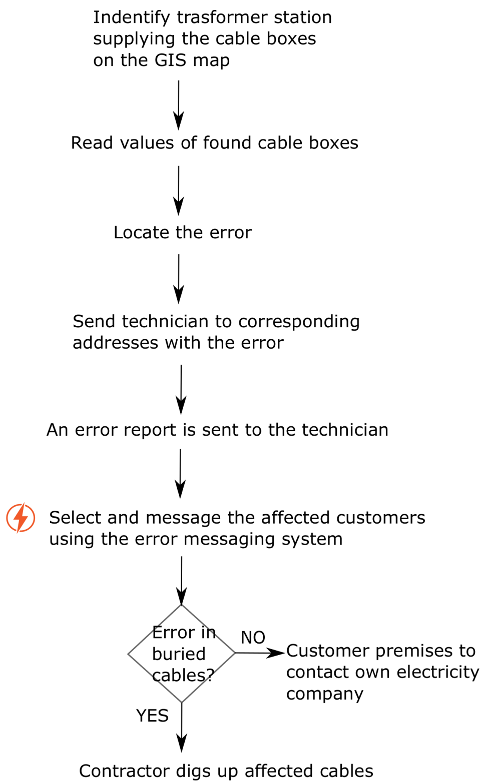


Fig. D.2: Sequence model representing the DSOs' actions for informing customers about an affected area in the electrical grid.

This is the procedure followed by the DSOs after they have identified that they are responsible for an error, as presented in Figure D.2. This is done either by detecting the error through a similar sequence as in Figure D.1 or by a private technician informing them. Determining the exact area that has been affected by the error is mainly done using the AMI interface and the map. When the area has been identified, a technician is sent to address the issue and the affected consumers are informed through the error messaging system.

This use case is mainly presented to strengthen the motivation for why there is a need for creating inter dependencies between different software tools utilized by the DSOs, for example between the map and the error messaging system. There is still the risk of incorrectly selecting and messaging the affected consumers, due to the independence among these tools.

The total time spent on identifying a grid area with failures (T_{msg}) can be modeled as in Equation D.3:

$$T_{msg} = t_{GIS} + t_{read} + t_{err} + t_{tech} + t_m + t_{dig} \tag{D.3}$$

Where the variables denote:

- t_{GIS}: time to identify a transformer station on the map;
- t_{read}: time to read cable box values;
- t_{err}: time to locate the error;
- t_{tech}: time in which the technician reaches the address which issues errors, while receiving the error report;
- t_m: time to message affected consumers via the error messaging system;
- t_{dig}: time to dig the affected cables.

The total T_{msg} in Equation D.3 depends on the number of meters/customers (denoted as m) which are affected by an error and it is as well influenced by the human impact in the debugging process.

The system operators use four different programs to assess different aspects of the low voltage grid: the AMI interface, the map, the billing management tool and the calculation software. This may seem as a time-consuming way of monitoring the grid, due to the lack of efficient integration among these programs. Furthermore, it is observed that in the case of the Excel calculation, the DSOs need to keep track of different parameters, such as customer and cable types, which the DSOs have to manually look through. An advantage would be to have a visual overview over different stages in the debugging process, in connection with speeding up the error identification. Due to the subjective nature of these sequence models, the following implementation aims to:

- Minimize the risk of incorrect actions during the DSOs working routine;
- Propose an data-driven integrated visualization solution as alternative to the current manual procedures.

IV Implementation

This section contains a description of a visualization prototype implementation, which aims to bring more automation to the procedures identified in Section III.

The test area in the electrical grid is located Thisted region, in Denmark [8]. The corresponding data is managed in a PostgreSQL database which contains medium and low-voltage grid topology (geographical information), attributes of the different entities and nodal measurements. The implementation is based on ASP.NET developing stack to create the WebGIS visualization application.

An overview of the WebGIS menu application is presented in Figure D.3 for the test area. The map menu layer contains:

- customers - households and companies from the low-voltage layer;
- secondary substations - belonging to the medium-voltage layer;
- cable boxes - entity containing the electronic equipment which links the medium to the low-voltage levels;
- low-voltage (LV) wires - cables showing the interconnections between customers and secondary substations;
- medium-voltage (MV) wires - cables showing the interconnections among substations;
- high-voltage (HV) wires - cables showing the interconnections between primary and secondary substation. No such information was available in the data, but the implementation is ready for including this as well.

Fig. D.3: Overview of the WebGIS application and its features. The medium-voltage part of the visualization includes secondary substations and their interconnections

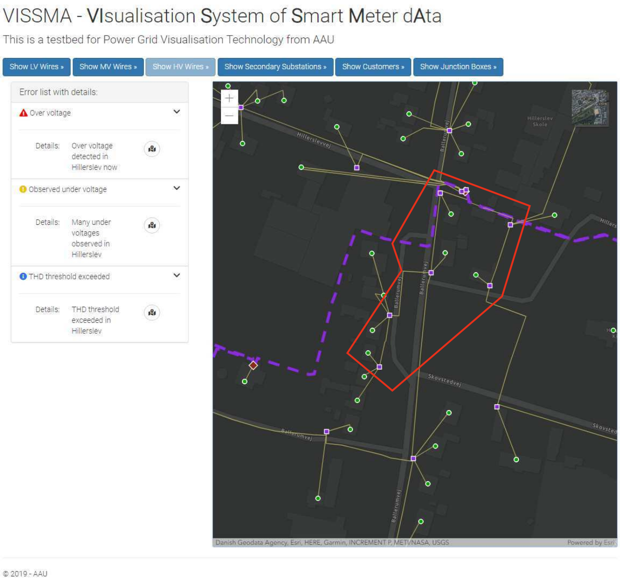


Fig. D.4: Interconnections between the different kinds of nodes - from MV to LV. The group of connected cable boxes is useful for rerouting algorithms.

An error listing functionality is also included in the left-hand side of the map. Possible errors are prioritized and represented according to their severity as: errors, warnings and other informative alarms. A description about their nature can be obtained by clicking on the down arrow to inspect the possible causes of the errors (Figure D.4).

The purpose of this application is manifold, providing the means for grid planning, monitoring and prediction.

A. Planning

As desired, the WebGIS page allows for selecting a certain layer of the electrical grid, just as it is done in Figure D.3 depicting the MV grid (secondary substations) and their interconnections. Different grid operations require various possible ways of displaying the data. In this case, having only the graph gives the opportunity to evaluate the current medium-voltage topology and assess the possibility of extending or changing the current topology by applying graph networks algorithms. This view is mostly usual for grid reinforcements and planning, for example in the case of light flickering (Figure D.1) where optimization procedures might be needed for the medium voltage part of the grid. The same procedures can be applied to the low-voltage grid if they are considered necessary.

B. Monitoring

An alternative to the graph view is adding an Open Street Map (OSM) layer to the WebGIS. This is done by selecting the view on the view button in the upper right corner of the map, obtaining an Earth outlook as shown in Figure D.5. This representation is mostly useful for grid monitoring, as the landscape information can be mined with the existing topology and measurements.

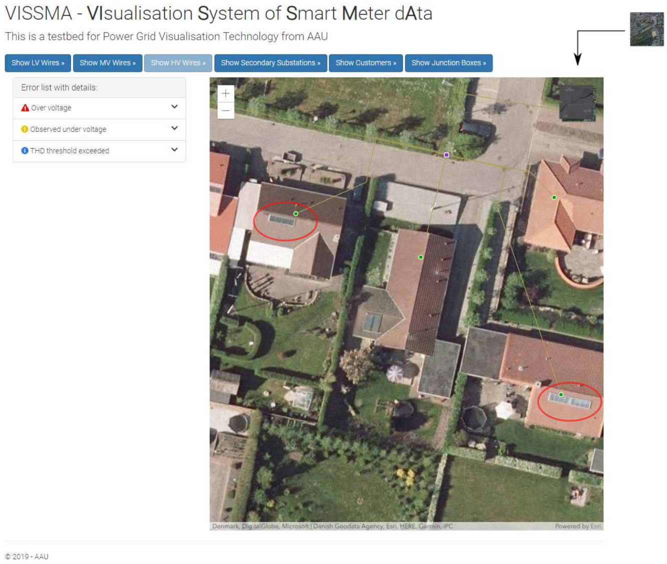


Fig. D.5: Earth view of a customer with a solar panel.

For example, it can be seen that the household shown in Figure D.5 has a solar panel installed on the roof. This is valuable information for the DSOs, especially in case they register inexplicable changes in the customers' generation and consumption patterns. It is possible that the some of the topology data does not contain information about whether a certain customer has renewable energy resources (RES) and, like in this case, the DSOs can visually evaluate it on the map.

More information can be obtained by clicking on the node associated to that customer. A window like in Figure D.6 reveals details about the selected node. In this case, the field denoted "Prosumer Generator Type" confirms that this specific household has a solar panel, meaning that this information was already updated in the initial data set.

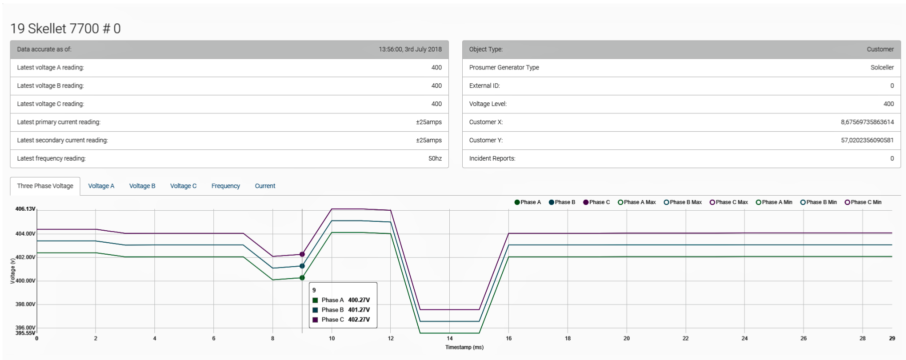


Fig. D.6: Window pop-up showing the last measured values and historical time-series plots of a certain node.

The pop-up window comprises both the latest measurement readings and historical time-series plots for values measured in the past 30 milliseconds. This granularity can be adjusted according to what is relevant for the DSOs in each application.

The ID of the selected node is denoted in the left upper corner with "#", after the address ("19 Skellet 7700"), and it can also be found in the "External ID" field. The date and time of the latest registered measurement are also displayed. As it was required by the DSOs to visualize the three-phased voltage values in every node, these are as well recorded and laid out as both latest value and historical data. Each phase measurement can be individually analyzed at any point in time by moving the mouse cursor on the plot (Figure D.6).

Current and frequency values are also available and can be analyzed if relevant, in the same manner as it is done with the voltage.

The "Voltage level" field denotes the layer of the grid to which the node belongs. The value of 400 V indicates that the node is placed in the low-voltage level, while values of 10 and 60 kV would refer to nodes in the medium-voltage level.

Information about certain events is also available in the pop-up in "Incident Reports". It contains the numbers of the incidents that have been reported for a selected node and it is related to the error tab available on the front page. A recorded event may refer to an error in the measurements (i.e. negative values of consumption energy), an alarm triggered by a voltage measurement outside the limit of ± 10 % of the nominal voltage value of 230 V (Figure D.1); or alarms and warnings generated by meters recording 0 V measurements for a long period of time (i.e. cable faults which can be caused by someone digging in the ground).

Other nodes on the map are represented by cable boxes and secondary substations. Their detailed information can also be accessed through a similar window pane as in Figure D.7. "CP3" is the identifier for a cable box and "Holtab 630" for a secondary substation.

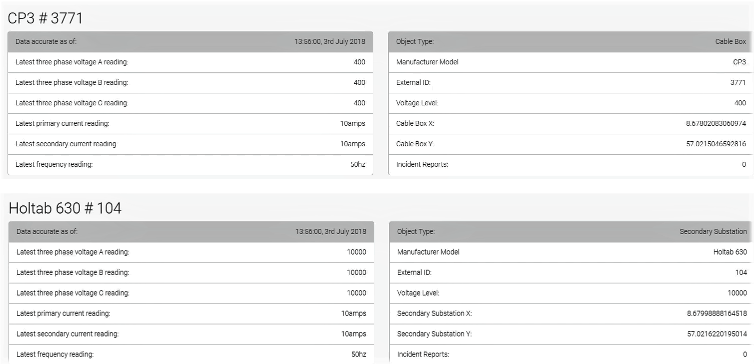


Fig. D.7: Detailed information about a cable box and a secondary substation.

C. Prediction

The linear representation of the grid topology may be utilized for prediction purposes, provided the knowledge extracted from the planning and monitoring. Figure D.4 depicts the connectivity among a group of cable boxes. Such information is useful for manipulating the existing tabular information of the grid layout and exploring the possibilities and benefits given by new topologies. Therefore, load analysis can be obtained by rerouting the information in the electrical grid network based on the different possible layouts. From this, power redistribution algorithms can be designed to predict, for example, the most vulnerable areas in the power grid with respect to power failures.

V Assessment

As the purpose of this study is to explore how the DSOs working routine can be optimized, this section contains the assessment of the WebGIS application presented in Section IV, by evaluating how the different implemented features have improved the debugging process. This application was already presented in a conference workshop [11], where both vendors for smart metering solutions and DSOs were present.

The diagram in Figure D.8 shows the new work flow process of the DSOs using the WebGIS application. It can be noticed that this process model links the individual procedures presented in Figures D.1 and D.2. As a result of a more efficient software integration, events and errors are identified automatically, making the operators aware of them before customers calling in. By integrating the different computational models (i.e. calculating the distance

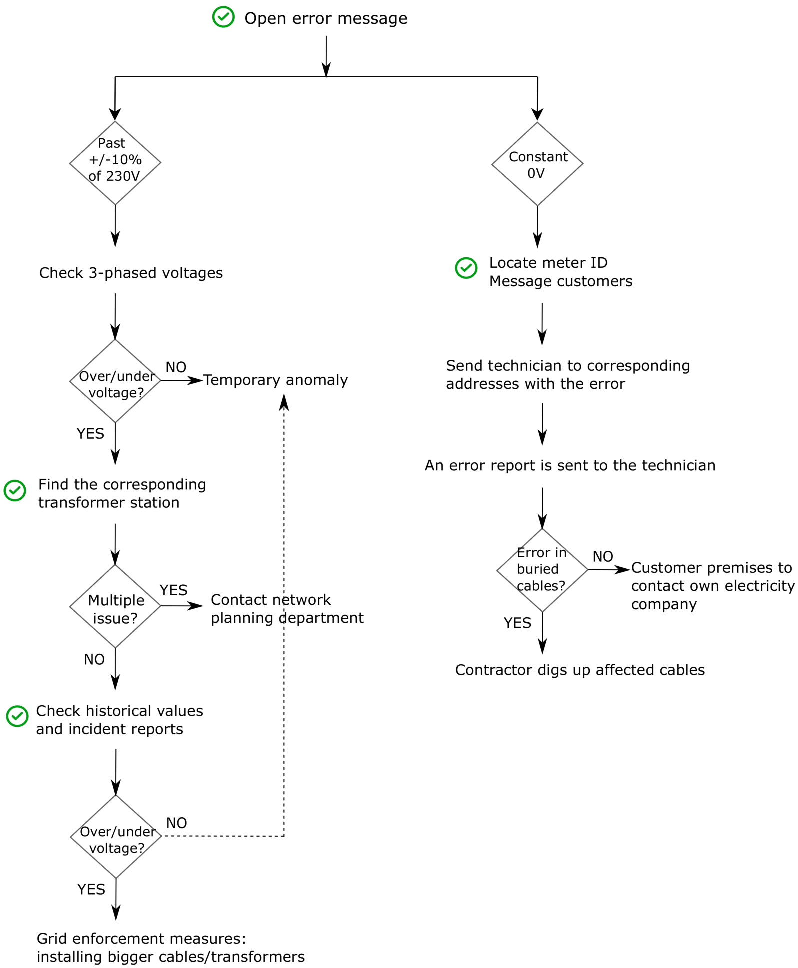


Fig. D.8: Updated sequence model for procedures of resolving two error cases using the WebGIS prototype.

from a node to the nearest transformer station) with the spatial awareness achieved via GIS, clear improvements have been achieved in the process flow in Figure D.8, where they are denoted with a green check mark:

- Fast error identification, particularly in case of collective issues (from MV to LV) - most relevant for grid *monitoring*;
- Instant visual correlation of a meter to its corresponding transformer station (and vice-versa), due to the connectivity information in the grid topology - most relevant for grid *monitoring*, as well as *planning*;
- Integrated calculation of the theoretical metering values and compari-

son with the actual measurements (both near-real-time and historical), where mismatches are shown as incident reports - most relevant for *prediction* of potential incidents;
- Efficient use of data structures to quickly identify a desired node ID without any manual searching - relevant for *data integrity* and grid structure *maintenance*.

By eliminating the loops in the new process model, the number of debugging actions performed is also reduced in each type of error, thus minimizing the total debugging time. This can be modeled by Equation D.4.

$$T_{debug} = \begin{cases} t_{i_m} + t_{tran} + t_{hist} + t_{install}, & \text{if } V \notin (\pm 10\% \cdot 230) \\ t_m + t_{tech} + t_{dig}, & \text{if } V = 0 \end{cases} \tag{D.4}$$

Where the variables denote:

- t_{i_m}: time to check the voltage values;
- t_{tran}: time to find a transformer station on the map. Corresponds to t_d in Equation D.1;
- t_{hist}: time to check historical measurements and incident reports. Corresponds to t_{rec} in Equation D.1;
- $t_{install}$: same as in Equation D.1;
- t_m: time to locate a meter and message the corresponding customer. Corresponds to t_m in Equation D.3;
- t_{tech}: same as in Equation D.3;
- t_{dig}: same as in Equation D.3.

The influence of the human factor is minimized due to integrating the computational models into the same application and thus reducing the total time dependency per work flow to four variables in the over/under voltage case. A cutback in the number of operations can also be seen in case of recording 0 voltage measurements for a long period of time.

Visual spatial awareness eliminates the risk of selecting and messaging the irrelevant customers and helps the operators keep track of the current state of the grid, while focusing on the most vulnerable areas. The spatial benefit was also acknowledged by both the smart meter vendors and the DSOs present in the SmartGridComm workshop [11], as it gives the possibility to further exploit the potential of the existing metering infrastructures, which are currently used only for billing.

This implications of the visualization prototype are summarized in Table D.1.

Table D.1: Benefits and drawbacks of the proposed visualization application.

Benefits	**Drawbacks**
• improved accuracy and process redesign	• migration requirements to data-driven systems
• system redesign for optimization of the DSOs' workflows	• design standardization of the model
• improved OpEx by integrating data-driven business processes and analytics	• CapEx investment required for data storage and processing power (either physical or cloud-based)
• unified software solution for grid operation	

VI Conclusion and Future work

Visualization and interpretability in practical applications must take into account the human cognitive factor that any knowledge extraction process entails, according to the specific requirements of the application area. A key idea in this study is that the main reason for a continuous integration (CI)-based system is the manual and time consuming error debugging process. For the Danish DSOs, interpretation is a tool to optimize their current workflow procedures and to acquire new knowledge about the low-voltage grid, which can be used for monitoring, planning and event prediction.

Many of the present-day big data management solutions based on cloud applications can accommodate for varied and voluminous data to be stored, processed and analyzed, but also offer system interoperability between vendors and DSOs. At the same time, WebGIS applications are flexible to be implemented according to the users' specific requirements, different profiles and working procedures and they are suited for mobile-based solutions.

Additionally, migrating towards a fully integrated system takes advantage of feature synergies and opens possibilities for development and adaption of enhanced functionalities, and as a result simplifying daily operations and workflows. This is done by integrating a grid model, continuous smart meter monitoring, geographic information and visualization techniques, thus offering spatial and situation awareness in power grid management. Data-driven system ingration results in eliminating the need for an operator to manually switch, transfer or compare data between different systems, resulting in reduced error rates via automatic data validation as well as improved workflow efficiency by minimization of operator processes.

The WebGIS application presented in this work brings benefits for low-voltage grid data management, by decreasing anomaly detection time and, as a consequence, by reducing the power grid OpEx.

A feasible approach in the future work may imply more iterations in the user analysis study, using the current prototype as basis for new on-site interviews with the DSOs and upgrading the application design accordingly. Moreover, the transition to mobile-based WebGIS should be taken into consideration, particularly for the on-call DSOs.

VII Copyright statement

A. Copyright

Acknowledgment

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 related to paper D

[1] S. S. Bhalshankar and C. S. Thorat. Integration of smart grid with renewable energy for energy demand management: Puducherry case study. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1–5, Oct 2016.

[2] Deborah Byrd. Bill meehan on using gis to help create a smart grid, `https://earthsky.org/human-world/bill-meehan-on-using-gis-to-help-create-a-smart-grid-2`, 2011.

[3] V. Dehalwar, A. Kalam, and A. Zayegh. Infrastructure for real-time communication in smart grid. In *2014 Saudi Arabia Smart Grid Conference (SASG)*, pages 1–4, Dec 2014.

[4] Charles H. Drinnan. Smart grid 2.0: The role of gis, `https://electricenergyonline.com/energy/magazine/621/article/Smart-Grid-2-0-The-Role-of-GIS.htm`, 2012.

[5] A. El Khaouat and L. Benhlima. Big data based management for smart grids. In *2016 International Renewable and Sustainable Energy Conference (IRSEC)*, pages 1044–1047, Nov 2016.

[6] Esri. Gis for smart grid, `https://www.esri.com/library/brochures/pdfs/gis-for-smart-grid.pdf`, 2012.

[7] Meiken Hansen and Bettina Hauge. Scripting, control, and privacy in domestic smart grid technologies: Insights from a danish pilot study. *Energy Research and Social Science*, 25:112 – 123, 2017.

[8] Ruben Sánchez Martin-Loeches, Florin Iov, Mohammed Seifu Kemal, Maria Stefan, and Rasmus Løvenstein Olsen. Observability of low voltage grids: actual dsos challenges and research questions. In *Proceedings of the 2017 52nd International Universities' Power Engineering Conference (UPEC)*. IEEE Press, 8 2017.

[9] Patrick Matschoss, Benjamin Bayer, Heiko Thomas, and Adela Marian. The german incentive regulation and its practical impact on the grid integration of renewable energy systems. *Renewable Energy*, 134:727 – 738, 2019.

[10] Amr A. Munshi and Yasser A.-R. I. Mohamed. Big data framework for analytics in smart grids. *Electric Power Systems Research*, 151:369 – 380, 2017.

[11] Rasmus Løvenstein Olsen. Challenges and solutions in future distribution grids. In *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, `https://sgc2018.ieee-smartgridcomm.org/challenges-and-solutions-future-distribution-grids`, 2018.

[12] Thomas J. Overbye and James Weber. Smart grid wide-area transmission system visualization. *Engineering*, 1(4):466 – 474, 2015.

[13] Miloš Radenković, Jelena Lukić, Marijana Despotović-Zrakić, Aleksandra Labus, and Zorica Bogdanović. Harnessing business intelligence in smart grids: A case of the electricity market. *Computers in Industry*, 96:40 – 53, 2018.

[14] S. Roy, B. Bedanta, and S. Dawnee. Advanced metering infrastructure for real time load management in a smart grid. In *2015 International Conference on Power and Advanced Control Engineering (ICPACE)*, pages 104–108, Aug 2015.

[15] Maria-Angeles Sanchez-Hidalgo and Maria-Dolores Cano. A survey on visual data representation for smart grids control and monitoring. *Sustainable Energy, Grids and Networks*, 16:351 – 369, 2018.

[16] Lea Schick and Christopher Gad. Flexible and inflexible energy engagements—a study of the danish smart grid strategy. *Energy Research and Social Science*, 9:51 – 59, 2015. Special Issue on Smart Grids and the Social Sciences.

[17] Maria Stefan, Jose G. Lopez, Morten H. Andreasen, Ruben Sanchez, and Rasmus Olsen. Data analytics for low voltage electrical grids. pages 221–228, 01 2018.

[18] Maria Stefan, Jose G. Lopez, and Rasmus 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)*, pages 3543–3548, 2018.

[19] A. Vaccaro, I. Pisica, L.L. Lai, and A.F. Zobaa. A review of enabling methodologies for information processing in smart grids. *International Journal of Electrical Power and Energy Systems*, 107:516 – 522, 2019.

[20] Tom Wilcox, Nanlin Jin, Peter Flach, and Joshua Thumim. A big data platform for smart meter data analytics. *Computers in Industry*, 105:250 – 259, 2019.

[21] Kehe Wu and Zhibin Zhang. Research and implementation of smart transmission grids based on webgis. In *2010 Second International Conference on Communication Systems, Networks and Applications*, volume 1, pages 302–307, June 2010.

[22] Lin Xiqiao and Yang Zhou. Analysis of large-scale electric vehicles charging behavior using data visualization. In *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, pages 384–388, May 2017.

[23] Soo Wan Yen, Stella Morris, Morris A.G. Ezra, and Tang Jun Huat. Effect of smart meter data collection frequency in an early detection of shorter-duration voltage anomalies in smart grids. *International Journal of Electrical Power and Energy Systems*, 109:1 – 8, 2019.

Paper E

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

A data sanity analysis

Maria Stefan, Jose Gutierrez, Pere Barlet, Oriol Gomis
and Rasmus L. Olsen

The paper has been submitted to:
The Journal of Applied Energy, 2019

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.

I 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 [3]). Currently, the power grid supports up to 43% power generation from wind turbines [9] [7] and can operate with up to 50% power supplied by combined heat and power (CHP) plants [1]. Consequently, as the penetration rate of renewable energy sources (RES) intensifies, the low-voltage power grid becomes active as consumers evolve into prosumers, emphasizing decentralization, influencing daily grid operation and management for the distributed system operators (DSOs).

Currently, the incoming metering data is used only for billing purposes [13]. 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 [16] . 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 consumers'/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.

II 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. In general, understanding the energy behavior at the low-voltage grid can be done by clustering households by specific attributes, which are 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 [10].

For forecasting purposes

Electricity short term load forecasting (STLF) applied to historical customer data is addressed in [5], 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 [4]. The model was obtained through data preprocessing, correlation clustering and discrete classification NN (Neural Network). The total energy demand was adjusted according to the total energy forecast.

For monitoring purposes

The study in [15] 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 operational planning is addressed in [11], resulting in the following services:

- 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;
- 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 [8]. The gained knowledge gives useful information regarding energy fraud detection, home invasion or children behavior. However, this method does not address the privacy issues that arise from the communication network infrastructure or from storing household-related data.

III A data sanity study for low-voltage electrical grids

A. Data system and data flow

The information exchange in the electrical grid corresponding to the RemoteGRID project [13] [14] is depicted in Figure E.1. The three actors defined

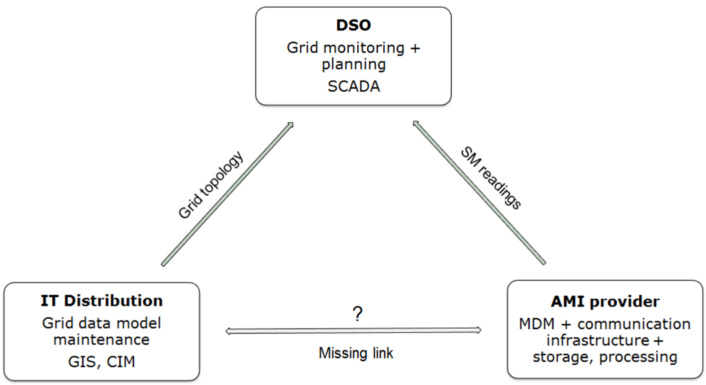


Fig. E.1: Information exchange among the meter provider, IT distribution and the DSO.

in Figure E.1 show the relationships between the DSO, AMI provider and IT distribution. The AMI provider [12] 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 [2] 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 [6] 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.

B. Data types

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

- *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 Figure E.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.

- *Time series data*
  Active and reactive energy measurements are provided for a period of one year (from April 2017 to April 2018 inclusively), with a granularity of 15 minutes, which is defined by the current metering infrastruc-

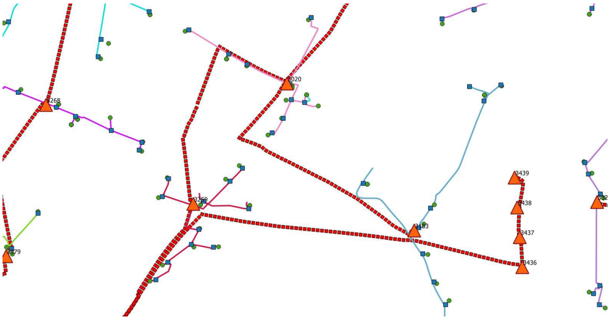


Fig. E.2: Medium and low-voltage grid topology sample.

ture. The data model for the time-series measurements is shown in Figure E.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. This topic will be covered in Subsection D..

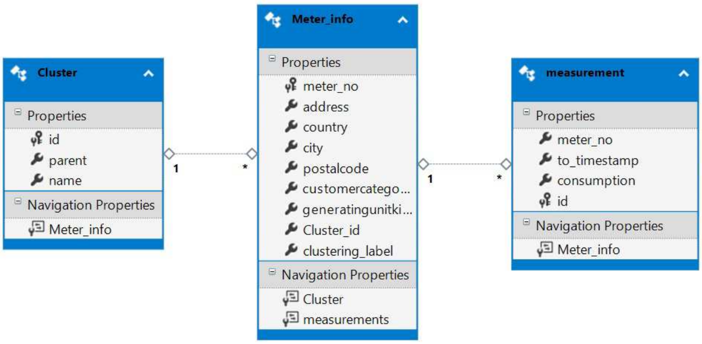


Fig. E.3: Data model for the time-series measurements.

The meter ID field (obtained from the distribution company involved in the study [2]) 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.

C. User profile analysis - labeled data

This subsection illustrates the customer profiles obtained from data related to substation 3011, which is shown in Figure E.4. This substation is chosen due to the presence of PVs, as well as for comprising of both apartment buildings and stand-alone houses. The blue lines represent the connections between the substation and its consumers.

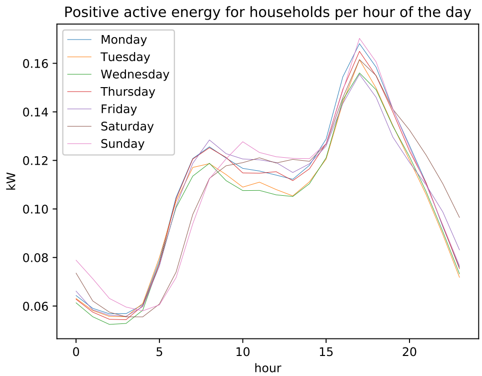
Two types of labeled customers have been identified - households and companies, which will be used as starting point in the analysis. Some of the companies are labeled with PVs, but there is no information about RES at household level.

Fig. E.4: Low-voltage grid topology information for secondary substation 3011.

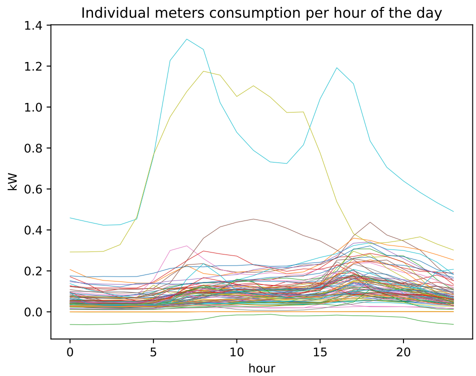
- Household profiles

  The plot is Figure E.5 shows the household consumption statistics (in kWh). Subfigure E.5a 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).

  Subfigure E.5b 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. This information can also be evaluated statistically in Table E.1.



(a) Households' average consumption patterns per hour.



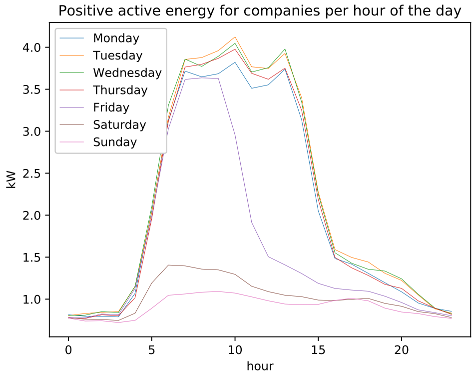
(b) Individual households' consumption patterns per hour.

Fig. E.5: Positive active energy plots for labeled households

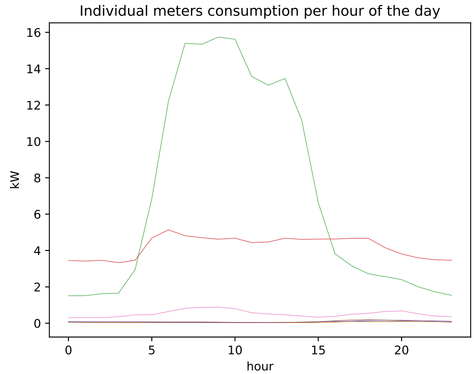
Table E.1: Statistics related to households and companies.

Label	Mean (kWh)	Min (kWh)	Max (kWh)	Variance	Standard deviation
House 4	0.127	0.01	3.41	0.028	0.166
House 7	0.156	0.00	5.13	0.198	0.445
House 15	0.787	0.11	8.84	1.029	1.014
Comp 9 (solar)	0.521	0.00	6.41	0.342	0.585
Comp 58 (not solar)	4.211	2.34	9.20	0.688	0.830
Comp 87 (not solar)	7.011	0.45	39.82	74.867	8.653

- Company profiles
  Similarly, the company consumption profiles (kWh) are presented in Figure E.6. The average consumption per year for every week day is shown in Subfigure E.6a. 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 Subfigure E.6b, 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 (Table E.1).



(a) Companies' average consumption patterns per hour.

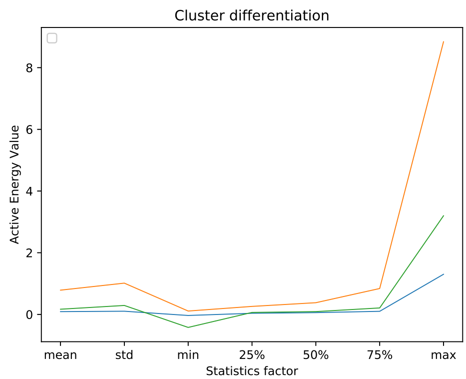


(b) Individual companies' consumption patterns per hour.

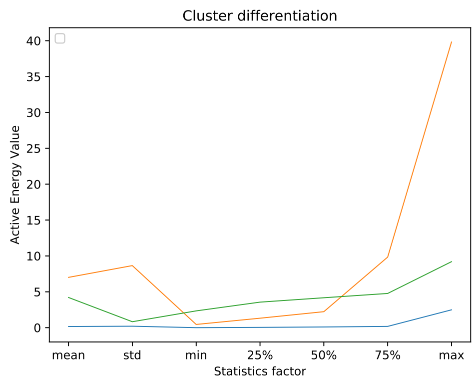
Fig. E.6: Positive active energy plots for labeled companies

D. 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 Figure E.7, for households (Subfigure E.7a) and companies (Subfigure E.7b).



(a) Cluster differentiation for households.



(b) Cluster differentiation for companies.

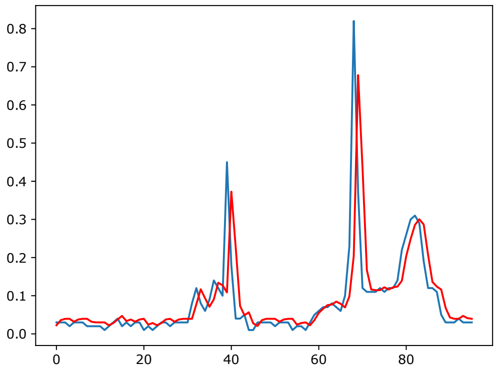
Fig. E.7: Clustering method based on active energy consumption values for labeled households and companies

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 Subfigure E.7b. 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.

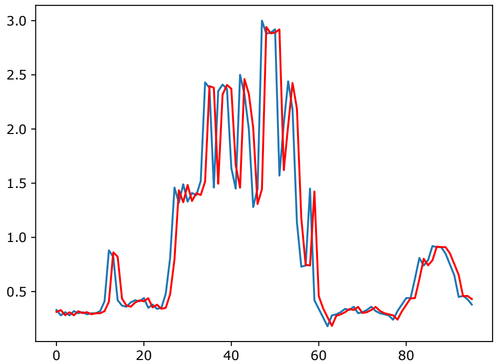
E. Customer behavior prediction

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 Figure E.8 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 Subfigures E.8a and E.8b 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. E.8: ARIMA predictions based on clustering for households and companies

IV Discussion

The data analysis presented in Section III 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-preocessing method, due to its abil-

Table E.2: 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, related to 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 certain 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

ity to correctly place the six different individual user consumption patterns into six corresponding clusters - three in each labeled category, household or company.

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 (weekdays/weekends, holidays and seasons) 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 subjective to a large extent. This study is important for the DSOs when taking into consideration future electrical grids consisting of 100% renewable and distributed energy resources, due to some 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 E.2.
- *Scalability* - the amount and diversity in the data incoming from the different distribution energy resources (DER) calls for a scalable and flexible data analysis solution. At the same time, the scalability may also refer to a collective group of operators (heat, water, transmission system operators) who need to use their 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 E.2, along with the corresponding analytical methods that can help overcome them.

These methods are useful as a data sanity checkup in the different situations where the available data is not labeled, 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.

V 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.

Acknowledgment

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. The authors would like to thank Eduardo Prieto, Assistant Professor at Universitat Politecnica de Catalunya, who helped defining the overall problem statement and interpreting the results.

References related to paper E

[1] Danish Energy Agency. Statistik, data, nøgletal og kort, February `https://ens.dk/service/statistik-data-noegletal-og-kort`, 2018.

[2] Thy-Mors Energi Elnet A/S. Beskrivelse af forskningsprojekter i elnettet, `https://www.tme-elnet.dk/om-os/forskningsprojekter`.

[3] Danish Energy Association. Smart grid in denmark 2.0, `https://www.usef.energy/app/uploads/2016/12/Smart-Grid-in-Denmark-2.0-2.pdf`, 2016.

[4] Christopher Bennett, Rodney Stewart, and Junwei Lu. Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system. 2014.

[5] Tobias Boßmann, Joachim Schleich, and Robert Schurk. Unravelling load patterns of residential end-uses from smart meter data. 06 2015.

[6] DAX. Network model management (nmm), `https://dax.dk/nmm-dansk-introduktion/`.

[7] Energistyrelsen. Månedlig og årlig energistatistik, `https://ens.dk/service/statistik-data-noegletal-og-kort/maanedlig-og-aarlig-energistatistik`.

[8] V. Ford and A. Siraj. Clustering of smart meter data for disaggregation. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 507–510, Dec 2013.

[9] Magnus Hornø Gottlieb. Danmark sætter ny rekord i vind, `https://www.danskenergi.dk/nyheder/danmark-saetter-ny-rekord-vind`, 2018.

[10] S. Haben, C. Singleton, and P. Grindrod. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid*, 7(1):136–144, Jan 2016.

[11] B. P. Hayes and M. Prodanovic. State forecasting and operational planning for distribution network energy management systems. *IEEE Transactions on Smart Grid*, 7(2):1002–1011, March 2016.

[12] Kamstrup. Intelligent solutions for electricity utilities, `https://www.kamstrup.com/en-en/electricity-solutions`.

[13] Ruben Sánchez Martin-Loeches, Florin Iov, Mohammed Seifu Kemal, Maria Stefan, and Rasmus Løvenstein Olsen. Observability of low voltage grids: actual dsos challenges and research questions. In *Proceedings of the 2017 52nd International Universities' Power Engineering Conference (UPEC)*. IEEE Press, 8 2017.

[14] Rasmus Løvenstein Olsen. Remotegrid - project description, `https://www.en.remotegrid.dk/projectdescription/`.

[15] M. Reis, A. Garcia, and R. J. Bessa. A scalable load forecasting system for low voltage grids. In *2017 IEEE Manchester PowerTech*, pages 1–6, June 2017.

[16] Maria Stefan, Jose G. Lopez, and Rasmus 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)*, pages 3543–3548, 2018.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-445-4

AALBORG UNIVERSITY PRESS