Jesse Christensen

CUSTOMER SPECIFIC LOAD MODELLING AND FORECASTING FOR SHORT-TERM OPERATIONAL PLANNING

- the use of hourly smart meter data

Master's thesis
Faculty of Information Technology and Communication Sciences
Pertti Järventausta
Kimmo Lummi
December 2023

ABSTRACT

Jesse Christensen: Customer Specific Load Modelling and Forecasting for Short-Term Operational Planning – the Use of Hourly Smart Meter Data
Master's thesis
Tampere University
Degree Programme in Electrical Engineering
December 2023

The ongoing energy transition towards sustainable energy, aiming to mitigate the effects of climate change is heavily transforming the entire electricity distribution sector. The surge of distributed, intermittent generation and electric vehicles along with increasing demand flexibility complicates the management of the entire electrical grid. Many tasks distribution system operators face in their daily work depend on modelling and forecasting network loads. Due to fundamental changes in electricity consumption patterns, load modelling is more difficult than ever, and existing load models have turned obsolete. Many distribution system operators in Finland and globally still rely on simple peak load estimation models and load profiles based on outdated measurement data in modelling and forecasting network loads, showcasing a need for new solutions.

Meanwhile, the transition towards smarter grids has opened new opportunities in load modelling, especially via the replacement of traditional electricity meters with smart meters, giving remote access to accurate load measurements from the entire network. The use of metering data for creating improved load modelling and forecasting methods has been researched, and many studies show promising results. This thesis employs a constructive research approach to develop a novel method for estimating and forecasting the loads of individual low-voltage network customers by analysing historical measurement data. The main objective of this work is to determine whether accurate estimations and forecasts can be created merely by analysing customer-specific historical load measurements.

MicroSCADA X is a product family by Hitachi Energy, comprising of multiple software and hardware solutions for distribution system control and supervision. DMS600, a software package of the product family, includes a distribution management system and a network information system, used by distribution system operators to document, monitor and control their networks. Since the existing load modelling functionalities of DMS600 are outdated, the system offers a good platform for testing the newly developed load model in a real-life environment.

In this work, the necessary theoretical background to understand the framework of this work is first covered. This includes examining the electricity metering process and presenting existing load modelling methods along with their use cases in distribution system management. The DMS600 software package is also introduced, and the weaknesses of its current load modelling functionalities are addressed. Once a theoretical background has been established, the process of designing the novel load modelling algorithm is reviewed. The algorithm is implemented into DMS600 and the accuracy of the new model is compared to the existing load profile model by creating an extensive set of short-term load forecasts for multiple real-life LV networks. The test set includes load forecasts for all seasons in three separate, relatively large urban LV networks.

The results prove that directly utilizing historical measurements in creating customer-specific load forecasts can provide significant improvements to short-term load forecast accuracy when compared to tradition load profile based models. In addition, the work shows that this type of a model can easily be integrated into distribution management software. However, the developed model had a slight tendency to underestimate the loads. Also, due to the limited time frame, some important variables could not be incorporated into the model. This left room for future development, while the results were extremely promising.

Keywords: electricity distribution, load modelling, load profiling, short-term load forecasting, smart grids, smart meters

TIIVISTELMÄ

Jesse Christensen: Asiakaskohtainen kuormitusten mallintaminen ja ennustaminen lyhyen aikavälin käyttötoimintaa varten – älymittareiden tuntienergiamittausten hyödyntäminen
Diplomityö
Tampereen yliopisto
Sähkötekniikan DI-ohjelma
Joulukuu 2023

Käynnissä oleva energiamurros kohti kestävämpää energiantuotantoa ilmastonmuutoksen hillitsemiseksi muuttaa rajusti koko sähkönjakelualaa. Hajautetun tuotannon, sähköautojen ja kulutusjouston lisääntyminen vaikeuttaa sähköverkon hallintaa. Monet jakeluverkkoyhtiöiden päivittäiset tehtävät riippuvat sähköverkon kuormitusten mallintamisesta ja ennustamisesta. Perustavanlaatuiset muutokset sähkön kulutustavoissa ovat tehneet kuormien mallintamisesta vaikeampaa kuin koskaan, ja monet kuormitusmalit ovatkin nykyään vanhentuneita. Monet jakeluverkon haltijat luottavat edelleen yksinkertaisiin huippukuormamalleihin ja vuosikymmeniä vanhaan mittausdataan perustuviin kuormitusprofiileihin kuormien mallinnuksessa ja ennustamisessa. Tämä kertoo uusien ratkaisujen tarpeesta.

Siirtymä kohti älykkäämpiä sähköverkkoja on toisaalta avannut uusia mahdollisuuksia kuormien mallinnukseen, etenkin kun perinteiset sähkömittarit on vaihdettu älymittareihin, jolloin tarkat kulutusmittaukset ovat jatkuvasti saatavilla koko verkosta. Mittarointidatan käyttämistä uusien kuormitus- ja ennustusmallien kehittämiseen on tutkittu ja tutkimustulokset ovat olleet lupaavia. Tässä diplomityössä hyödynnetään konstruktiivista tutkimustapaa uuden kuormitusmallin kehittämiseksi, jonka tavoite on arvioida ja ennustaa yksittäisten pienjänniteverkkojen asiakkaiden kuormituksia historiallisen datan perusteella. Työn päätavoite on selvittää, voidaanko tarpeeksi tarkkoja arvioita ja ennusteita muodostaa puhtaasti asiakaskohtaisen historiallisen kulutusdatan avulla.

MicroSCADA X on Hitachi Energyn tuoteperhe, joka muodostuu useista jakeluverkkojen hallintaan ja ohjaamiseen suunnitelluista ohjelmistoista ja laitteistoista. DMS600, joka on yksi tuoteperheen ohjelmistopaketeista, sisältää käytöntukijärjestelmän ja verkkotietojärjestelmän, joilla jakeluverkon haltijat voivat dokumentoida, valvoa ja hallita verkkojaan. Koska DMS600:n nykyiset kuormitusmallit ovat vanhentuneet, se tarjoaa hyvän alustan tässä työssä kehitetyn kuormitusmallin testaamiseen todellisessa ympäristössä.

Tässä työssä käydään ensin läpi työn viitekehyksen ymmärtämiseen vaadittava teoreettinen tausta. Tähän sisältyy sähkön mittausprosessin tarkastelu sekä nykyisten kuormitusmallien ja niiden käytännön käyttötarkoitusten esittely. Lisäksi esitellään DMS600-ohjelmistopaketti ja sen nykyisten kuormitusmallien heikkoudet. Teoreettisen taustan jälkeen käydään läpi ja perustellaan uuden kuormitusmallin suunnitteluprosessi. Algoritmi implementoidaan DMS600-järjestelmään ja uuden mallin tarkkuutta vertaillaan nykyiseen kuormitusprofiileihin perustuvaan malliin luomalla kattava otanta lyhyen aikavälin kuormitusennusteita useille pienjänniteverkoille. Testijoukko sisältää ennusteita kaikille vuodenajoille kolmessa erillisessä urbaanissa pienjänniteverkossa.

Tulokset osoittavat, että historiallisten mittausten suora käyttö yksittäisten asiakkaiden kuormitusten ennustamisessa voi tuottaa merkittäviä parannuksia ennustustarkkuuteen verrattuna perinteiseen kuormitusprofiileihin perustuvaan malliin. Lisäksi työ osoittaa, että tämäntyyppinen malli voidaan helposti integroida osaksi jakeluverkkojen hallintaohjelmistoja. Uusi malli osoitti kuitenkin taipumusta kuormien pieneen aliarviointiin. Lisäksi käytettävissä olevan ajan rajallisuuden vuoksi joitakin merkittäviä muuttujia ei voitu sisällyttää malliin. Nämä tekijät jättivät tilaa jatkokehitykselle, vaikka tulokset olivatkin todella lupaavia.

Avainsanat: sähkönjakelu, kuormitusten mallinnus, kuormitusten profilointi, lyhyen aikavälin kuormitusennustaminen, älykkäät sähköverkot, älymittarit

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

PREFACE

This Master's thesis was written for Hitachi Energy at the company's Tampere office in Hervanta. The work was examined by professor Pertti Järventausta, who I would like to thank for his feedback during the writing process and for providing excellent sources and above all, expert and friendly guidance. Also, from the company's side, I want to thank product manager Matti Kärenlampi for providing me with a very interesting subject and essential support during the process. I also want to express my gratitude to my colleagues at the Tampere office for great daily talks that contributed to this work.

This thesis would have been impossible without the measurement data provided by Koillis-Satakunnan Sähkö. Thanks for providing an extensive data set for my research.

Lastly, a special thank you to my partner, family, and friends for the continuous support during the entire process, especially during times when the work felt overwhelming.

Tampere, 15 December 2023

Jesse Christensen

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

AMI	Advanced Metering Infrastructure
AMR	Automatic Meter Reading
ANN	Artificial Neural Network
CIS	Customer Information System
DMS	Distribution Management System
DMS600 NE	DMS600 Network Editor
DMS600 WS	DMS600 WorkStation
DSSE	Distribution System State Estimation
DSV	Delimiter-separated Values
EV	Electric Vehicle
GPRS	General Packet Radio Service
LV	Low Voltage
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MV	Medium Voltage
NIS	Network Information System
P2P	Point to point
PLC	Power-line Communication
RE	Relative Error
RF	Radio Frequency
RMSE	Root Mean Squared Error
SAF	Standard ASCII Format
SCADA	Supervisory Control and Data Acquisition
SQL	Structured Query Language
TSO	Transmission System Operator

1. INTRODUCTION

During the past decade, electrical grids and electricity consumption habits have been undergoing arguably the greatest transformation since the widespread inception of electric power in the late 19th century. The recent changes stem from the ongoing energy transition from fossil fuels to renewable energy resources, as an effort to combat global warming. The transition takes place over numerous domains of the society, with the energy sector leading the way. Not only is electricity production switching from coal and oil to wind and solar, but electrical grids are also becoming smarter and more controllable, electric vehicles are taking over the market, and individual consumers are becoming small-scale producers of renewable energy.

While society becomes ever more dependent on electricity and the world population continues to grow, overall electricity consumption is on a steady rise [1]. Even more importantly, consumption habits are changing rapidly. As storing electricity in large scale is still not feasible, the intermittency of renewables is channeling consumption to times where electricity is available. This elasticity of demand is primarily controlled by the fluctuating prices of electricity. Along with major consumers, individual households are starting to adapt their consumption in accordance with electricity prices. Besides demand elasticity, electric cars and new heating solutions are rapidly changing electricity consumption patterns. These changes are making it ever more difficult to model and predict electrical loads in networks.

Successful management of distribution networks requires accurate modelling and forecasting of electrical loads across the grid. Load models are the foundation of *distribution system state estimation* (DSSE) and load forecasts, which are needed for network planning and operational planning, among other uses. In Finland, load estimation and forecasting has traditionally been based on customer class load profiles, where customers are divided into groups based on their type and all customers are assumed to conform to the group average. These load profiles are mostly based on old load research studies and as consumption patterns have changed, the profiles have largely turned obsolete and no longer accurately represent actual consumption. While recent changes have made predicting loads more challenging, advancements in technology have also offered new tools for monitoring what happens across the grid, such as automatic meter reading. Nowadays, practically all consumption locations are equipped with smart electricity meters that provide plenty of useful data for *distribution system operators* (DSOs), such as accurate load data, electricity quality measurements and information on outages. When it comes to load modelling, measurements from smart meters offer great potential in improving existing methods.

This master's thesis aims to study how the electrical loads of individual low voltage network customers could be estimated and forecast using smart meter data. While using smart meter measurements to improve load modelling has been studied in literature, using the historical consumption of individual metering points to create customer-specific estimations and forecasts has not been proposed. Using constructive research as a methodology, this work seeks to develop a straightforward customer-specific load modelling algorithm. To test the developed model, the algorithm will be implemented into MicroSCADA X DMS600, a software package designed for DSOs to manage their networks. Developed by Hitachi Energy (previously ABB Power Grids), the software consists of a *network information system* (NIS) and a *distribution management system* (DMS). The model will be tested in a real-life distribution network using actual smart meter measurements from a Finnish DSO, which will be imported into DMS600 database. The accuracy of the newly developed load model will be compared to existing load modelling methods. To provide a necessary theoretical framework, the thesis first examines how electrical loads are generally modeled as part of distribution network management, and how electricity consumption data is collected in the first place.

The conducted research does not aim to create a comprehensive solution to load modelling using smart meter data, but rather study whether estimating individual loads purely based on historical data could be used to improve the current load estimations used in management software. Due to the limited time frame and the technical difficulty of implementing a functional load model into network management software, the research focuses solely on modelling low voltage network loads and creating short-term forecasts, which are most useful in operational planning. The abovementioned restrictions also subject limitations to the developed load model, which will be addressed later in the thesis. Ultimately, this thesis aims to provide foundation for further research on the utilization of smart meter data in load modelling and intends to prove that even basic metering data analysis can provide improvements to existing load modelling methods.

2. BACKGROUND TO LOAD MODELLING

This chapter aims to provide necessary background knowledge to understand the framework of this thesis. In the beginning, the process of measuring electricity usage is explained, and recent evolution of electricity usage patterns is analysed. Subsequently, different load modelling methods are briefly introduced. The last subchapter will finally provide a practical view on how load profiles and other load modelling methods are used by distribution system operators in managing electrical grids.

2.1 Electricity metering

Measuring electricity usage is fundamental for the operation of distribution and electricity retailing companies. Traditionally, electricity usage measurements were only used for billing purposes. The early electricity meters were analogue, meaning that physically visiting each measurement location was necessary to record the reading. Manually reading thousands of meters was extremely labour-intensive and costly. The measurements were thus read very infrequently: as rarely as once per year. Customers had to be billed with an estimation of their consumption and the bills had to be corrected with each reading of actual measurement data. As the old meters only measured cumulative electricity usage, the operators couldn't easily get information on electric power quality or faults in low voltage networks. They also had no accurate information on the consumption patterns of individual customers. Eventually, the digital revolution and development of wireless communication technologies paved the way for the inception of remotely readable, smart electricity meters.

2.1.1 Automatic meter reading and smart metering

Automatic meter reading (AMR) refers to the technology that allows automated collection of consumption data from electricity meters. While often used synonymously in everyday language, a distinction can be made between AMR devices and smart meters. While AMR as a term only addresses remote reading of meters, smart meters are usually defined as being capable of two-way communication [2]. This means that smart meters not only send data but can also receive it. Modern smart meters can receive direct commands, allowing for example remote disconnection and reconnection, and remote activation of load control relays [3]. The current generation of advanced metering solutions has also given birth to a new concept, *advanced metering infrastructure* (AMI). AMI can be seen as the top-level definition for a complete advanced metering system, containing everything involved in the metering process such as all hardware and software [2], [4]. Two-way flow of information is generally seen as a prerequisite for advanced metering systems. As this thesis purely depends on measurement data received from electricity meters and practically all locations in Finland are nowadays equipped with smart meters, the term smart meter will be used here onward when referring to electricity meters and related measurement data.

The introduction of remotely readable meters completely revolutionized the electricity measurement process, eliminating the need for physical visits to premises. AMR systems started to become more and more common in the beginning of the 21st century and soon governments and regulatory bodies also started to recognize the immense benefits of these systems, such as reduced costs, improved customer service and the opportunities for enhancing grid management. In 2009, the European Union issued a directive for member states to deploy smart meters under EU energy market legislation. In practice, the target was to provide 80% of end consumers with smart electricity meters by 2020 in countries where the effects of the roll-out were deemed positive in a cost-benefit analysis. [5] As a result, the replacement of traditional meters with new smart meters began very quickly. By the end of 2014, nearly all electricity usage points in Finland were already equipped with smart meters.

2.1.2 Power balance and measurement intervals

At all times, the production and consumption of electricity must be equal. In Nordic countries, this power balance is traditionally settled hourly, which means that for each hour, electricity market operators attempt to balance their production and consumption in advance. However, the actual outcome will always have deviations to the plans and as a result, there is a separate balancing power market for acquiring or selling the necessary power needed to balance the actual production and consumption. In the Nordic power system, the *transmission system operator* (TSO) of each country is responsible for maintaining the nationwide electrical balance. The power balance at each time can easily be monitored by measuring the frequency of the electricity grid. When there is more production than consumption in the grid, the frequency rises over the nominal frequency of 50.0 Hz. Correspondingly, the frequency drops if there is more consumption than production. [6]

After electricity delivery, market operators are required to submit the actual production or consumption data to the TSO for balance analysis. In Finland, the data is submitted to a system called *Datahub*. Launched by the Finnish TSO Fingrid in February 2022, Datahub presents a centralized information exchange system for all electricity market operators. The system offers an equal and secure platform for exchanging data between all electricity market operators. [7], [8] For DSOs, this means providing actual measurements from electricity meters. Since the power balance in the grid is traditionally observed hourly, electricity usage is usually recorded in an hourly interval as well. Many modern smart meters already record usage on shorter time intervals, such as 5 or 15 minutes. In this case, the time series are combined into an hourly series in the control system. However, as intermittent production such as wind and solar power becomes more widespread, electricity production becomes more prone to rapid fluctuations. As a result, all of Europe is now transitioning towards a more accurate, quarter-hourly imbalance settlement period. The new 15-minute settlement period will be rolled out in phases. The initial introduction of the new settlement period took place in May 2023 in Finland, while the day-ahead market is scheduled to switch to a 15-minute time interval in 2025. In the process, all smart meters will also be reconfigured to measure electricity usage in 15-minute windows. [6], [9]

2.1.3 Measurements and communication

Typically, smart meters are configured to measure energy flow and certain power quality factors. The flow of energy is measured bidirectionally: energy inserted from customer to the grid is recorded along with energy taken from the grid. In terms of power quality, typical measurements are voltage, current and frequency, as well as active and reactive power. In addition, smart meters record some predetermined events. These events can for example be outages of different duration, significant voltage drops and surges, or zero faults. The meters store all information until it is read by the control system. [3], [10]

Multiple different technologies and configurations can be used for communication between remotely read meters and the central database. The most cost-efficient and reliable solution for each case depends on many factors, such as geographical location and the coverage of applicable networks. The most frequently used communication technologies are *PLC* (Power-Line Communication), *cellular networks*, and *RF* (radio frequency) [2], [3], [11].

PLC is a wired communication method where data is carried on a conductor using a standardized frequency area. In other words, the data is transferred in the physical electrical grid. The main advantage of PLC is the cost-efficiency of utilizing existing infrastructure. The downside of PLC is that the connection to the meter breaks in case of an outage. The connection also breaks if the meter is installed behind the customer’s main switch and the switch is turned off. For this reason, new meters are usually installed before the main switch so that the meter can stay online even if the switch is turned off.

Cellular networks are the primary technology used for wireless metering data transfer. Data transfer over mobile networks allows reliable communication over long distances and mobile networks typically have excellent coverage. GPRS is the prevalent technology in current implementations. It represents a second generation (2G) service for transferring data in the GSM network. While GPRS is currently by far the most used wireless metering data communication method, the newer generation technologies, mainly 4G and 5G, are expected to become more common with the development of smart grids. [2], [11]

Smart meters can communicate with the central remote reading system either directly or indirectly through a concentrator hub. Direct communication is commonly referred to as P2P (point to point) and is mostly used in sparsely populated areas. In densely populated areas, a common configuration is to first transfer measurement data from a group of meters to concentrator hubs, and then wirelessly transfer the data from the hub to the remote reading system. [2], [11] An illustration of a widely used PLC/GPRS communication setup is presented figure 1.

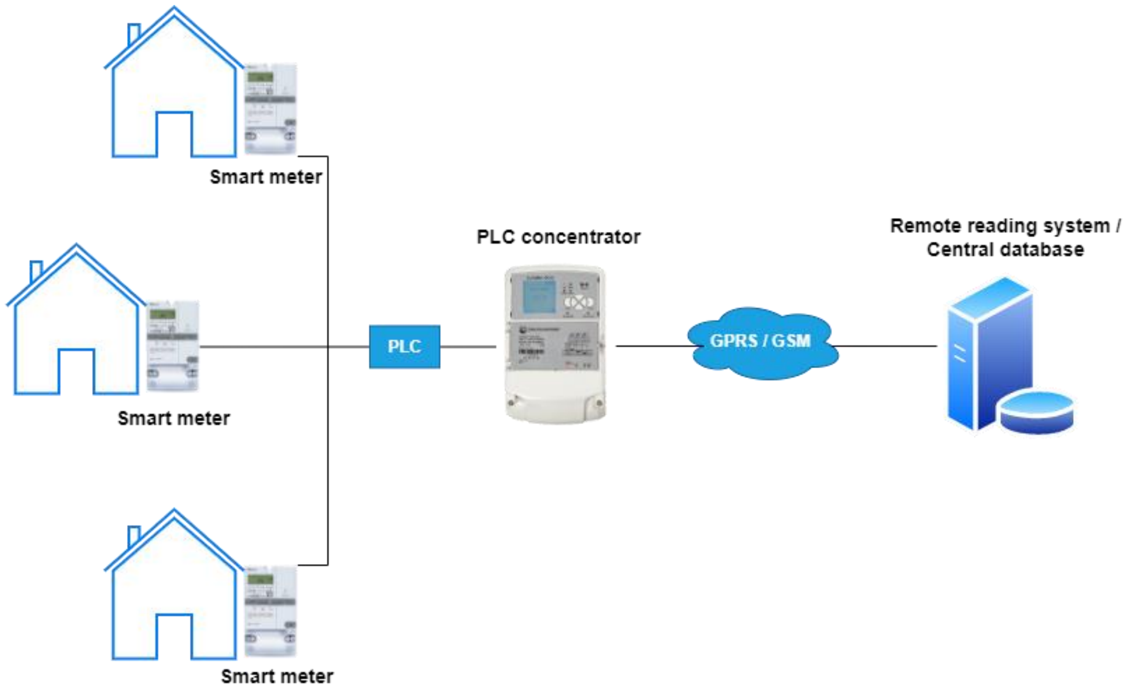

Figure 1: *Smart meter communication using a combination of PLC and GPRS*

In the system depicted by figure 1, PLC is first used to transfer measurement data from a set of consumers to concentrator hubs, often located in distribution transformers. GPRS is then used to further transfer the data wirelessly to the central remote reading system.

2.2 Recent changes in electricity consumption patterns

Electricity consumption habits have changed drastically during the past decades. This transformation has taken place across every sector, producing lots of uncertainty and complexity in terms of load forecasting, or modelling electrical consumption patterns. These changes are attributable to numerous phenomena, such as the evolution of heating solutions, electric vehicles, distributed generation, electricity price volatility and shifts in different domains of industry. As the state of distribution networks becomes dependent on an increasing amount of variables and unpredictability, system operators are forced to develop their practices in actively controlling the grid. [12] This subchapter aims to offer a brief overview on why and how the consumption patterns have recently undergone a vast shift, mostly focusing on changes in Finnish consumption habits, although most trends are universal and thus taking place all over the world.

Heating is globally the largest end-use of energy with around a 50% share of global final energy consumption [13]. While the majority of global heating energy demand is still covered with fossil fuels, most of the heating energy in Finland originates from renewable resources. In Finland, district heating is the largest source of heat energy for residential and service buildings. While district heating traditionally depends on fossil fuels, most of the heat is nowadays obtained from biomass. Direct electricity heating follows district heating as the second largest source of residential heating in Finland. [14] The heating industry has lately been revolutionized by the surge of heat pumps as a more efficient alternative to heating. While heat pumps effectively utilize electricity to transfer heat, their efficiency is often much higher than the one of direct electricity heating. Measured by *coefficient of performance* (COP), heat pumps are usually able to transfer more thermal energy than the work it requires. For example, with a COP of 4, a heat pump could transfer 4 kWh of thermal energy by using only 1 kWh of electrical energy. Thanks to their relatively low investment cost and ability to be used also for cooling purposes, air-source heat pumps are currently a very common addition to heating solutions in residential and service buildings.

In Finland, the total number of air-source heat pumps has steadily increased from virtually zero to over 500 thousand units since the beginning of the millennium [15]. Although the growth has started to level in Finland, global heat pump sales are still climbing.

Figure 2 shows the annual growth rate of air-source heat pump sales globally and in Europe in 2021 and 2022.

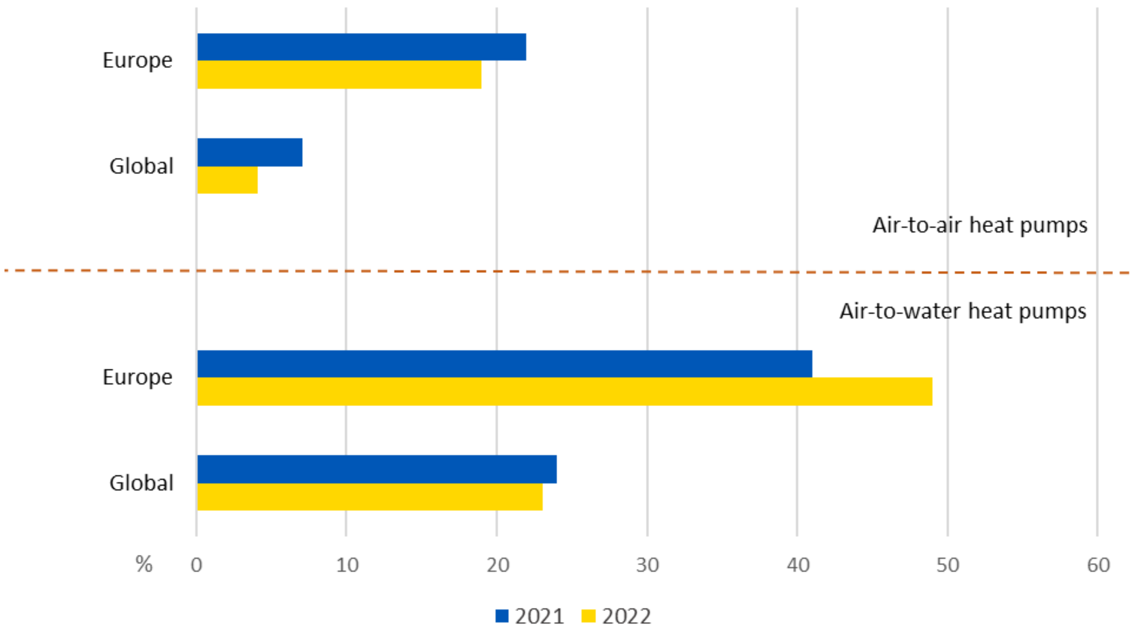

Figure 2: *Air-source heat pump sales growth rates in 2021 and 2022. Modified from [16]*

Due to their limited capacity, however, air-source heat pumps are often used alongside traditional direct electricity heating. Despite having a larger investment cost, ground-source heat pumps have also surged as a stand-alone solution for residential heating, already surpassing direct electricity in new residential buildings [14]. Due to the increased efficiency, the use of heat pumps can decrease the amount of electricity needed for heating, while the use for cooling might lead to increased electricity usage during hot periods. However, the efficiency of air-source heat pumps decreases significantly when outside temperature drops, which limits the benefit of heat pumps systems during cold periods [17]. One study even reported higher electricity consumption in households with heat pumps compared to households with direct electricity heating during the coldest periods of the year [18]. Nevertheless, heat pumps generally provide an improvement to heat transfer efficiency and lower overall electricity consumption. Heat pumps are also easy and rapid to control, providing the possibility for consumers to optimize their usage to times of low electricity prices.

One of the largest forces to alter electricity consumption patterns is the proliferation of electric vehicles (EVs) and the related infrastructure. The electrification of transportation affects the electricity consumption patterns in many ways across the entire grid. Firstly, EVs have very powerful batteries and thus require a lot of electricity to charge. This leads to increased overall electricity consumption. Secondly, to enable tolerable charging times, especially public charging stations need a very high charging current. For example, a single Tesla Supercharger can have an electric power supply of up to 250 kW [19], so it's clear that a widespread charging infrastructure immensely increases the stress on the distribution networks. On the other hand, EVs could potentially be used to push electricity back to the grid from the battery at times of excessive demand. This technology is referred to as vehicle-to-grid (V2G). While the technology is still at a very early stage, it could potentially be used to widely balance the grid in the future. Even before a breakthrough in V2G, smart charging alone provides a degree of demand flexibility, affecting the load profile of EV households. In smart charging, the charging process is controlled to balance the load demand curve. [20]

EVs also cause a significant time shift in electricity usage. Most EV owners charge their vehicles at home overnight, which greatly increases electricity consumption during nighttime. The advantage of added nightly consumption is that it can increase the balance between electricity production and usage as night hours are generally a time of low usage and lower electricity prices. However, if many EVs are charged simultaneously in the same area, local congestion problems, such as unacceptable voltage deviations, could still occur. Conversely, the use of public speed charging stations might lead to an increase in electricity demand during peak hours, which could again lead to congestion problems. These problems could be mitigated with smart charging. [20]

Similarly testing the strength of distribution networks, *distributed energy resources* (DER), especially wind and solar power, are transforming electricity production from centralized, large power plants to small-scale generation scattered all around the grid. As intermittent energy sources depending on weather conditions, distributed generation complicates maintaining the balance between load and production and introduces many power quality risks into the grid. In addition, the addition of wind and solar power decreases the amount of inertia in the electrical grid. Traditional power plants produce inertia through the kinetic energy stored in the massive generators. Inertia keeps the grid frequency more stable when load or production changes, adding stability to the power system. Wind and solar power do not produce inertia to the grid, thus decreasing the amount of inertia. [21] While the abovementioned issues complicate distribution network operation, the phenomenon that alters consumer electricity consumption patterns is the growing number of private *photovoltaic* (PV) systems. As PV technology has advanced, the investment cost of a small system has reduced to a point where acquiring a system is feasible for individual customers. Along with the increased solar cell efficiency, even a rather small system can produce a useful amount of power to supplement the ordinary grid connection, or even replace it in secondary residences. Naturally, the addition of a PV system changes the consumption pattern of a customer significantly. Not only does the customer take less electricity from the grid, but the PV production is likely to intermittently exceed the personal load which in practice means that the customer will insert electricity into the distribution network. The intermittent production also complicates modelling the loads of customers with PV systems.

Mainly due to the intermittency of wind and solar power, the electricity prices in Europe have seen a steep volatility increase over the past few years. In 2022, already 23,5% of EU’s net electricity generation were covered by wind and solar, which means that a large portion of the total capacity relies on the weather conditions [22]. At times of excellent wind conditions, the electricity prices inside the Nord Pool exchange have been extremely low, and vice versa. For example, the system price of the Nord Pool day-ahead market has recently gone negative on multiple occasions [23]. This price volatility has further been emphasized by Europe’s challenging geopolitical situation which has directly affected typically stable resources, such as natural gas [22]. The fluctuation of prices naturally channels consumption into times when the electricity prices are low. In Finland, the electricity prices of most individual consumers have traditionally been tied to a fixed-price contract, but contracts tied to the Nord Pool day-ahead prices have become more and more common [24]. This goes in tandem with the availability and ease of implementing home automation: as technology advances, it has become easier for consumers to concentrate their loads to times of low electricity prices. From a power balance perspective, the responsiveness of consumer consumption to market prices is beneficial, as it acts as a form of demand response. However, if the loads of individual customers are no longer tied to personal habits but fluctuating market prices, estimating and forecasting the loads becomes extremely difficult.

The transforming consumption patterns are also attributable to various other changes. These include for example:

- Shifts in different industries, such as agriculture, retail, transportation, and construction: In developed countries, the number of farms has plummeted due to the increased efficiency requirements for sustaining a profitable business, favouring larger farm sizes with advanced mechanization. Retail is shifting from brick-and-mortar stores to e-commerce. Transportation is ongoing an increase in electrification, also in public transport. Modern buildings are designed for maximal energy efficiency, focusing on insulation and minimizing all energy waste. Similar trends are visible in nearly all industries, each contributing to the shift in consumption patterns.

- Technological advancements: Appliances have become more energy-efficient and LED lighting has become predominant, which reduce electricity consumption. At the same time, many mechanical appliances and machines have turned electrical, which in turn increases consumption.
- Political decisions and greener policies: As an effort to combat global warming and increase sustainability, governments and political entities have introduced a wide variety of regulations to incite energy efficiency at both consumer and business levels. These regulations have a second-hand impact on nearly all sectors of the society.

The constantly evolving electricity consumption highlights the need for flexible and dynamic load models that no longer rely on conventional tendencies.

2.3 Overview on load modelling

In electricity distribution, load modelling often refers to the mathematical representation of the network parameter relationships (such as power, voltage, and frequency) [25]. However, in the context of this thesis, the goal of load modelling is to help estimate the current network loading state, as well as to forecast network loads. The need to model current loads arises from the fact that the exact state of each point in a distribution network is never known with certainty. Receiving exact loads in real-time from every electricity meter is not currently feasible, and measurement points are mostly present in major network nodes, such as substation feeders. Load models are used to gain the best possible estimate of the network loads using the available information and tools. Often a combination of real-time measurements, historical data analysis and forecasting techniques is needed to achieve the most accurate possible system state estimation. Forecasting future loads is a somewhat similar task, as is also requires equivalent load models. Depending on the application, models are expected to answer different questions: often the interest lies in peak loads, but sometimes minimum loads or average power over a certain period are also needed. In this subchapter, some of the most common traditional and state-of-the-art load modelling methods and tools will be presented.

2.3.1 Velander's formula

Velander's formula is a traditional and simple method for estimating peak loads using annual energy and predetermined, customer group specific coefficients. It is best suited for estimating the peak load of a large group of consumers because in practice, the loads of individual customers do not conform to the assumptions of Velander's formula. The peak load of a customer group is calculated with the formula

$$P_{max} = k_1 \cdot W + k_2 \cdot \sqrt{W}, \tag{2.1}$$

where P_{max} is the group's peak demand in kilowatts (kW), k_1 and k_2 are the Velander coefficients for the group, and W is the group's annual consumption in megawatt hours (MWh). The Velander coefficients are determined with practical experience and measurements to best reflect actual loads. Different customer groups can be assigned with separate coefficients to factor in consumption habit differences. [26], [27]

When different customer groups are assigned with separate coefficients, one line section might feed customers with varying Velander's factors, whereupon Velander's formula cannot be used directly to calculate the section peak power. The peak power of such a line section can be calculated with the equation

$$P_{max} = \sum_{i=1}^{n} k_{1i} W_i + \sqrt{\sum_{i=1}^{n} k_{2i} W_i} \tag{2.2}$$

where P_{max} is the section's peak load, k_{1i} and k_{2i} are each group's Velander coefficients and W_i is the annual consumption of each group.

Since Velander's formula only estimates the peak power and doesn't give information about how the load varies in quantity and over time, it is most useful for network planning, where each network component needs to be chosen so that they can withstand the maximum load that could occur, with some additional buffer.

2.3.2 Load profiling

In general, a *load profile* depicts electrical loads in discrete time intervals. A load profile can portray any source of consumption with any time interval, for example the average daily electric load of Finland could be shown for a period of one year. Load profiles are typically displayed as graphs with the electrical load on the vertical axis and time on the horizontal axis. Distribution system operators are often interested in knowing hourly loads for either individual customers or, for simplicity's sake, a group of similar customers. Consequently, the term load profiling usually refers to a statistical method where customers with similar consumption patterns are categorized into groups, often referred to as customer classes, and a general representation of the group's load curve is established.

An example of a graphical *load curve*, representing hourly electrical loads over a period of one week, is shown in figure 3.

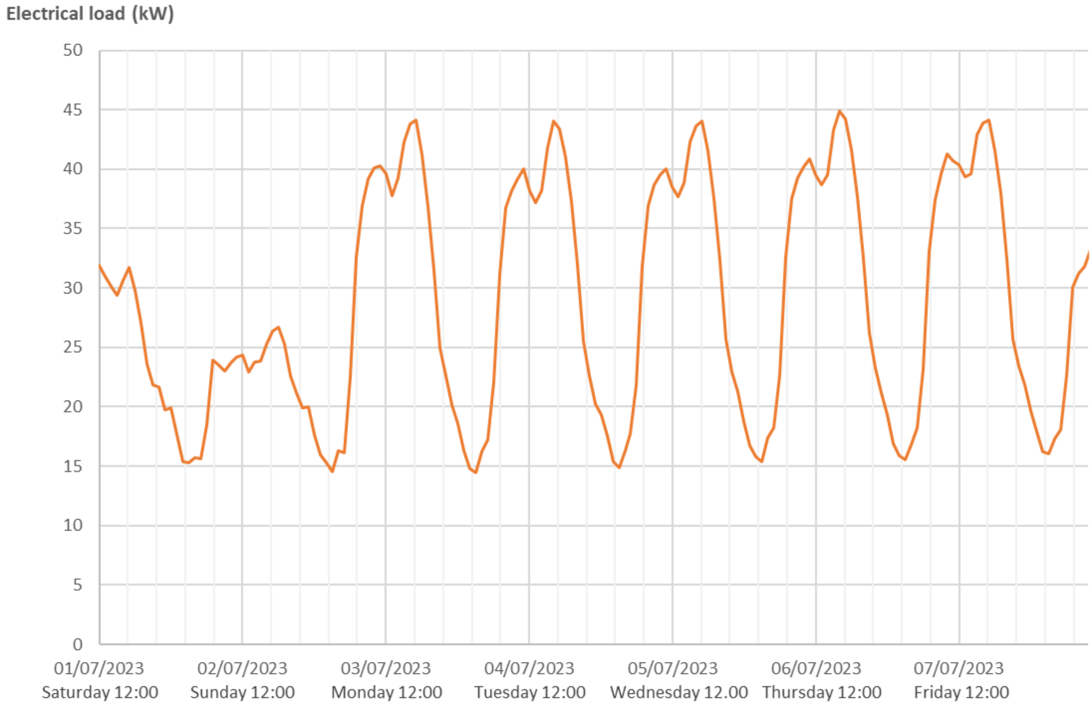

Figure 3: *Example of a graphical load curve. Data retrieved from DMS600.*

Because the example load curve is based on actual consumption data from a group of customers, some qualities typical to electricity consumption are clearly visible: It is easy to see that the consumption is very different during the weekend and weekdays. Meanwhile, the consumption pattern is extremely similar between weekdays. Also, a characteristic load peak is visible on weekday mornings and evenings. In Finland, grouping customers with similar consumption patterns and analysing the groups' behaviour is a well-established means of load modelling. Next, the history and current state of customer classification and customer class load profiles in Finland will be briefly presented.

Before the introduction of automatic meter reading, load research was very laborious as it required collecting data manually from individual customers. The first nationwide load research project in Finland began in 1983, organized by the *Finnish Association of Electricity Supply* (currently named *Finnish Electricity Association*, operating under *Finnish Energy*). A total of 42 electricity utilities collaborated in this research, in which the hourly consumption of almost 1200 customers were collected. The research included two measurement periods that took place during the 1980's. Based on their consumption habits, customers were divided into multiple classes based on their purpose, such as housing, industry, service, and agriculture. Each category was further divided into subclasses: for example, housing was divided into different subclasses based on building type and heating solution. The objective of the classification was to divide consumers into groups where electricity usage is similar enough that it depicts the consumption of an individual customer with decent accuracy. After extensive data analysis, 46 unique customer class load profiles were eventually published in 1992. The load profiles include expected hourly loads, load deviations and temperature dependencies for each customer class. [26]

When the measurements were analyzed, the average power for each two-week period of the year, as well as relative two-week indices, were calculated for each customer class. In other words, the year was split into 26 two-week periods. The relative two-week index gives the percentage difference between the average power of the two-week period compared to the average power of the entire year. For example, if the two-week index is 125, the average power of this two-week period is 25% higher than the average power of the whole year. In addition to the two-week periods, hourly indices were calculated for each day of the week, separately for summer and winter. Since the consumption on different weekdays is very similar, all weekdays are assumed to be identical in order to limit the data that needs to be processed. Saturdays (or eves) and Sundays (or public holidays) are analyzed separately. [26]

The correlation between electrical load and outside temperature is called *temperature dependency*. The dependency was factored in the load profiles using the following formula:

$$q_a(t) = q_0(t) + \beta \cdot \Delta T(t) \tag{2.3}$$

where $q_a(t)$ is the actual measured electric power at time t, $q_0(t)$ is the electric power at normal temperature at time t, β is the temperature dependency coefficient and $\Delta T(t)$ is the temperature difference between measured and normal temperatures at time t [26].

The customer class load profiles published in 1992 remain the only publicly available extensive set of load profiles. As a result, they are still widely used by distribution system operators. Afterwards, multiple load research projects have been conducted for individual operators or consortiums. These projects have resulted in improved and updated load profiles, but their use is limited to the participants. Many companies have also created individual load profiles for major customers with smart meter data. While the 1992 load profiles have been extremely useful for distribution companies and electricity market operators, they have become obsolete because the electricity consumption habits have changed significantly over the past 30 years, as examined in the previous subchapter.

The concept of load profiling is not strictly limited to Finland. Other countries have conducted similar load research studies in order to create equivalent load curves that represent typical electricity consumption of certain customer groups. For example, in 1991, the *Swedish Association of Electric Utilities* (*Svenska Elverksföreningen*) published a report which contained load curves for a large set of customer classes. The load curves were obtained after analysing metering data collected from electricity end users. [28]

As another example, load profiles have long been used for facilitating imbalance settlements in the United Kingdom. Traditional electricity meters were unable to measure electricity consumption in accordance with the half-hour imbalance settlement period used in the UK. To avoid having to install half-hourly electricity meters to all locations, eight basic load profiles were created which would be used along with readings from existing electricity meters to create half-hourly settlements. This method was applied to all customers below 100kW maximum demand. Today, the load profile assisted imbalance settlements are still used in the UK, although for an increasingly smaller group of customers. A transition is underway to measure electricity consumption in accordance with the half-hourly period in all locations. [29]

2.3.3 Clustering

Clustering is a general data analysis technique which aims to divide data points into groups based on similarities. A group where each object is similar to each other is called a *cluster*. As clustering is a high-level concept used in many fields of science for various data types, hundreds of different clustering algorithms have been developed. Clustering can be used for numerous applications, such as customer segmentation, trend detection, data summarization and network analysis. Certain algorithms are best suited for certain types of applications, although finding and selecting the perfect algorithm for a specific purpose is often a near-impossible task. Due to the vast amount of distinct algorithms, they are often categorized by their main notion. For example, based on the used technique, clustering algorithms could be classified to *probabilistic techniques*, *density-based techniques*, *distance-based techniques* or *hierarchical techniques*. [30]

In the context of electricity distribution, clustering can be used for grouping customers based on their consumption patterns. Instead of relying on predetermined customer classes, different clustering algorithms are used to identify customers that share similar load behavior. The main advantage of clustering compared to traditional customer class profiling is that customers can be classified purely based on their load patterns rather than being automatically grouped by their basic type [31]. When new load profiles are created for each cluster, the new profiles can directly replace old customer class load profiles.

Perhaps the most comprehensive literature review on the use of clustering for electricity consumer classification was conducted by DSc Antti Mutanen in his dissertation [32], where he also developed a novel method for creating and improving load profiles based on metering data from electricity meters by utilizing clustering algorithms. In his review, Mutanen noted that multiple different algorithms have been proposed and studied for segmenting electricity consumers, but no consensus has been achieved on which method is best suited for the task, although the *k*-means algorithm produced good results in multiple studies. Eventually, Mutanen ended up using a two-stage weighed *k*-means algorithm for the new method. Different algorithms were also studied, but the weighed *k*-means came out as the best overall solution. As the developed clustering method produced very promising results in load profiling, the key points of the model are shortly presented which also gives a solid example on how clustering can be used as a tool in load modelling.

k-means is a distance- and centroid-based clustering method. In this type of a method, each cluster is represented by a *centroid*, in other words a central vector which can for example correspond to the mean of all the cluster data points. The similarity of data points is determined by their distance, most often the Euclidian distance, from the centroids. In weighted k-means, each pattern vector is weighted. In this case of electricity customer classification, the vectors were weighted with the corresponding annual energies. Thanks to its simplicity and large experimental success, *k*-means is one of the most widely used clustering methods. One of the weaknesses of *k*-means is that the number of clusters needs to be determined beforehand, which is often difficult. The centroids must also be initialized and in practice, the initialization affects the final result accuracy. [30], [32]

Through extensive iterative and practical analysis to mitigate possible sources of inaccuracy, Mutanen proposed the following simplified workflow for creating new cluster load profiles from metering data:

> First, metering data is read, pre-processed, and validated. Then, temperature normalization is performed to the data before calculating next year energy forecasts. From the temperature normalized time series data, the pattern vectors can be calculated, which describe the average hourly loads of each hour of the year. Next, before the first weighted *k*-means clustering stage, the largest customers are separated from the data as they shall be calculated an individual load profile. After the first clustering stage, clear outliers are filtered and selected for individual load profiling. Then, a second weighted *k*-means clustering stage will be performed. After outliers are classified to their closest clusters, new load profiles can be formed from each cluster centroid. Once temperature dependency parameters and standard deviations are calculated for all load profiles, the method is completed. [32]

While it was found that the optimal number of clusters cannot be unquestionably determined beforehand, the above method produced very good results in improving the accuracy of load profiling. It was also shown that the developed model can easily be implemented into existing distribution system management software. [32]

The model developed by Mutanen acts as a great example of how clustering can be used as a tool to create improved load profiles. As pointed out, clustering is an extremely widespread concept, and many different approaches could be taken on the use of clustering for classifying electricity customers.

2.3.4 Artificial neural networks

Artificial neural networks (ANNs) are a branch of machine learning that have quite recently risen to everyone's attention due to the groundbreaking technological advancements they have produced. From revolutionizing image recognition and machine translations to creating remarkably advanced artificial intelligence models, neural networks are the backbone of many applications exploited in our everyday lives. Neural networks are inspired by the structure of brains: a neural network is essentially a graph, where nodes, called artificial neurons, are connected to other neurons. Neurons can transmit data to the neurons it is connected to. The signals are actual numbers, and some neurons can process the data between signals. Neural networks can be trained with a large set of sample input data. During the training phase, the difference between the desired output and the actual network output is minimized. The use of neural networks for electricity customer classification as well as creating load forecasts has been studied. Customer classification with the help of neural networks is essentially a form of complex clustering. Perhaps a larger benefit could be achieved in load forecasting since forecasting electrical loads is notoriously difficult due to the stochasticity of electricity consumption, and because neural networks are a rather well-established tool for nonlinear regression. How the general neural network concept is turned into a functional model capable of solving complex tasks is complicated and not relevant for this thesis. Instead, the key points of two studies where neural networks are utilized in load forecasting are shortly introduced as examples.

In [33], an ANN model was developed for creating 24 hour forecasts for MV/LV substations loads. Load measurements with a 30-minute interval were collected from two substations in France for a total of 540 days between years 2009 and 2011. Two separate models were developed: one that forecasts the daily average load, while the other forecasts the intraday loads in intervals of 30 minutes. Measurements from the first one-year period were used in training the models and the remaining data was used as a test set.

Different input variables were given to the two models. The average power model used only load and temperature variables, while the intraday model used cycle variables and day type variables in addition to the load and temperature variables. To assess the forecast accuracy, the *mean absolute errors* (MAE) and *mean absolute percentage errors* (MAPE) were calculated and compared between the neural network model and two simple reference models. The reference models used were a *naive model* and a *time series model*. The naive model used real historical consumption from a similar day as a direct forecast, while the time series model was a regression model that combined a day type variable with a temperature dependency model and a Fourier component periodic model. The results showed that the neural network model produced more accurate forecasts than the reference models, with MAPE being 12.9%, 11.0% and 10.3% for the naive model, time series model and neural network model, respectively.

Another similar but more recent study was conducted in [34]. This study proposed a more sophisticated *deep learning* model for short-term load forecasting, combining two neural network concepts, *convolutional neural network* (CNN) and *long short-term memory* (LSTM), into one model. Hourly measurement data collected from Italy from a period of three years (2015-2017) was used in training and testing the model. Data from the first two years was used in training, leaving the data from the last year for testing purposes. As input variables, load data from the past 21 days was used for producing a forecast for the following 24 hours. The forecast accuracy was compared against well-known machine learning algorithms, namely *random forest* (RF), *decision tree* (DT) and *DeepEnergy* (DE). Across 8 separate test periods, the average MAPE was 3.96%, 4.37%, 4.79% and 5.78% for the new model, DeepEnergy, DE and RF, respectively. Therefore the new, more sophisticated model produced an improvement in load forecasting accuracy when compared to popular machine learning algorithms.

Although neural networks are considerably more complex than traditional load forecasting approaches, they arguably provide better accuracy in load prediction. However, their very complexity makes neural network models difficult to implement in actual power grid management systems. For this reason, distribution system operation still relies on more traditional load forecasting methods.

2.4 Load models in distribution system operation

Load models have a wide range of use cases and applications in distribution network operation and related businesses. Earlier in this chapter, different techniques used to model, estimate, and forecast electrical loads were presented. Next, the practical applications of these methods are examined.

One of the key purposes of load modelling is distribution system state estimation (DSSE), often performed by the DMS. As mentioned earlier, the exact state of each node in the network is not known because of the limited number and accuracy of measurements. Like the name suggests, the objective of DSSE is to estimate the most likely present state of the network. DSSE typically utilizes two types of data in determining the system state: actual real-time measurements and *pseudo-measurements*. Real-time measurements are typically received through SCADA and originate from primary substations and occasional measurement points across the grid. The measurements are typically either voltage and current magnitudes, or active and reactive power pairs. Due to the limited number of actual measurements, the state estimator requires additional data to fill the gaps. This is achieved with pseudo-measurements, which are not actual measurements but are used in state estimation like real-time ones. [12], [35] In a load profile based model, the load estimations derived from the profiles can be directly used as pseudo-measurements in DSSE. The accuracy of DSSE can be further improved by verifying that the sum of pseudo-measurements from a feeder match the real-time measurement from the same feeder, and correcting the pseudo-measurement values if the total load does not match. Different algorithms can be used for the state correction. [12], [36]

Load modelling methods are also very useful for the generation of short-term load forecasts and their subsequent use in operational planning. In distribution network operation, short-term forecasting typically means forecasting loads for up to one week into the future [34]. The forecasts can be used for planning maintenance outages and congestion planning, for example. To minimize the disturbance experienced by network customers, distribution system operators naturally want to schedule maintenance outages to times of minimal consumption. Short-term forecasts can be used to detect risks of congestion, i.e. times when the grid capacity isn’t sufficient to distribute power to all locations. Network overloads could cause violations in power quality, so it’s important to identify congestion risks in case load shedding or curtailment is necessary. [37] Since somewhat reliable weather forecasts are available for one week, they can be used to improve the accuracy of short-term load forecasts.

In addition to short-term forecasts, DSO’s sometimes need longer-term forecasts for different purposes. This refers to forecasting loads over 1 week into the future, which is analogously based on load modelling. Most commonly, long-term forecasts are needed for network planning. Loading scenarios throughout the entire year are needed when the network structure changes to ensure that the network is adequately dimensioned. Also considering load or generation changes that could possibly take place during the foreseeable life cycle of the network components is important. Overestimating the loads results in economically wasteful investments, while underestimating the loads potentially leads to challenges in power supply. [38] For this reason, the accuracy of the forecasts is of major concern to distribution system operators. Traditionally, the interest is in the network peak loads, as they set the minimum capacity of each network component. However, with the increasing number of distributed generation, DSOs are nowadays more and more interested in times of lowest loads, as low loads coupled with distributed generation could lead to power quality issues, namely overvoltage, in the network [39].

In addition to system state estimation and load forecasting, load models and especially customer type load profiles also serve more mundane purposes. Firstly, customer class load profiles are the foundation of tariff design, for both DSOs and electricity retailers. This is an extremely important function; not only must the fees be fair and equal for a public commodity, but tariff design can also contribute to grid control, for example by factoring peak loads and reactive power usage in billing. Secondly, when new electricity connections are added to the distribution network, the operator must estimate the customer's annual energy consumption since historical data does not yet exist. Load profiles can also be used by businesses to identify potential customers for load optimization solutions, such as peak load management or reactive power compensation services. Overall, load modelling has a large variety of use cases, some of which require classifying customers based on their load patterns, while some might benefit from individual load estimation.

3. MICROSCADA X PRODUCT FAMILY

MicroSCADA X is a product family developed by Hitachi Energy, designed to be a complete solution for electrical network automation and control. The product portfolio includes multiple software solutions as well as supporting hardware. In this chapter, the software relevant for this work, DMS600 and SYS600, are briefly introduced. These applications are often used hand in hand to manage the entire network. Since this thesis includes implementing a new feature into DMS600, this software package is presented in more detail and only a short introduction is given to SYS600. In the DMS600 introduction, the current implementation for load modelling in DMS600 is presented and its weaknesses are analysed.

MicroSCADA X SYS600 is a *supervisory control and data acquisition* (SCADA) software used for real-time high-level monitoring and control over an electrical network. DMS600 on the other hand is a distribution management software package consisting of two separate programs: *DMS600 Network Editor* (NE) and *DMS600 Workstation* (WS). DMS600 NE is a network information system (NIS) used primarily for modelling and documenting the electrical network, while DMS600 WS is a distribution management system (DMS), which in turn is used for monitoring and controlling the state of the network. [40] The basic structure and communication between the applications is presented in figure 4:

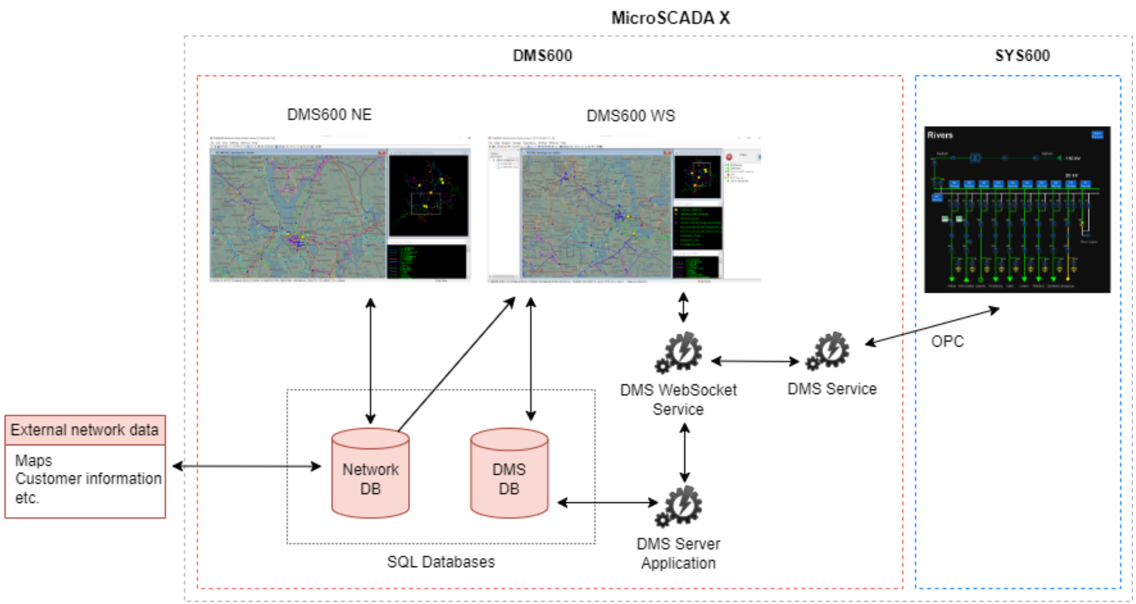

Figure 4: *Communication and basic structure of a DMS/SYS setup*

DMS600 applications are designed to run on Microsoft Windows or Microsoft Server operating systems. Data is stored into two separate SQL databases. As the name implies, the network database is used for storing all information on the network components, whereas the DMS database contains primarily operational information, such as data on outages and switching states. DMS600 utilises Windows Service based background add-ins for both internal and external communication. In this figure, internal communication means sharing data between different DMS600 instances and modules, primarily handled by DMS WebSocket Service, which uses a secure WebSocket protocol. [40] External communication refers to communication between DMS600 and several external interfaces. Most importantly, DMS600 communicates with SYS600, which is implemented with an OPC connection between DMS Service and SYS600. DMS Service includes a specific OPC communication module called ExternalOPCDAClient, which connects to the OPC server on the side of SYS600. This allows real-time data transfer between the two systems. In addition to SYS600, external interfaces can include for example a trouble call centre, an AMI system, or a field crew tracking interface. While the MicroSCADA X applications are designed to run in tandem, they can also be used separately in joint with systems from other suppliers. For example, multiple Finnish customers operate a NIS developed by TietoEVRY along with the DMS600 Workstation. Similarly, different SCADA applications can be used with the DMS600 package using the previously mentioned OPC data access. Different modules have been designed for data import from external applications to facilitate integration between these systems. [40], [41]

3.1 MicroSCADA X SYS600

As mentioned earlier, SYS600 is the SCADA system of MicroSCADA X family. By far the most widespread application of the product family, SYS600 is used globally by thousands of customers. SYS600 includes many functionalities for monitoring, controlling and analysing the status of distribution networks. The main view is the process display, which shows substation diagrams and connections between substations. The display can be used to operate remotely controlled switching devices in real-time. In addition to the network topology, the display can also show alarms, faults and important measurements.

The process display is presented in figure 5.

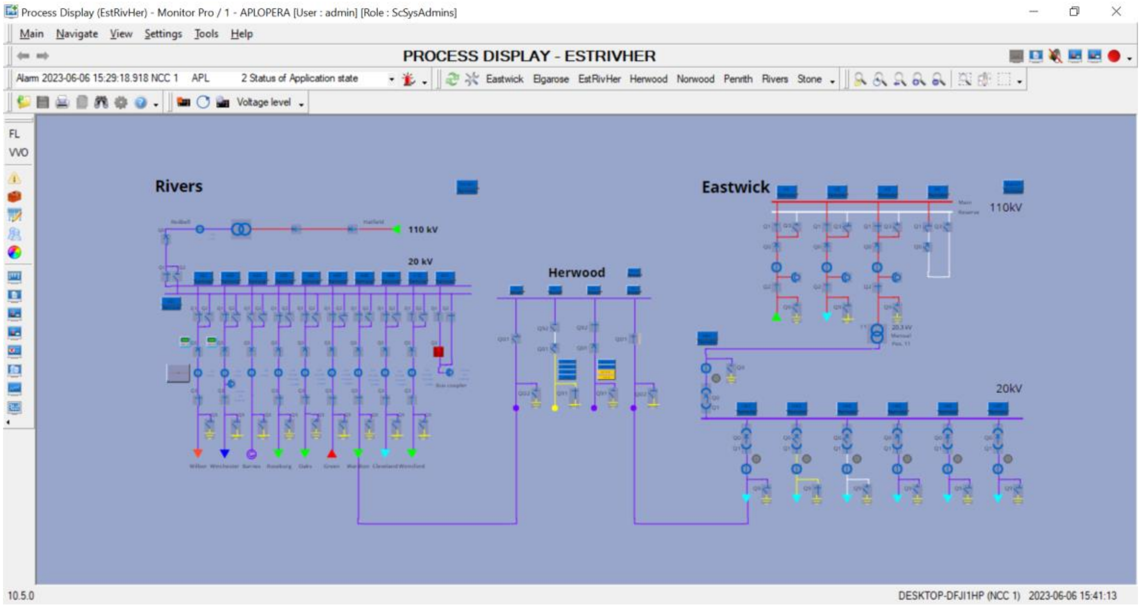

Figure 5: *The process display of SYS600 with three separate substations*

Besides the process display, SYS600 also offers separate displays for events, alarms and trends. The trends display shows different measurements, such as power, voltage and current, as a function of time. Together with measurements reports, this information can be used by the operator to monitor electrical quality and analyse the cause of anomalies, for example.

3.2 MicroSCADA X DMS600

The DMS600 software package consists of two main applications, DMS Network Editor and DMS Workstation, usually abbreviated simply as NE and WS. Next, the main functionalities of NE and WS will be presented before getting into the current load modelling implementation of DMS600.

3.2.1 DMS600 NE

The most important function of NE is to generate the network model into the network database. NE also creates a separate binary file of the network model. This binary file can be read by WS to recreate the same network model. [40] As a network information system, NE also includes many functionalities for network management, such as network calculation, network planning, reliability analysis and asset management. As shown earlier in figure 2, external data can also be imported to the network database. This commonly includes background maps and data from an external *customer information system* (CIS). NE supports both raster and vector maps, and different background maps can be configured for each zoom level. Besides map settings, NE is used to configure many general settings used in both NE and WS, such as component symbols and network colouring preferences. [36]

The main user interface of NE is presented in figure 6:

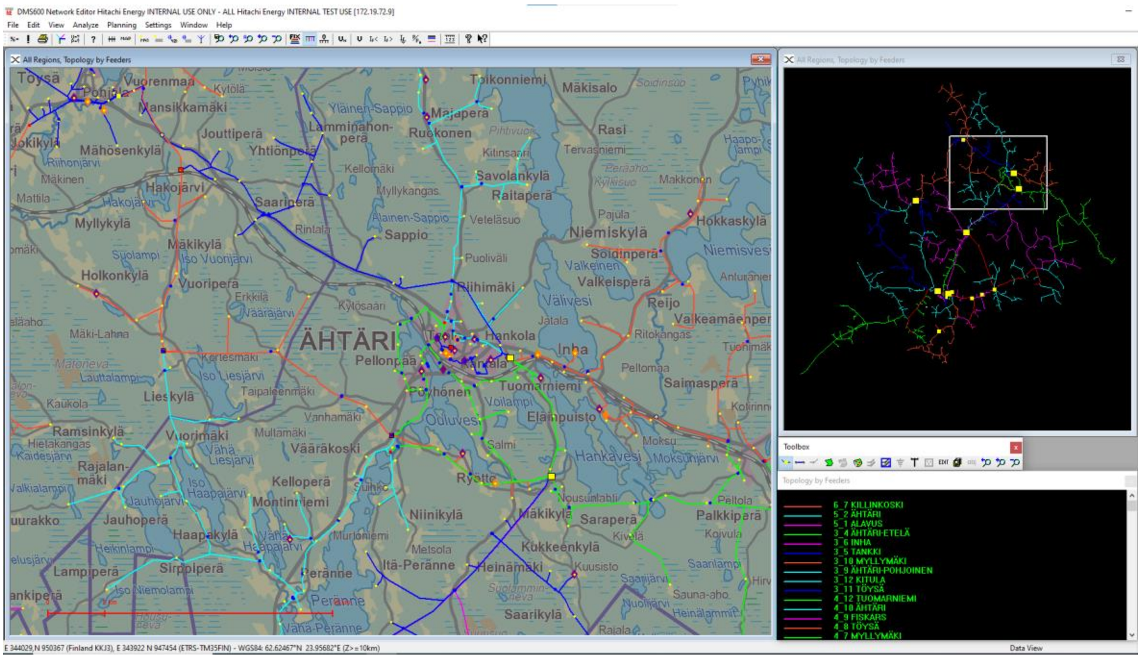

Figure 6: *Overview of NE displaying a medium voltage network*

The main window displays the entire *medium voltage* (MV) network by default, including all substations, line sections and components (disconnectors, circuit-breakers, fuses etc.). *Low voltage* (LV) networks can be loaded either separately for each MV/LV substation or as a cluster from a designated area. By using the embedded editing mode, the user can freely modify the network in the main window by for example removing or adding different nodes or line sections. A separate planning mode can be used for creating network plans. Each plan is stored separately in the database and can be executed at any time to insert the changes into the general network model. [36]

3.2.2 DMS600 WS

Workstation serves the operative side of electrical network management. Typically utilised by the operative personnel of distribution network companies, WS offers various functionalities for monitoring and controlling medium and low voltage networks. Some of the main functionalities offered by WS are fault management, operational simulations, switching planning, field crew management, alarms, and reporting services. [40]

As previously stated, WS reads the binary network file created by NE to recreate the network model. The main user interface is very similar to NE as can be seen in figure 7.

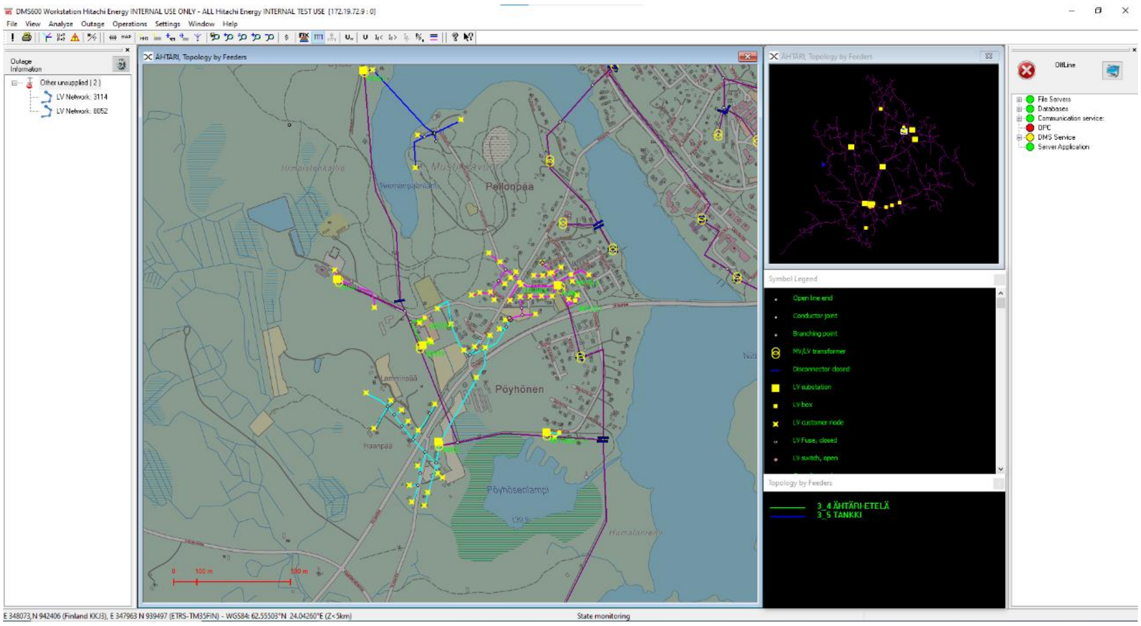

***Figure 7:** Overview of WS displaying both MV and LV networks*

Despite sharing the same network model, WS offers very different functionalities than NE. When linked with a SCADA application such as SYS600, WS receives switch states from remotely controlled devices in real time via OPC connection. Switching states can also be controlled in WS and the state changes are sent to remote devices by first notifying SCADA, which in turn sends the control command to the physical device. For remote controlled switches, clicking a switch in WS opens a control dialog identical to that of SYS600. [36]

One of the most important functionalities of WS is advanced fault management. Nowadays, WS offers a completely automated solution for fault management called FLIR (Fault Location, Isolation, and supply Restoration). FLIR utilises fault current calculations and fault detector data to automatically create sequences that can isolate the fault and restore electricity to as many customers as possible before the fault can be repaired. FLIR is also capable of automatic trial switchings that can further help locate the fault in the case where fault current and detector data is insufficient in definitively locating the fault. Even without FLIR, WS can perform fault location calculations without trial switchings and propose basic restoration sequences. [36]

Another fundamental segment of WS is operational planning. In distribution network operation, intentional or unexpected outages are an everyday occurrence. Unexpected outages are often caused by rough weather conditions or human error, while planned outages are sometimes needed for maintenance work, for example. By preparing switching plans in advance, it is ensured that switchings are performed in a safe and reliable manner. In WS, the operator can also analyse the expected loads to determine the least disturbing time for a planned outage and use the switching planning functions to create and execute necessary switchings. [36]

3.3 Load modelling in DMS600

Currently, DMS600 features two methods for load modelling: Velander's formula and load curves. The desired load model is typically chosen during installation, but the selection can also be later changed in NE settings. [36] In practice, Velander's formula is typically used by non-Finnish DMS600 operators and load curves by domestic operators. The reason for this is simple: most countries have no history in creating load curves from measurement data and the DMS default curves are based on Finnish consumption habits and thus unprofitable for foreign operators.

3.3.1 Velander's formula

Velander's formula was theoretically presented in chapter 2. As mentioned, the Velander coefficients could be defined separately for multiple different customer groups. However, DMS600 only allows defining one set of Velander coefficients for calculations. The default values for k_1 and k_2 are 0.28 and 0.08, respectively. These default values can also be changed in NE settings. [36] As Velander is not useful for estimating the loads of individual customers, load forecasts are only available in MV networks when the Velander model is in use.

3.3.2 Load curves

Compared to Velander's formula, load curves provide a more comprehensive solution for load modelling. While Velander's formula can only be used to estimate peak loads of entire customer groups, load curves model the electricity consumption of individual customers, which varies in quantity and over time. The principles of load curves were already presented in chapter 2 so we will now review how the use of load curves is implemented in DMS600.

The default installation of DMS600 contains a small set of load curves from the 1992 load research study. However, since the predefined load curves are obsolete and increasingly inaccurate, practically all domestic DMS600 operators have created their own set of load curves from metering data. The importation and creation of new load curves is supported in DMS600. In addition to creating general, customer class specific load curves, it's typical to create individual load curves for major customers. The largest consumers often have quite unique consumption habits that wouldn't fit very well to a general customer class load profile. Moreover, the electricity demand patterns of these customers are often more constant than those of typical households and thus an individual, measurement history based load curve can be a very accurate way of modelling their consumption. Due to the creation of individual load profiles, many DMS operators have dozens or even hundreds of separate load curves in use.

For load estimation purposes, DSOs ordinarily assign each customer to a predefined group based on their type, such as function and heating solution. For example, apartment houses with district heating and without an electric sauna could be one customer group. Each group has their own load curve. This information is stored in the operator's customer information system. DMS600 receives this information along with the general customer information importation and saves it to the customer table in the network database.

The load curve itself consists of two binary files that contain the load data from the corresponding customer or customer group. All load curves are scaled to a total consumption of 10 MWh (10,000 kWh). One binary file contains hourly load data from the entire year, while the other binary file includes hourly deviation data. The deviation curve provides information on how much the demand on similar times fluctuates. DMS600 base installation includes a separate tool for examining and editing load curves, called *LoadCurve.exe*. The load curve tool features three separate views: day, week and the whole year. The week view is presented first in figure 8.

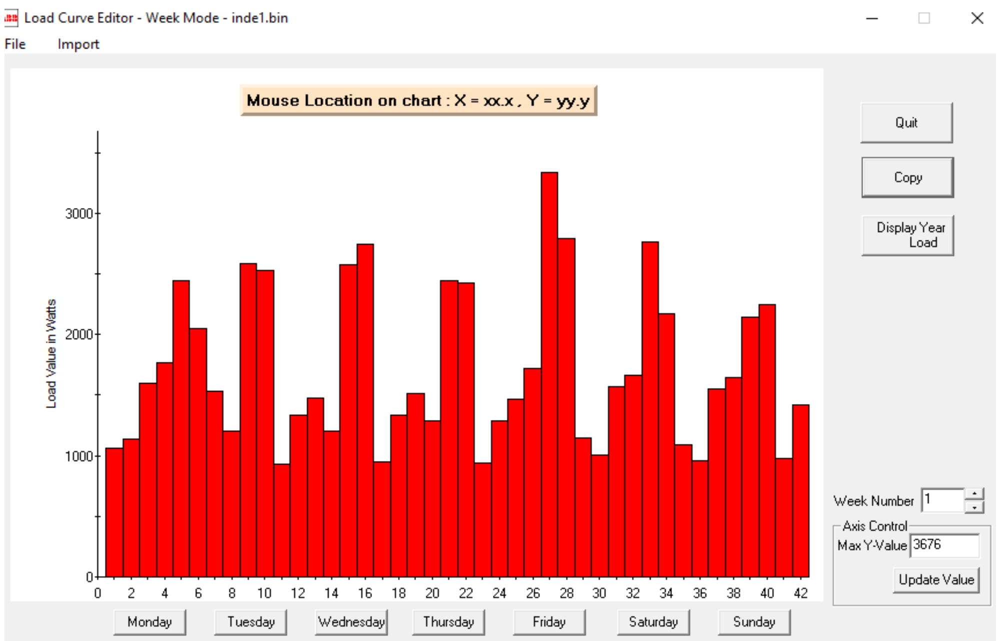

Figure 8: *Week view of an hourly load data file opened in LoadCurve.exe*

When opening a load curve, the first week of the year is shown by default. The week view splits the week into 42 sections, covering 4 hours each. In other words, each column tells the average load of a 4-hour period. The user can select which week is displayed in the *Week Number* control. Clicking one of the weekday buttons will bring forth the day view of the corresponding day of the week in question, as displayed in figure 9.

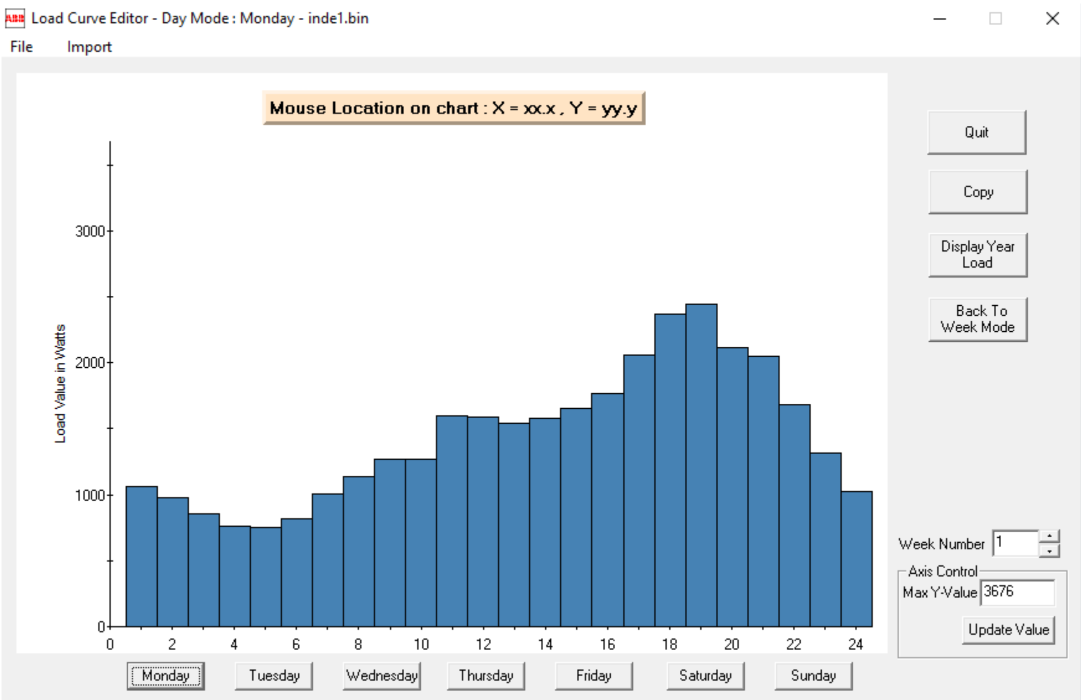

Figure 9: *Day view of an hourly load data file opened in LoadCurve.exe*

The day view is more intuitive: it shows the load of each hour of the day. Lastly, by clicking the *Display Year Load* button, the tool displays the average load of each week of the year. Figure 10 presents the year load view.

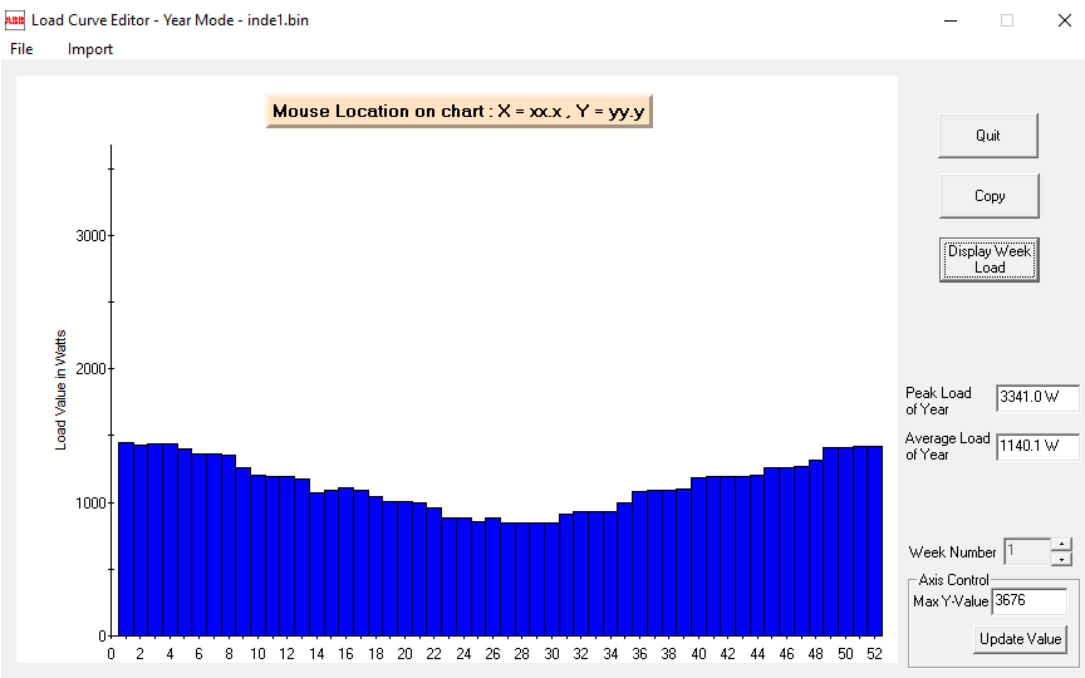

Figure 10: *Year view of an hourly load data file opened in LoadCurve.exe*

The basic functionality is same across all views. Exact loads are displayed by hovering over the columns. The load values can be edited by simply dragging columns to the desired values.

When load curves are in use, both NE and WS offer the possibility to see a graphical load forecast for any node or node section for the upcoming week. In WS, the user can select to forecast the next 1-168 hours for MV networks while in NE, the whole week is always displayed. In MV networks, the graph consists of four curves: voltage, reactive power, active power, and active power with estimation correction. The active power curve is based on the sum of customer class specific load profile curves which are scaled with the customer yearly energy consumptions. The expected value for each exact hour is presented, which corresponds to a 50% confidence range. The estimation curve displays forecasted active power calculated by *Esti*, an estimation program that uses real-time power measurements to improve the accuracy of the load profile based modeling. Esti is only used in MV networks. [36] Example of a MV section load curve (forecast) can be seen in figure 11:

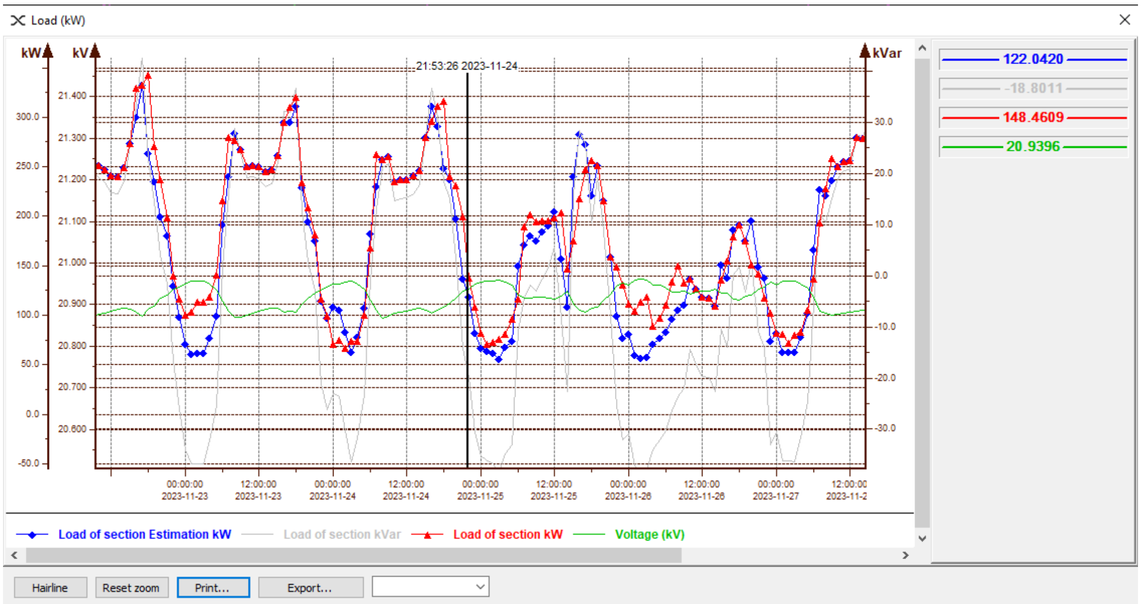

Figure 11: *Load curve (forecast) for a MV network node section*

The user can freely edit the graph settings, including the color, markers and width of each curve, the axes scaling and the graph type. For example, a bar chart can be selected instead of the default plot line. In figure 11, the active power curves are purposefully highlighted.

When a load forecast is requested in LV networks, the whole week is always displayed. Since Esti isn't applicable to LV networks, the graph doesn't contain an estimation curve. Instead, a load statistical confidence curve is displayed. This curve displays a confidence coverage according to a statistical factor. The default statistical factor is 1.6, which means that the load displayed by the confidence curve is the expected value + 1.6 * standard deviation [36]. In traditional load profiling, the load dispersion of similar customers at a specific time is assumed to be normally distributed [26]. This means that the factor 1.6 corresponds to around a 90% confidence range. In other words, the actual load is below the shown value with 90% certainty. The statistical factor can be freely selected in NE settings. Figure 12 shows an example of a load forecast for a LV network node section.

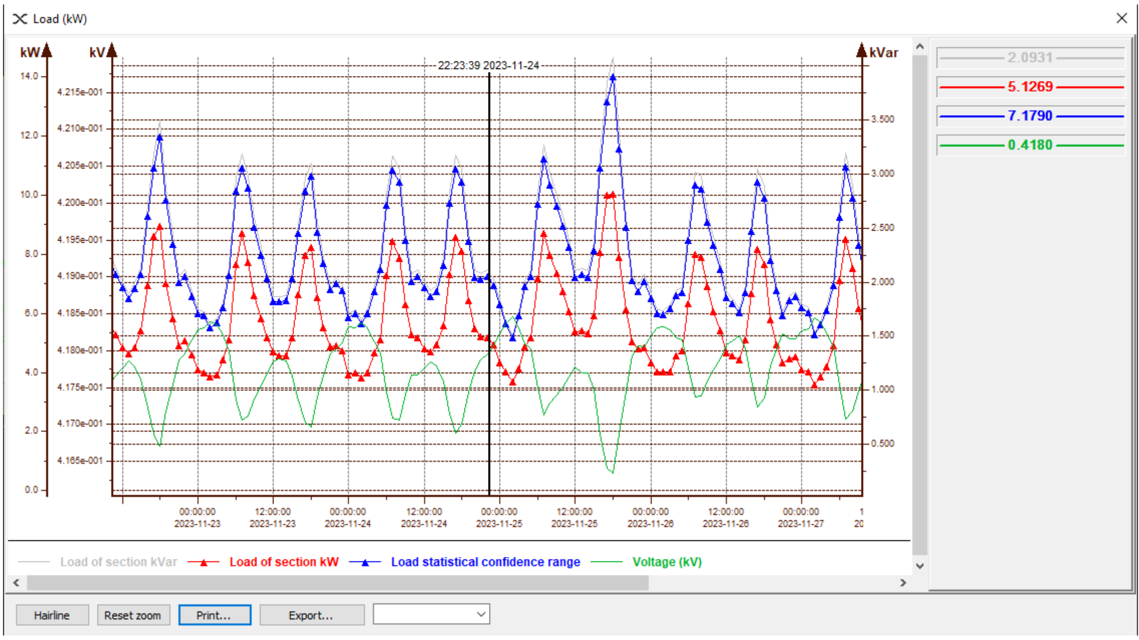

Figure 12: *Load curve (forecast) for a LV network node section with a 1.3 statistical factor*

This graph uses a statistical factor of 1.3, which corresponds to around an 80% confidence range. Since the active power curve displays a 50% confidence range, the confidence curve always shows a higher value, as it presents a higher confidence range. With a statistical factor of 1.6, the confidence curve values would naturally be even higher.

The load curves together with the estimation tool offer the most accurate method for system state estimation in DMS600. As examined in chapter 2, distribution system state estimation and load forecasting are extremely important in network operation.

3.3.3 Weaknesses of current implementation

The current load modeling features in DMS600 that were presented earlier in this chapter are not ideal for estimating and forecasting modern day electrical load patterns. While acting as a novel conceptual approach to load modelling, the model that is developed as part of this work also aims to improve the load modelling functionalities in the DMS600 software. The main defects of the existing load modelling methods have already been examined so this subchapter will only shortly revise the weaknesses in DMS600 load modelling functionalities.

The DMS600 base installation only features a small batch of customer class load curves. These curves alone cannot encompass all the varying electricity consumption patterns present in today's electrical grids. As a result, customers are inevitably assigned to classes to which they don't fully adhere to. In addition, the existing load curves originate from a load research study from decades ago, so the load patterns no longer accurately represent the current electricity consumption. To mitigate this issue, practically all DMS600 operators have imported multiple additional load curves into the software. While the importation of external load curves is supported at a basic level, this creates tons of excessive work for the DSO.

Another deficiency of traditional load profiling is that a possible change in customer class is often left undetected. For example, if a customer with direct electricity heating installs a heat pump, their consumption pattern changes dramatically which would require moving the customer to another customer class. However, these types of changes are rarely reported to the DSO by the customers. As a result, an increasing number of customers remain assigned to wrong customer classes. Besides incorrect profile assignment, another issue with load profiling is that there is always a certain number of customers that do not conform to any customer class load profile. Especially large consumers, such as industrial customers, can have very distinctive consumption patterns. For these customers, creating an individual load profile is often required, which relies on manual labor in the current implementation.

All the abovementioned problems could be solved to some degree with a customer specific load model. When load estimation is done individually, purely based on historical consumption data, any possible changes in consumption habits are automatically considered, albeit with a certain delay. There would no longer be a need to classify customers to specific groups, which reduces the inaccuracies caused by generalization. Also, this would solve the problem with customers that do not clearly belong to any group.

Despite its deficiencies, load profiles offer a much more comprehensive solution to load modeling in DMS600 than the alternative: Velander's formula. Since the existing load profiles are only applicable to Finnish electricity consumers, foreign grid operators must rely on Velander's formula for load modelling. As explained earlier, Velander's formula is not the most versatile method as it is only suitable for roughly estimating the peak loads of larger customer groups. In theory, the model developed in this work could be used by foreign customers in addition to the domestic market, but it would require providing metering data in the specific format used by Finnish DSO's, which might not be feasible for many foreign operators.

4. DESIGNING A CUSTOMER SPECIFIC LV NETWORK LOAD MODEL

Based on the presented background knowledge on load modelling so far, the focus can be shifted into developing a novel algorithm for modelling and forecasting LV network loads with historical measurement data. In the beginning of this chapter, the process of importing metering data into DMS600 is briefly examined. This serves as the backbone for implementing the new load model and gives knowledge on how metering data is stored and transferred in Finland. Afterwards, designing the new algorithm and implementing the model into DMS600 is examined, and a straightforward test scenario is proposed to determine the accuracy of the newly developed model.

4.1 Smart meter data

Before a new load modelling and forecasting algorithm can be implemented and tested in a real-life network, time series data needs to be imported into the DMS600 database. This subchapter will briefly introduce the file format uniformly used to transfer time series data in Finland, which will also be used to import data to DMS600. Furthermore, a short overview of how metering data files can be imported and stored in the DMS database is given. For this thesis, an extensive set of actual smart meter measurements was received from a DSO operating in Western Finland. This included metering data from their entire network for the past 4 years, adding up to a total of around 500 million separate measurements.

4.1.1 Transfer file format

In order to collect, transfer and handle measurement data from smart meters, a somewhat standardized data format is needed. As measurement data is essentially tabular data in huge quantities, a *delimiter-separated value* (DSV) format is a natural approach. When Fingrid's Datahub project began, a need for a documented, clearly standardized data format for exporting and importing time series data emerged. Eventually, a new data format called *standard ASCII format* (SAF) was introduced. Today, all Finnish DSOs are required to submit measurement data from their networks to Datahub as SAF files. Since all Finnish DSOs are already required to compile their metering data into SAF files, the same format will be used in importing the data to DMS600.

An example of a SAF file is presented in figure 13:

```
EXH;2;20010301200020+00;
TSH;AT004000050200000000000001050884;1;15;MIN;kWh;4;200111200100+00;200111
200200+00;Ref1;Ref2;Ref3;Ref4;Ref5;Product;MPCode;Meter code;Meas code;
TSV;1;200111200100+00;10.2111;Measured;
TSV;2;200111200115+00;10.2222;Measured;
TSV;3;200111200130+00;10.2333;Measured;
TSV;4;200111200145+00;10.2444;Measured;
TSH;AT004000050200000000000001050500;1;1;HOUR;MWh;8;200111200000+00;200111
200800+00;Ref1;Ref2;Ref3;Ref4;Ref5;Product;MPCode;Meter code;Meas code;
TSV;1;200111200000+00;20.2111;Measured;
TSV;2;200111200100+00;30.2222;Measured;
TSV;3;200111200200+00;40.2333;Measured;
TSV;4;200111200300+00;50.2444;Measured;
TSV;5;200111200400+00;60.2111;Measured;
TSV;6;200111200500+00;70.2222;Measured;
TSV;7;200111200600+00;80.2333;Measured;
TSV;8;200111200700+00;90.2444;Measured;
EXT;
```

Figure 13: *Example of SAF file contents, directly from [42]*

The same SAF file format is also used for transferring gas grid time series data into Gas Data Hub, the gas counterpart of Datahub [43], [44]. For this reason, the format offers a wide variety of data fields and formatting options, not all of which are relevant for electricity metering data.

As the example shows, SAF is one type of DSV with each value separated with semicolons. The purpose of each row is indicated with a three-letter abbreviation. The following abbreviations, also present in the data example above, are commonly used for electricity metering data:

- EXH – Export header
- TSH – Time series header
- TSV – Time series value
- EXT – Export trailer

Each file always begins with an *export header* row (EXH) and ends with and *export trailer* (EXT) row. The export header mainly tells the export processing time. All times in SAF files are in the format YYYYMMDDhhmmss+XX, where XX is the UTC offset. [42]

The function of *time series header* (TSH) rows is to give information about the following *time series value* (TSV) rows. Among other details, a TSH row tells the start and end time of the reporting period, the metering point code, as well as the unit and time step of the measurements. The TSV rows then simply include the start time of the interval and the associated value as a decimal number. [42]

4.1.2 Database solution

As mentioned earlier in chapter 3, DMS600 almost always features two separate databases, one for operational data and one for network data. Since the metering data is a somewhat separate entity and the amount of data is very large (dozens of gigabytes), creating a separate database for the data could be reasonable to avoid flooding already large databases. However, that would complicate implementing both the data importation and application extensions that access and handle the data, since all existing DMS600 functionality is based on a two-database system. For this reason, it was decided to store the metering data in the existing network database of DMS600.

Not every piece of data included in the SAF files is relevant for calculating a load estimate or forecast. The necessary information, along with the structure of the new table created into the network database for storing the measurements is presented in table 1. The data types were selected to match the specifications given in the technical description of SAF files [42].

Table 1: *The structure of the measurements table*

Column name	Data type	Additional information
MeteringPointCode	nvarchar(35)	Primary key, not null
MeasurementDateTime	datetime	Primary key, not null
Value	float	Not null
Unit	nvarchar(6)	Not null
LoadType	nvarchar(35)	Primary key, not null

The metering point code can be used directly to identify the customer and the grid connection in DMS600. The unit and load type are relevant information because the metering data can also include reactive power measurements alongside active power measurements. The unit used to measure reactive power is *volt-ampere reactive* (var), while active power is measured with *watts* (W). The load type is reported separately in the TSH rows, with P+ or P- signifying active power measurements, and Q+ or Q- signifying reactive power measurements. P+ indicates ordinary active power taken from the grid, while P- is designated for production, i.e. active power fed into the grid. Meanwhile, Q+ refers to *inductive power* (positive reactive power caused by inductive loads) and Q- refers to *capacitive power* (negative reactive power caused by capacitive loads).

The table's primary key was determined to be a composite, consisting of the metering point code, measurement date and time, and load type, since the combination of these three values is always unique. Also, these are the main filters used when performing queries into the table, and a clustered index is automatically created on the primary key, which speeds up the queries.

4.1.3 Importing metering data

It was briefly mentioned in chapter 3 that DMS600 includes a service framework called DMS Service. In practice, DMS Service runs and keeps track of many separate background services that are responsible for a wide variety of tasks, related to for example communication, fault management and timed data transfer. It is often necessary to import or export data between DMS600 and different customer systems and for this reason, many different modules for data transfer have been implemented into DMS Service. Consequently, implementing a new service module for handling the importation of metering data seems reasonable. The service framework also includes ready-made functionality for automated operation, as well as logic for connections to DMS databases.

As suggested, a new service module was written in C# to handle metering data importation. Through an existing graphical interface called Service Monitor, the user can configure the importation process. For example, the user can select whether outdated measurements are automatically deleted from the database after the import, as they are no longer useful for the new load model. One at a time, the import module reads all SAF files from a folder path given by the user. The file is first read to memory into a list of rows. The data is then parsed row by row and saved to temporary tables in the C# code in batches of 5000 rows. Each batch is then saved to the database table described earlier. However, each SAF file is saved to the database in one transaction, so that in case of a failure in reading, parsing, or saving the data inside one file, the transaction can be rolled back. The user is presented with progress loggings on the service monitor.

Reading, parsing, and saving hundreds of millions of rows can be very time consuming. For this reason, the efficiency of the importation is vital but even more importantly, normal operation must be able to continue despite an ongoing import. Consequently, nearly all functions in the import module were written to be asynchronous, meaning that the processing thread can run other tasks in parallel, and normal system operation can continue.

Naturally, the large bunch of historical data only needs to be imported once, when the system is taken into use. Afterwards, the import can be scheduled to run for example every night or once a week to bring in new measurements. With a much smaller data set, the import is no longer time and performance consuming.

4.2 Developing a smart meter data based load model

Now that the necessary measurement data can be saved into the database, a novel load model can be designed and implemented into the DMS600 software for testing. In this subchapter, the general algorithm is first designed and reasoned. The algorithm is then turned into a model that is implemented into DMS600 Workstation.

4.2.1 Mitigating load stochasticity

As examined in chapter 2, the use of smart meter measurement data has been studied in literature to improve load modelling. However, the focus has mainly been in developing new load profiles, improving existing profiles and classifying customers to clusters based on their consumption habits. Directly using the historical measurements of individual customers to create customer specific load estimates and forecasts has not been proposed, albeit in his dissertation, DSc Antti Mutanen noted that past year measurements should not directly be used to estimate individual loads as they fail to factor in temperature differences, calendar changes or the stochastic nature of electricity consumption [32]. In this thesis, estimating individual loads purely using historical data is attempted while trying to mitigate some of the defects stated by Mutanen. In short, this means using statistical analysis to calculate an estimate for the load of one hour by using directly comparable historical measurements as data points.

Maybe the greatest difficulty in estimating the average load of a single hour based on historical data is indeed the stochasticity of electrical loads. While some network customers, such as industrial consumers, might adhere to a stable and repetitive load curve, most consumers are individual households that have rather unpredictable consumption patterns. Many home appliances operate on a relatively high power (multiple kilowatts), which drastically changes the load on a one-hour period. For example, an ordinary electric sauna stove might have a maximum power of 10 kW. This effect is emphasized in apartments with district heating since their base consumption is usually low. One can imagine that this leads to a scenario, where comparing, say, the hour between 19:00 and 20:00 every Thursday, gives very different results based on whether a stovetop or an electric sauna was used or not. The sporadic nature of individual consumption is demonstrated in figure 14. The figure shows the hourly loads of three randomly selected household customers from one week in July 2023. The loads are actual measurements. All three customers are from the same LV network, and they are all apartments with district heating. The figure also includes the statistical load curve of this specific customer class.

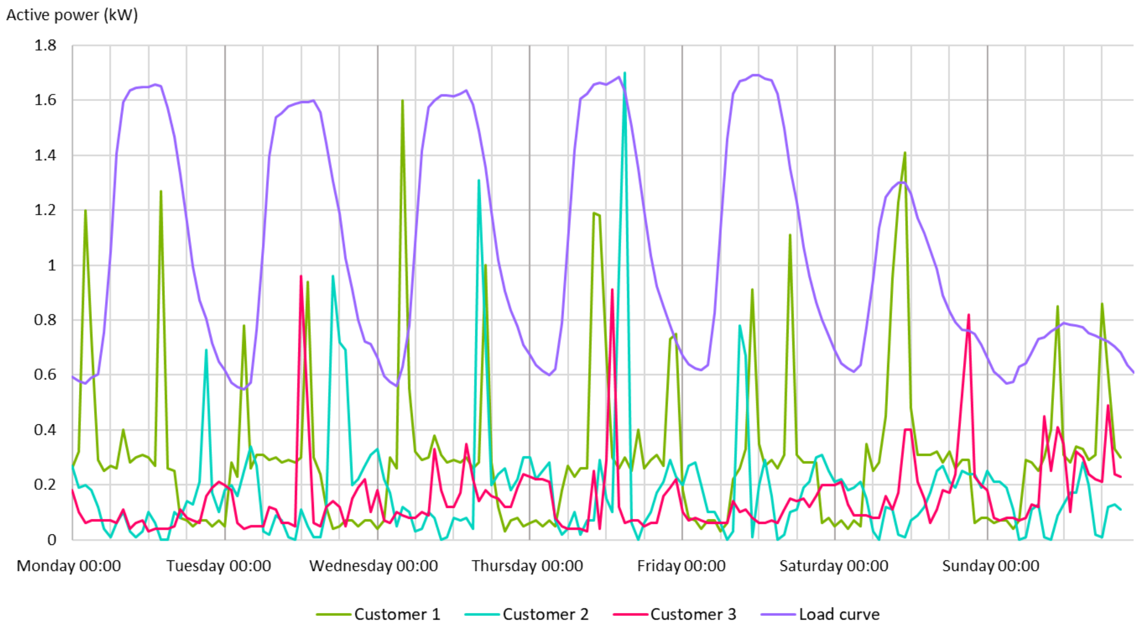

Figure 14: *Average hourly loads of three similar customers and the customer class load curve. Data from July 2023.*

The fact that the loads of all three customers are systematically lower than the associated load curve is somewhat irrelevant and mostly because the customer class load curve is scaled to an annual consumption of 10,000 kWh. The more important takeaway is that the shape of the customer consumption curves is nowhere near the load curve; they are also very different with each other. The base load of each customer can be observed to be low, while each series contains multiple distinct load spikes at varying times. The peak power is thus very high compared to the average power, but the duration of each peak is very short. This type of behavior is typical to household electricity consumption. The unpredictability balances out when the number of consumers increases as chance variation decreases, which is also the base of load profiling. The sum of 100 similar consumers is already very likely to mirror the customer class load curve behavior. [26] However, the problem persists if we attempt to forecast the loads of an individual consumer.

While the problem of stochasticity cannot be fully mitigated, using multiple data points for each customer will inevitably lower the relative error, similarly to summing up the loads of multiple customers to form a load curve. This raises a new question: which time points are directly comparable with each other? It is apparent that the same exact hour of the same day from last year is a valid data point. For this thesis, it was decided that data from the past 4 years is usable. Older data becomes progressively more unreliable, and it would unnecessarily increase the amount of raw data that needs to be imported and stored. Meanwhile, using less data would decrease the amount of data points, also adding unreliability to the analysis.

Traditional load profiling utilizes a three-way model where the week consists of weekdays, Saturdays and Sundays/holidays [26]. In other words, all weekdays are assumed to be identical, and all public holidays are assumed to produce similar load patterns than Sundays. Using this same approach provides us with more data points for load calculations: If weekdays are assumed identical, it is safe to assume that the load of the same hour from yesterday or two days ago is comparable to the estimated hour. Seasonal variance does however start to play a role as we move further from the observed point in time. While the average temperature is not likely to have a significant difference in a span of one week, there is likely a meaningful difference in a span of three weeks. Similarly to using data from previous years, a balance between more data points and the accuracy of each point needs to be balanced. For the new load model, it was decided to use the measurements from past three weeks.

So far, it was determined to utilize the corresponding hours from the past three weeks, and the exact same hours from the past 4 years as the base data for the new load estimation algorithm. In practice, this means that for weekdays, we get 19 data points: 15 from the past three weeks and 4 from the past years. For Saturdays and Sundays, we only get 7 data points, as there are only three corresponding days in the previous three-week period. Hypothetically, this would suggest that the estimation will be more accurate for weekdays. Also, due to the higher number of data points, the measurements from past weeks have more weight for weekdays.

4.2.2 Eliminating outliers

As stated, electricity consumption has a large variance and thus historical data is likely to include extreme values that should be discarded from the calculations. A common approach to finding outliers or extreme values in a data set is to calculate the standard deviation of the data and determine outliers based on their distance from the mean or expected value. However, the presence of extreme values in the data strongly affects the standard deviation and mean value, which makes this method unreliable when analyzing a small data set. The amount of data points in this case is very small which means that even one outlier would significantly affect the mean and standard deviation value.

The data is also not normally distributed. To reliably find outliers from such data, we could use the *median absolute deviation* (MAD) measure.

MAD is a robust method for measuring statistical dispersion in a data set that is not normally distributed. It is not dependent of mean and standard deviation and thus not affected by the presence of extreme values. [45] The use of MAD for outlier detection is demonstrated in the following example:

> Consider the following data set that consists of 9 values: (0, 2, 3, 4, **4**, 5, 7, 8, 12). The median value of this set is 4. As such, the absolute deviations of each number from the median are (4, 2, 1, 0, 0, 1, 3, 4, 8), which is sorted into (0, 0, 1, 1, **2**, 3, 4, 4, 8). The median of this set, in other words the median absolute deviation of the original data, is 2.

To determine which values in the set are outliers, a boundary needs to be set. For example, one could determine that all values that are over 2 MADs away from the median should be flagged as outliers. With this threshold, only the value 12 would be considered an outlier. This seems like a reasonable conclusion when looking at the data set. The cutoff boundary can be freely selected. An iterative approach is often needed to find the best solution for each data type.

4.2.3 Weighting data points

After the outliers have been flagged and removed from the data set, all remaining data points can be assumed valid. The value of each data point still needs to be considered. It is evident that a data point from yesterday reflects the current status better than a data point from four years ago. Assigning weights to each data point and calculating a weighted average is a straightforward way to assess the data. Analogously to the boundaries of outlier elimination, an iterative approach is often the easiest solution for determining optimal coefficients. Consequently, an educated guess was made. To begin with, the weights were assigned based on data recency as displayed in table 2.

Table 2: *Initial weights assigned to data points*

Data age	< 1 week	1-2 weeks	2-3 weeks	1 year	2 years	3 years	4 years
Weight	1.0	0.6	0.2	0.8	0.6	0.4	0.2

Data from the previous week was deemed most reliable. Firstly, it is most recent and thus not affected by fundamental changes in consumption patterns or seasonal changes. Secondly, data from less than one week ago is likely to reflect the prevailing weather trend at least with some accuracy. As we move further away from the present date, data reliability quickly deteriorates due to possible seasonal changes. Thus, the data between 2 and 3 weeks of age is assigned a rather low weight. When it comes to data from previous years, the main sources of unreliability are weather variance and fundamental changes in consumption habits. Even if the data from previous years presents the same day of the year, the prevailing weather trends can be very different between separate years, particularly during spring and autumn when large temperature variances are frequent. However, previous year data still provides a highly comparable baseline since the time of the year is exactly the same. As for fundamental changes, for example the acquisition of a heat pump would significantly change the consumption pattern, thus turning older data somewhat useless. For this reason, the weights are systematically reduced according to the data recency. Once the data is prioritized, a weighted average can be calculated.

4.2.4 Model implementation

To summarize, the following superficial algorithm is proposed for creating a customer specific load forecast:

1) Fetch metering data from the past three weeks and past four years for the corresponding hours.
2) Find and eliminate outliers using median absolute deviation (MAD).
3) Assign weights for different measurements, prioritizing more recent measurements separately for data from past weeks and past years.
4) Calculate a weighted average from the measurements separately for each hour. This will be the load estimation for each hour.

When fetching past data, the same weekday is searched from the database. For example, the last year’s and last week’s equivalent for Friday 12th of May 2023 would be Friday 13th of May 2022 and Friday 5th of May 2023, respectively. Public holidays occurring on weekdays could skew the results and should optimally be considered, even if the outlier elimination might remove some of the deviating data. Due to the limited implementation time frame, the consideration of public holidays was left out of the forecast algorithm. However, the author suggests a high priority for this as a future improvement.

The above algorithm shall be the base of first model implementation into DMS600. The model was first implemented into DMS Workstation’s LV network calculation. In DMS, LV networks are separately loaded into memory by command. When an LV network is being loaded, the load of each node section in the network is calculated. Traditionally, the program first loads all the electricity connection objects into memory. Each object contains information about how many customers belonging to each load curve are contained in that connection, as well as each customer's annual energies. Many connections only have one customer, but for example apartment buildings contain multiple customers, which could potentially belong to different load curves. With the load curve numbers and annual energies, the load of each node can be estimated. These can then be summed to estimate the load of each section in the network. With the metering data model, when an LV network is loaded, all metering point codes that include measurements in the database are saved into the LV network object. The new estimation algorithm is then used to separately calculate a load estimate for each customer.

Explained with more detail in chapter 3, a short-term load forecast can be requested for any node or section in the LV network in DMS600 Workstation. The program then presents a graph that includes a forecast for active power, active power statistical confidence, reactive power, and voltage for the next 168 hours from the calculation time, in other words for the following week. The metering data model will replace the calculations behind the active power curve in the backend. Since the developed model does not contain information about load statistical dispersion, the active power statistical confidence curve will be identical to the standard active power curve. If the metering data model fails to calculate a reliable estimation, the old load curves will be used, which means that the statistical confidence curve will also be visible for the faulty period. An example of a successful metering data estimation is presented in figure 15.

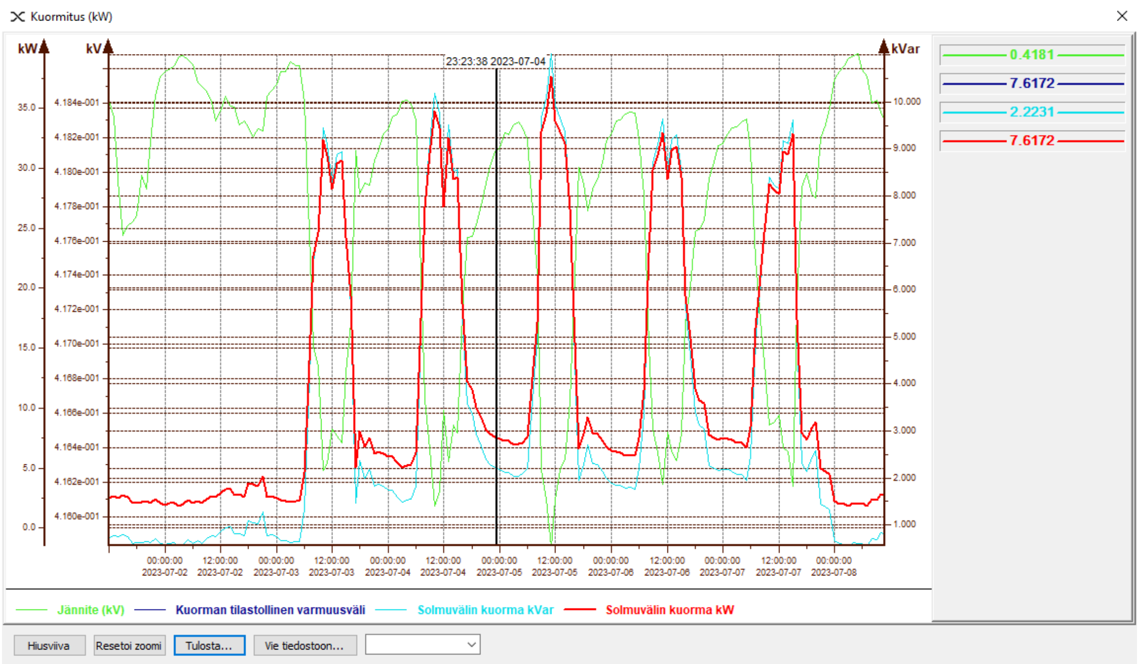

Figure 15: *One week load forecast in Workstation using metering data calculations*

As can be seen, the statistical confidence curve is identical to the active power curve and thus not visible. From a user experience perspective, the best approach would be to not include the statistical curve at all when it serves no purpose. However, the curve is an easy way to instantly tell whether the metering data calculations were successful or not, and thus it was kept to facilitate this work.

In this work, the model is only implemented into LV network calculation and short-term forecasts in DMS600 Workstation. While in theory, the same logic of calculating the loads of each individual customer could be extended to MV network load calculations, it was quickly noticed that for DMS600, the implementation would be very laborious and time consuming. As explained in chapter 3, DMS600 calculates MV network loads by combining load curve information with estimation correction, the latter of which uses real-time measurements to adjust the load curve estimation. The entire MV network is always calculated at once so due to the massive number of customers, separately calculating loads from metering data for each customer in the whole network is very likely too inefficient. Instead, the calculation should be performed for a limited area, such as one substation. The basic estimation derived from load curves could be replaced with a sum of all downstream customer load estimations calculated with the metering data model. However, since the implementation for LV networks was already demanding, it was decided to keep the scope of this thesis on LV network modelling.

4.3 Designing a test scenario

Now that a functional model has been implemented, a test scenario can be designed to determine the accuracy of the new model. The short-term load forecast tool in Workstation contains the option to export the graph data into a csv file. This allows us to gather data from both the metering data model and the old load curve model. The CSV files can be taken to excel, where the data can be analyzed. Due to the large variance in electrical loads, mainly caused by load stochasticity and temperature differences, a large sample size is required to achieve reliable results. Data should also be collected from different seasons, as electricity consumption patterns differ systematically according to the season.

To address these preconditions, data will be collected from three large, arbitrary LV networks, from all four seasons. The networks 1, 2 and 3 have a total of 187, 175 and 166 customers, respectively. In other words, data from a total of 528 real-life customers shall be collected. Each LV network will be analyzed separately as a sum of all the loads inside the network. The time periods to be analyzed are October 2022, January 2023, April 2023, and July 2023. The metering data from the first week of these four months will be pulled from the database, and a short-term load forecast will be created for these four weeks using both the old load curves and the new metering data based model. This can be done by setting the calculation time in Workstation to for example 1/1/2023 12:00 and then requesting the load forecast for the entire LV network, which would then give a load forecast for the first week of January. The data will be exported to an excel spreadsheet where the data will be analyzed. Three key error metrics will be calculated and presented to offer a comprehensive evaluation of the model accuracy.

5. RESULTS AND ANALYSIS

Actual measurements, metering data based forecasts and load curve forecasts were all collected and analysed according to the methods described in chapter 4. Before the final results are displayed and dissected, some observations from the testing phase and consequent corrective actions are presented.

Initially, the threshold for outlier elimination was set to be 2 times MAD. This threshold is a rather standard approach in outlier elimination [46]. The initial tests showed that the calculated load forecasts were systematically lower than the actual loads. While the average error seemed reasonable, systematically providing too low load forecasts would be harmful, as loads should rather be overestimated than underestimated in operational planning. The reason for the systematic error was not obvious, so calculation parameters were one by one altered to find the cause. Eventually, it turned out that the systematically small estimations were caused by the outlier elimination. While the outlier elimination used the same threshold for eliminating relatively large and small results, the algorithm eliminated significantly more large values than small values, thus negatively affecting the load estimation. This is likely caused by the nature of electricity consumption: As examined earlier, in residential use, loads are at the base level most of the time. Larger spikes in loads are typically caused by the use of powerful household appliances, or in Finland, heating up an electric sauna. Powerful appliances are only used intermittently, albeit still regularly, so the load spikes they cause should not be entirely neglected in load calculations. However, as the load is near base level most of the time, the median value and the median absolute deviation stay rather low, which results in the elimination of almost any larger values. This hypothesis was proven right by temporarily removing the outlier detection from the calculations: this resulted in a much more evenly distributed error when the calculations were compared to actual measurements. However, to no surprise, the calculated values were now systematically a tad too high.

Figure 16 shows an example of the relative forecast error with the initial outlier elimination, as well as the relative error without any outlier elimination.

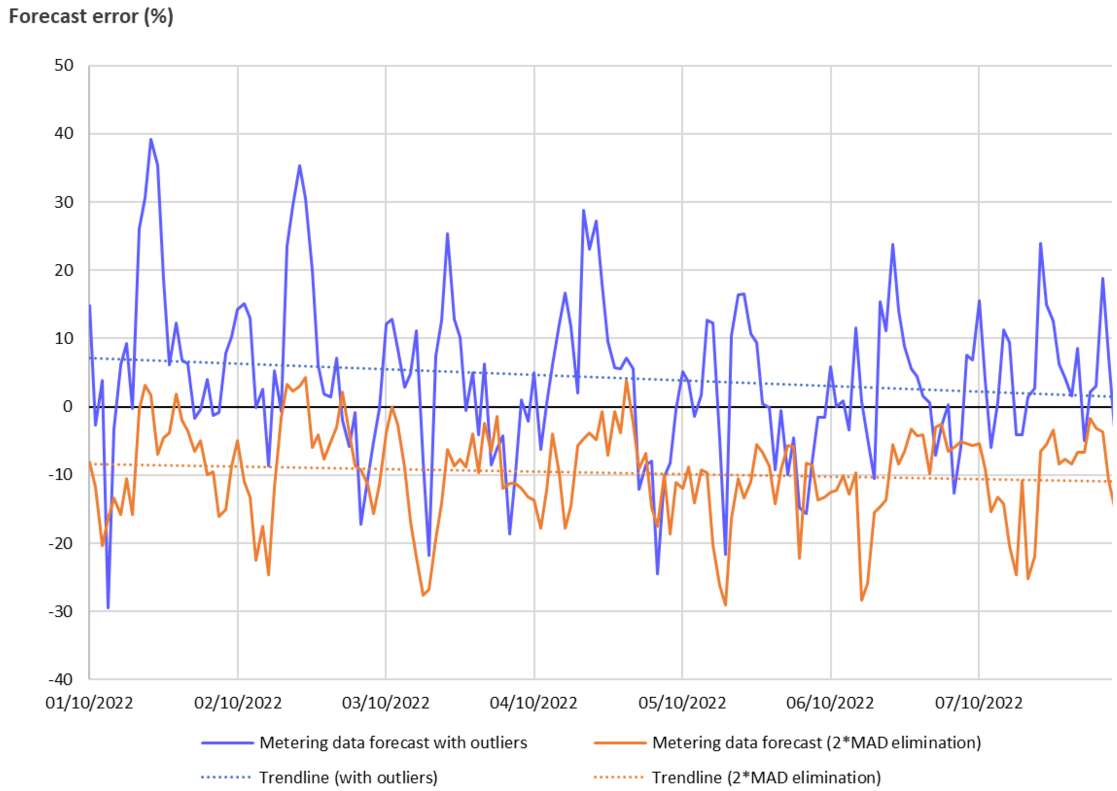

Figure 16: *Example of the forecast relative error with the initial outlier elimination and without any outlier elimination*

The abovementioned issues are clearly visible in the graph: with the initial outlier elimination, the forecasted values are systematically too low, but if outliers are not eliminated at all, the forecast includes frequent spikes where the loads are highly overestimated.

This suggested that some outlier elimination is still needed for abnormally large values. Simply raising the threshold for outlier elimination would likely provide an improvement to how large values are handled but would probably result in complete negligence of abnormally small values. A small example to depict the issue:

> Consider the following set of values: 0, 1, 1, 2, **2**, 3, 6, 7, 14. As explained in chapter 4, the median value of this set is 2 and the MAD would be 1. With a threshold of 2 * MAD, the values 6, 7 and 14 would all be eliminated. While the value 0 could be seen as an outlier since an electrical load cannot physically be any lower, it is not eliminated because the base load is so low. Meanwhile, values 6 and 7 seem like values that maybe should be included.

This phenomenon suggested that a different threshold should be used for small and large values. A narrower threshold is needed for small values and a broader one for large values. The optimal thresholds were tested iteratively to see which boundaries provided the most accurate estimations compared to actual measurements. In the end, a threshold of 1,5 * MAD was selected for the values smaller than the median and 5 * MAD for values higher than the median. If these thresholds would be used for the data set in the earlier example, the values 0 and 14 would get eliminated, while the previously eliminated values 6 and 7 would be included. This also seemed to produce the best results in real life scenarios, in terms of both the absolute error and the distribution of too low and too high estimations. For comparison to the initial threshold and to the lack of outlier elimination, figure 17 shows the new threshold forecast added to the example shown in figure 16.

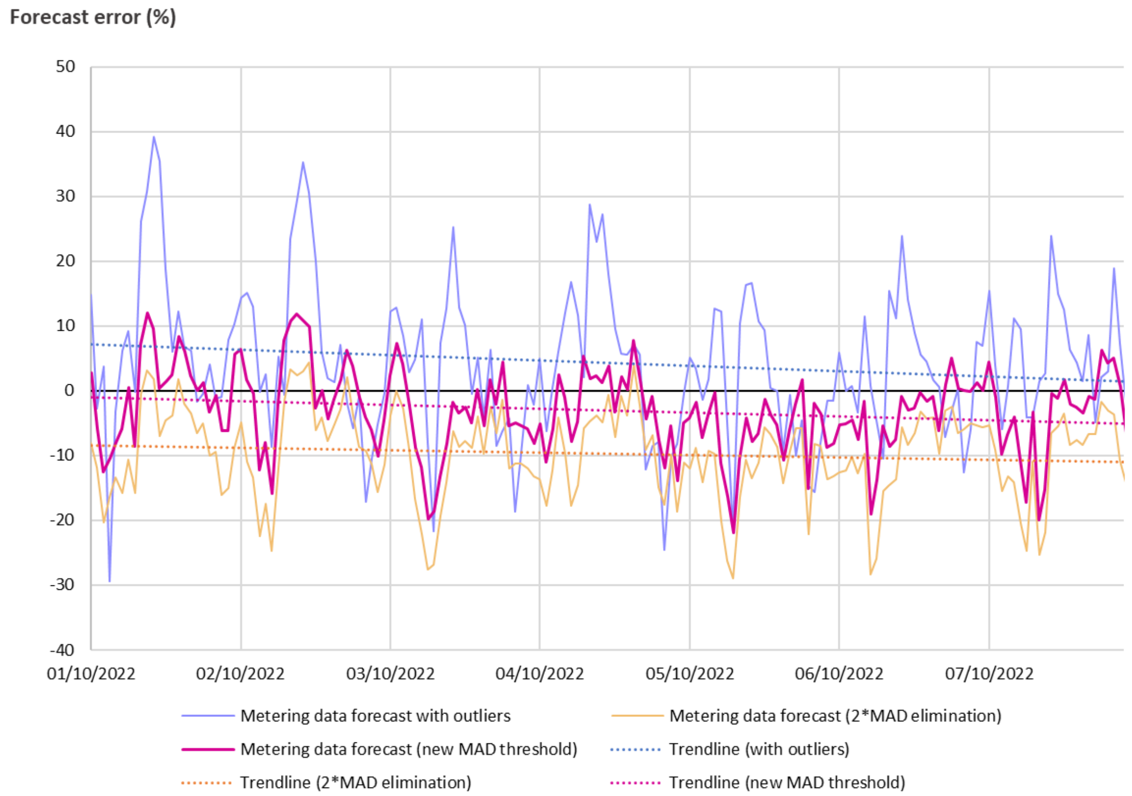

Figure 17: *The forecast relative error with the new MAD thresholds added to the example shown in figure 16*

The graph shows an improvement in both directions; especially the high error peaks of too high forecast values were largely removed. While the average error is still a tad negative, indicating that the newly forecasted values were a bit underestimated, the average error is smaller and more evenly distributed. Since this threshold combination produced best results, the final results were gathered using this modification to the initial algorithm.

To analyze the load model accuracy, *the relative error* (RE), *mean absolute error* (MAE) and *root mean squared error* (RMSE) was calculated for both the old load curves and the new load model. The errors were calculated separately for each hour of each one-week period for the three LV networks, and then combined into averages for each metric. The relative error tells how many percents the forecasted value differed from the actual value on average. The percentage is calculated from absolute differences. All average metrics are presented in table 3.

Table 3: *The average errors of load curve forecasts and metering data model forecasts, compared to actual measurements*

LV Network	Time period	Load curve RE (%)	Metering data RE (%)	Load curve MAE (kW)	Metering data MAE (kW)	Load curve RMSE (kW)	Metering data RMSE (kW)
1	July 2023	18.71	6.73	9.02	3.06	10.54	4.44
1	April 2023	21.67	8.09	11.74	4.49	13.96	6.06
1	January 2023	12.87	7.92	9.26	5.91	12.54	7.86
1	October 2022	18.50	5.57	10.76	3.54	12.55	4.92
2	July 2023	18.80	19.04	7.82	7.95	10.43	10.84
2	April 2023	20.06	16.34	10.31	7.97	12.87	10.68
2	January 2023	23.88	15.85	10.31	7.97	12.87	10.68
2	October 2022	23.93	15.85	9.20	7.31	11.82	10.02
3	July 2023	30.79	10.28	8.23	3.22	10.02	4.81
3	April 2023	25.34	12.31	8.68	4.53	10.29	5.98
3	January 2023	20.66	10.22	8.29	4.62	10.35	6.26
3	October 2022	26.47	11.23	8.98	4.44	10.94	6.29
	Avg error %	**21.81**	**11.62**	**9.39**	**5.42**	**11.60**	**7.40**

The results clearly indicate that the new load model produced significantly better forecasts than the existing load curves. In nearly all occasions, the accuracy of the metering data based forecast was better. As a hypothesis, it was suspected that the largest variance in the forecast accuracy would be caused by seasonal variance. For example, if the observed week in January would have been much colder than in previous years, the loads could have been noticeably higher than expected. However, the accuracy variance inside single LV networks is uniformly quite small. Instead, the largest factor to affect accuracy seems to be the observed LV network. With all metrics, the metering data model performed worst in LV network 2. This suggests that the historical consumption in this network was less stable than in the other networks. Conversely, the load curve forecast accuracy was not worse in LV network 2 than other networks.

Table 4 presents a summary of all the average errors for both forecast methods.

Table 4: *Summary of the forecast error metrics*

	RE (%)	MAE (kW)	RMSE (kW)
Load curve	21.81	9.39	11.60
Metering data	11.62	5.42	7.40

The relative error shows a decrease of almost 50% with the metering data model. On average, the load curve forecast was 9.39 kW off the actual value for each hour, while the corresponding error was 5.42 kW for the metering data forecast. In RMSE, the errors are initially squared, which emphasizes larger deviations. Compared to other metrics, the improvement provided by the new model was smaller when measured with RMSE. This suggests that the metering data model is slightly more prone to outliers. Regardless, the accuracy of the metering data model was still significantly better when compared to the load curves.

While the presented metrics give a decent overview of the forecast accuracy, they do not show how the errors are distributed around the actual value. If the forecasted values are consistently too small or too large, it is likely that the model has a systematic defect even if the error is small on average. Overall, the error was positive (forecasted value too high) in 31.3% of all data points for the new load model. For load curves, 51.4% of errors were positive. This result suggests that the load curves produce more evenly distributed errors, and that the metering data model produces an excessive number of low values. However, when looking at the relative error distribution, the load curve forecast relative error is 25.4% when the error is positive, and 17.4% when the error is negative. The corresponding figures are 10.2% and 11.8% for the metering data model. Therefore, despite having an almost even amount of positive and negative errors, the positive errors of the load curve forecasts are relatively larger. This aspect is more balanced with the metering data model. Taking that into account, the error distribution still suggests slight defects in the developed model.

Next, one measurement period from each network will be observed in more detail and distinctive forecast errors are analyzed. Figure 18 displays the hourly load curve forecast, metering data forecast and the actual measurements for network 1 (187 customers) for the first week of January 2023.

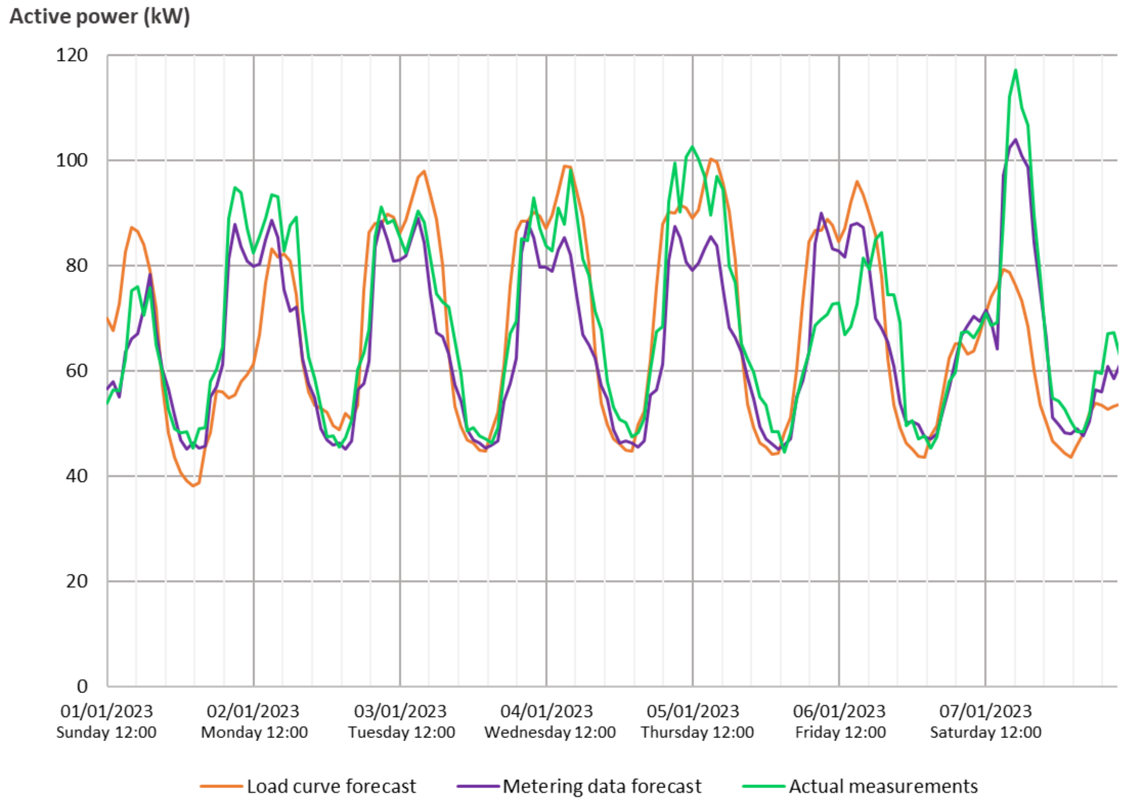

Figure 18: *Forecasts and actual measurements for network 1 in January 2023*

Compared to the stochasticity of individual customer loads shown in chapter 4, the load curve adheres to the actual consumption rather well now that a large of number of separate loads are summed up. However, the metering data forecast mimics the actual measurement curve slightly better. For this period, both forecasts were relatively accurate, as can also be seen from table 3. According to the *Finnish Meteorological Institute* (FMI), the week's end was rather cold [47], which might explain why the metering data estimates were slightly too low. However, most customers in this network have district heating, which greatly reduces the load temperature dependency. The loads of Friday are distinctively different from other weekdays; this is because Friday 6/1/2023 was Epiphany, which is a public holiday in Finland.

Figure 19 shows a similar graph for network 2 (175 customers), for the first week of April 2023:

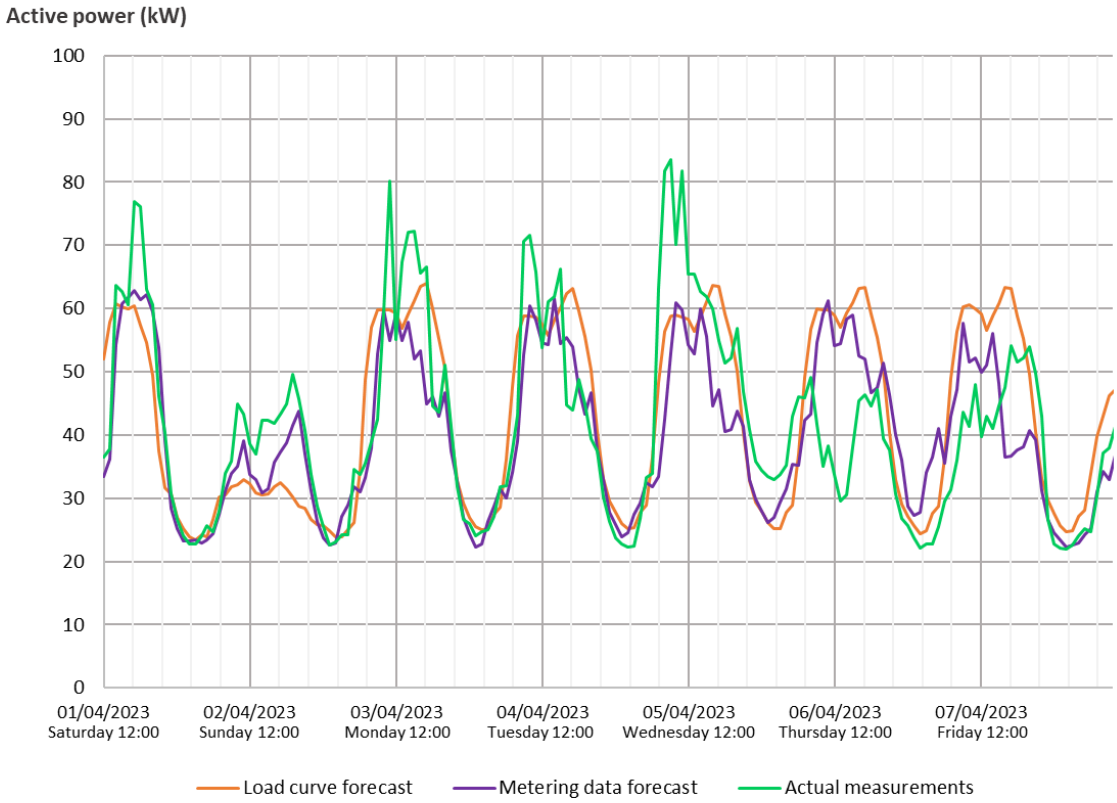

Figure 19: *Forecasts and actual measurements for network 2 in April 2023*

Compared to the previous example, both models struggle with forecasting the loads of this period, although the metering data forecast mimics the actual measurement curve's shape more closely, especially in the beginning. In 2023, Easter occurred on the first week of April, which likely explains the discrepancies towards the end of the week. Despite this, the forecasted loads are too systematically too low. While FMI reported colder night temperatures than usual during the first week of April [48], the nightly loads are quite accurately modeled before Easter and the largest differences occur at daytime. This suggests that abnormal weather conditions were not the culprit behind forecast error.

As the last example, figure 20 contains the forecasts for network 3 (166 customers), for the first week of July:

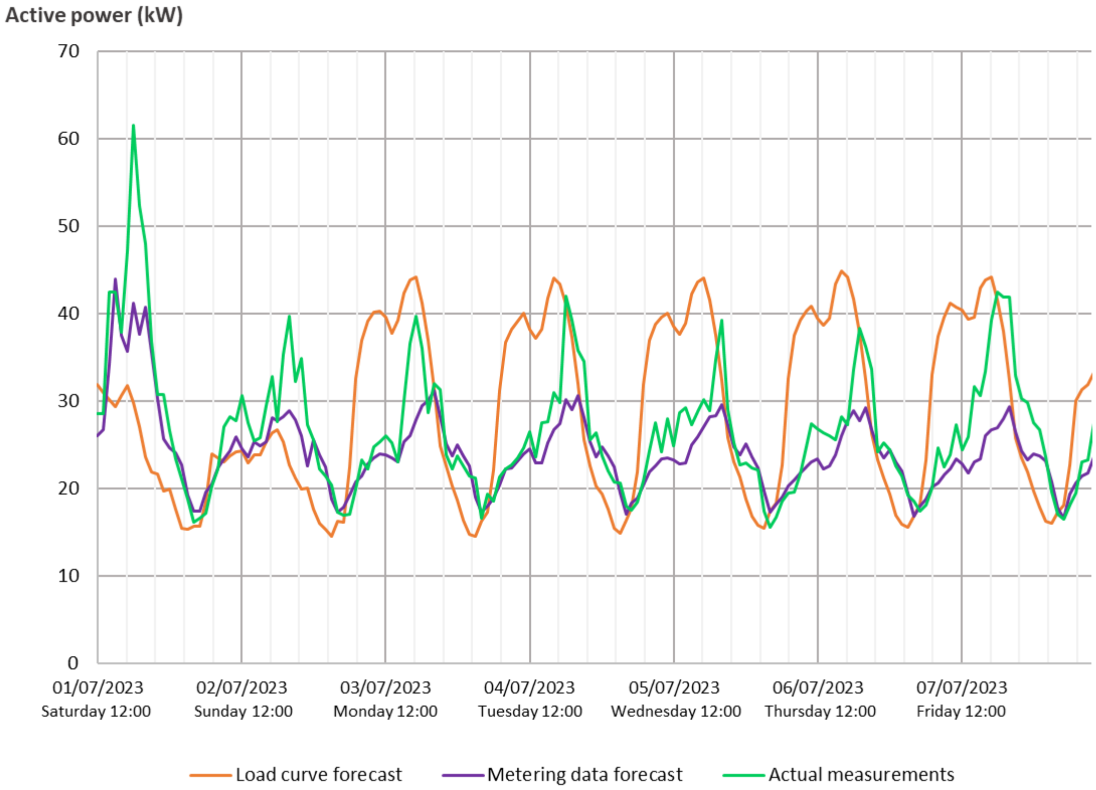

Figure 20: *Forecasts and actual measurements for network 3 in July 2023*

In this example, the load curve forecast deviates significantly from the actual measurements. With an average relative error of over 30%, this was the worst load curve forecast performance during the study. The metering data forecast quite accurately conforms to the measurement curve shape, although the forecasted values are again systematically a little low. The metering data model struggled particularly with the afternoon and evening load spikes that occurred daily. According to FMI, the first week of July was colder than usual in the whole country [49], but again, the nightly loads seem to be accurately forecasted which suggests that weather conditions were not the main source of error.

The deeper analysis of weekly forecasts showed that the largest forecast errors occur during peak loads, and in these scenarios the loads are usually underestimated. This is logical, since load peaks are very unlikely to repeat at the same exact time and with the same amplitude, so the calculation data likely includes smaller values, lowering the forecasted value. This deduction also suggests that the metering data model should overestimate some peak loads, i.e. the model should be likely to forecast higher peaks than are actually measured. While not visible in the graphs above, this was also a common source of error. This is also indicated by the earlier observation that over 30% of errors were overestimations and that the error magnitude was similar with positive and negative errors.

Overall, when short-term load forecasts were analyzed, a significant increase in load modeling accuracy was noticed with the metering data model. When the weekly forecasts were dissected, the developed model seemed to be slightly prone to underestimating the peak loads. This is problematic to some degree, since the main use case of short-term forecasts is operational planning, which specifically relies on peak loads. However, even if the peak amplitudes are slightly underestimated, the peak hours seem to be correctly identified, which is useful for operational planning.

6. FUTURE DEVELOPMENT POSSIBILITIES

The load model developed in this thesis provided significant accuracy improvements compared to the old load curves. However, the developed model is far from comprehensive; it includes simplifications and neglected factors, which were not possible to implement in the time frame of this thesis. This chapter briefly examines how the metering data based model could be improved in the future to further increase the estimation and forecast accuracy, and provide value in more areas than mere short-term load forecasting.

As mentioned earlier, the developed model does not factor in load temperature dependency. Temperature dependency is one of the most consistent and well-known factors that affect electrical loads. Weather forecasts are usually accurate to some degree for the following week, so adding the predicted temperatures to load forecast calculations could improve the accuracy of short-term forecasts. However, while the temperature dependency is rather easy to model with known equations, implementing it into the algorithm and further into the management software is not entirely straightforward. Since the metering data model is based on individual historical consumption, the temperature dependency of these past measurements also needs to be considered. Thus, obtaining the weather forecast for the upcoming week is not enough: historical weather data corresponding to the past measurements is also needed, or alternatively some temperature dependency correction needs to be embedded into the measurement data. While the implementation might be somewhat laborious, this would be a clear way of improving the model.

Another deficiency in the developed model that was already addressed earlier is that public holidays are not accounted for. The detailed load forecast graphs shown in chapter 5 demonstrate how largely a public holiday affects the load pattern of a weekday. Implementing this into the model is rather simple, as it would only require maintaining a list of the public holidays in the particular region and treating public holidays as Sundays in the model.

One factor that increasingly influences electricity consumption patterns is the price of electricity. Consumers are shifting to market-price electricity contracts with an accelerating pace, which means that more and more consumers will direct consumption to times of low electricity prices. While this phenomenon is still not very significant, it is likely to play a large role in the future. While the effect of electricity price could be feasibly implemented into the load forecast model, some further research would be needed on the relation between loads and electricity prices.

The results of this thesis showed that the developed model had a slight tendency to underestimate load peaks. This suggests that there is room for improvement in the load accuracy. The estimation algorithm used in this thesis was quite straightforward and likely a more sophisticated approach could yield better results. For example, the author suspects that the outlier elimination used in this thesis is not perfect, resulting in the loss of some of the higher load values, which could explain the struggle in peak load forecasting. Different forecast algorithms are highly studied in literature, and for example machine learning based solutions could provide more tools to mitigate the effect of load stochasticity on forecasting.

The model implementation into DMS600 only included the use of active power load measurements. While this was sufficient for conducting research on the model accuracy, factoring in all measurement types is essential for real-life distribution system operation. In addition to active power loads, the measurement files can also include active power production and reactive power measurements, as explained in chapter 4. The importation module created for this thesis is already capable of handling all measurement types so full support in DMS600 would only require some extensions to the load calculations in the program core. Taking the production measurements into account is very straightforward as the electricity production of a customer can simply be summed up with the corresponding load measurements. Meanwhile, factoring in the reactive power measurements in DMS600 is also relatively simple since the existing algorithm can also be used in reactive power calculations.

Lastly, the model in this thesis was only implemented and tested in LV networks and only for system state estimation and short-term load forecasts. The model could be developed further and extended to MV networks. In DMS600, MV network load calculations include the estimation tool that improves load model predictions by utilizing real-time load measurements mostly from feeders. Combining this estimation tool with the new load model could provide greatly improved MV network calculation and load forecasting. Also, the model could be extended to be able to calculate rational long-term forecasts. Long-term forecasts cannot rely on past week calculations, and only utilizing the past year measurements from the same hour might not be sufficient for creating accurate forecasts. Consequently, the algorithm would need some further logic to handle long-term forecasts.

7. CONCLUSIONS

Distribution system management is becoming ever more difficult due to the drastic changes required across the electricity grid to support energy transition towards more sustainable electricity production. Among other things, distributed generation and the growing number of electric vehicles and load automation complicate maintaining power balance and power quality across the grid. The related fundamental changes in electricity consumption patterns increase the stochasticity of electrical loads, making load modelling and forecasting increasingly difficult. Distribution system operators depend on accurate load models to estimate and forecast network loads in everyday network management. The existing load models are obsolete, and no longer match the present-day electricity consumption, creating a need for new load modelling solutions.

This Master's thesis studied whether the historical consumption of individual customers could be directly used in estimating customer loads and creating short-term load forecasts in low-voltage networks. A constructive research approach was taken to develop a novel load model, which was implemented into a real-life distribution management software. The software in question was the DMS600 distribution system management software package developed by Hitachi Energy. To support development and provide the necessary data for testing, smart meter measurements from the last 4 years were received from the entire network of a distribution system operator operating in Western Finland. Before developing the novel model, a necessary theoretical background was established. This included examining how electricity is measured and how loads are generally modelled and forecasted in distribution networks. The practical applications of these models were also explored. Lastly, the DMS600 software package and its load modelling functionalities were presented.

The developed algorithm is a rather straightforward statistical model that uses historical data to separately estimate customer-specific loads for each hour. In other words, if the load of an MV/LV substation is estimated, the load of each related customer is calculated separately for each hour and summed up to form an hourly load estimate for the substation. The historical data consisted of measurements from comparable hours: in practice, the loads from corresponding hours from the past three weeks, as well as the loads of the exact same hours from past 4 years were used. The algorithm first identifies and eliminates outliers using median absolute deviation, after which the remaining measurements are assigned with weights based on their recency. A weighted average is used to form the final load estimation. The outlier elimination boundaries were iteratively determined.

The algorithm was implemented into the DMS600 Workstation, which is a distribution management system that traditionally uses either Velander's formula or load profiles as a load modelling method. Velander's formula is only capable of estimating the peak loads of larger customer groups, while load profiles offer a more comprehensive solution. However, as load profiles only exist in countries where load profiles have been formed through load research, they are only available in a handful of countries. Since the implementation is based on load data provided in the standardized format used in Finland, the main contribution of the actual implementation serves Finnish operators. However, the general concept is applicable to load modelling in any country.

To study the accuracy of the newly developed load model, a test scenario was designed. The developed model and the existing load profile model of DMS600 were used to create one-week load forecasts across all seasons in three separate, relatively large urban low voltage networks. For each hour, the forecasted values by both models were compared to actual measurements to determine their accuracy. The selected error metrics were relative error, mean absolute error and root mean squared error. On average, the metering data model produced more accurate results than the load profile model by all metrics. The relative error was cut down to almost a half with the metering data model producing an average 11.6% relative error compared to an average relative error of 21.8% with the load profile model.

Despite the promising results, the developed model had some defects. The model had a slight tendency to underestimate loads, which is not optimal for operational planning. The largest errors occurred during peak loading hours. However, the model was able to correctly identify the load peaks, which is perhaps the most important factor for operational planning. Apart from a slight underestimation, the forecasted values were usually more accurate than the values forecasted by the load profile model. Due to the limited timeframe for this work, incorporating some variables with significant importance was not feasible. This included for example implementing load temperature dependency together with weather forecasts into the model, as well as factoring in public holidays occurring on weekdays. In terms of future development, the author highly recommends including these variables in possible future iterations of the load model for increased accuracy. While these factors are most influential, forecasting also involves various other variables that could be considered in future development.

Overall, the load estimation and forecasting model developed in this thesis showed promising results. This work proves that the historical measurements of individual customers can in fact be used to create customer-specific load forecasts that improve the accuracy over traditional load profiling. In addition, it was shown that this type of a model can easily be integrated into distribution management software, which is a key factor for real-life applications. However, the implemented model is still quite limited. Besides addressing the mentioned defects, expansion to estimate and forecast medium voltage network loads is highly recommended. This, in turn, raises a potential efficiency problem, as the number of customers and thus the number of necessary calculations increases heavily when shifting from low voltage to medium voltage networks. Utilizing recent smart meter measurements for continuous load estimation and forecasting is an area that is not extensively researched. The author hopes for further research on related topics along with concrete improvements to existing distribution management software.

REFERENCES

[1] Electricity consumption – Electricity Information: Overview – Analysis, IEA, website. Available (accessed 29.3.2023): https://www.iea.org/reports/electricity-information-overview/electricity-consumption

[2] G. van de Kaa, T. Fens, J. Rezaei, D. Kaynak, Z. Hatun, and A. Tsilimeni-Archangelidi, "Realizing smart meter connectivity: Analyzing the competing technologies Power line communication, mobile telephony, and radio frequency using the best worst method", Renewable & sustainable energy reviews, vol. 103, 2019.

[3] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status," Renewable & sustainable energy reviews, vol. 15, no. 6, 2011.

[4] H. Abu-Rub, M. Malinowski, and K. Al-Haddad, Power electronics for renewable energy systems, transportation, and industrial applications, 1st edition, Wiley/IEEE, 2014.

[5] European Parliament, Council of the European Union, Directive 2009/72/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC. 2009. Available (accessed 26.5.2023): https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009L0072

[6] Reserves and balancing power, Fingrid, website. Available (accessed 26.5.2023): https://www.fingrid.fi/en/electricity-market/reserves_and_balancing/

[7] Questions and answers about Datahub, Fingrid, website. Available (accessed 11.7.2023): https://www.fingrid.fi/en/electricity-market/datahub/questions-and-answers-about-datahub/

[8] Information exchange in the electricity retail market enters a new era – Datahub goes live successfully, Fingrid, website. Available (accessed 6.11.2023): https://www.fingrid.fi/en/news/news/2022/information-exchange-in-the-electricity-retail-market-enters-a-new-era--datahub-goes-live-successfully/

[9] Finland switching to 15-minute imbalance settlement, Fingrid, 2022, website. Available (accessed 26.5.2023): https://www.fingridlehti.fi/en/finland-switching-to-15-minute-imbalance-settlement/#3605e45a

[10] Sähkön mittauksen periaatteita, Energiateollisuus, 2022. Available (accessed 26.5.2023): https://energia.fi/files/7465/Sahkon_mittausohje_-_20220123_-_versio_20221101.pdf

[11] Älyverkkotyöryhmä: Seuraavan sukupolven älykkäiden sähkömittareiden vähimmäistoiminnallisuudet, Pöyry, 2017. Available (accessed 26.5.2023): https://tem.fi/documents/1410877/3481825/AMR+2.0+loppuraportti+15.12.2017/6a2df7e6-a963-40c0-b4d8-d2533fbca488/AMR+2.0+loppuraportti+15.12.2017.pdf

[12] A. Primadianto and Chan-Nan Lu, A Review on Distribution System State Estimation, IEEE transactions on power systems, vol. 32, no. 5, 2017.

[13] Heating - Fuels & Technologies, IEA, website. Available (accessed 29.5.2023): https://www.iea.org/fuels-and-technologies/heating

[14] Energiavuosi 2022 Kaukolämpö, Energiateollisuus ry, 2023. Available (accessed 26.10.2023): https://energia.fi/tilastot/energiavuosi-2022-kaukolampo/

[15] Suomen virallinen tilasto (SVT): Asumisen energiankulutus, 1. Asuinrakennusten päälämmönlähteiden kehitys 2010-luvulla, Tilastokeskus, 2018, website. Available (accessed 31.10.2023): https://www.stat.fi/til/asen/2018/asen_2018_2019-11-21_kat_001_fi.html

[16] Annual growth in sales of heat pumps in buildings worldwide and in selected markets, 2021 and 2022 – Charts – Data & Statistics, IEA, 2023. Available (accessed 2.11.2023): https://www.iea.org/data-and-statistics/charts/annual-growth-in-sales-of-heat-pumps-in-buildings-worldwide-and-in-selected-markets-2021-and-2022

[17] K. Sezen and A. Gungor, Performance analysis of air source heat pump according to outside temperature and relative humidity with mathematical modeling, Energy conversion and management, vol. 263, 2022.

[18] K. Kallioharju, M. Honkiniemi, J-P. Juvela, E. Lipsanen and S. Uusitalo, Vuoreksen olosuhde- ja energiaseurantahanke - Loppuraportti, 2015. Available (accessed 4.12.2023): https://projects.tuni.fi/uploads/2021/11/a140ece9-loppuraportti_vuores_09122015.pdf

[19] Supercharger, Tesla, website. Available (accessed 29.10.2023): https://www.tesla.com/supercharger

[20] A. G. Anastasiadis, G. P. Kondylis, A. Polyzakis, and G. Vokas, Effects of Increased Electric Vehicles into a Distribution Network, Energy procedia, vol. 157, 2019.

[21] S. Shivashankar, S. Mekhilef, H. Mokhlis, and M. Karimi, Mitigating methods of power fluctuation of photovoltaic (PV) sources – A review, Renewable & sustainable energy reviews, vol. 59, 2016.

[22] How is EU electricity produced and sold?, Council of the European Union, 2023, website. Available (accessed 31.10.2023): https://www.consilium.europa.eu/en/infographics/how-is-eu-electricity-produced-and-sold/

[23] See hourly day-ahead prices, Nord Pool, website. Available (accessed 31.10.2023): https://www.nordpoolgroup.com/en/Market-data1/Dayahead/Area-Prices/ALL1/Hourly/

[24] Pörssihintaisten sähkösopimusten osuus nousi lähes 14 prosenttiin vuonna 2022, Energiavirasto, 2023, website. Available (accessed 31.10.2023): https://energiavirasto.fi/-/porssihintaisten-sahkosopimusten-osuus-nousi-lahes-14-prosenttiin-vuonna-2022

[25] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, Load Modeling - A Review, IEEE transactions on smart grid, vol. 9, no. 6, 2018.

[26] E. Lakervi and J. Partanen, Sähkönjakelutekniikka, Otatieto, no. 609, Helsinki: Otatieto Helsinki University Press, 2008.

[27] E. Lakervi and E. J. Holmes, Electricity distribution network design, 2nd ed., IEE power engineering series, no. 21. London, 1995.

[28] M. Hemmingsson and M. Lexholm, “Dimensioning of smart power grids for the future,” Elforsk, 2013.

[29] Load Profiles and their use in Electricity Settlement, Elexon, website. Available (accessed 10/12/2023): https://bscdocs.elexon.co.uk/guidance-notes/load-profiles-and-their-use-in-electricity-settlement

[30] C. C. Aggarwal and C. K. Reddy, Data Clustering: Algorithms and Applications, 1st ed., vol. 31. in Chapman & Hall/CRC data mining and knowledge discovery series, 2014.

[31] G. Chicco, R. Napoli, and F. Piglione, Comparisons among clustering techniques for electricity customer classification, IEEE transactions on power systems, vol. 21, no. 2, 2006.

[32] A. Mutanen, Improving Electricity Distribution System State Estimation with AMR-Based Load Profiles, 2018.

[33] Ni Ding, C. Benoit, G. Foggia, Y. Besanger, and F. Wurtz, Neural Network-Based Model Design for Short-Term Load Forecast in Distribution Systems, IEEE transactions on power systems, vol. 31, no. 1, 2016.

[34] C. Tian, J. Ma, C. Zhang, and P. Zhan, A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network, Energies (Basel), vol. 11, no. 12, 2018.

[35] K. A. Clements, The impact of pseudo-measurements on state estimator accuracy, IEEE, 2011.

[36] MicroSCADA X DMS600 4.6 Operation Manual, Hitachi Energy, 2022.

[37] S. S. Reddy, Multi-Objective Based Congestion Management Using Generation Rescheduling and Load Shedding, IEEE transactions on power systems, vol. 32, no. 2, 2017.

[38] S. A. Soliman, Electrical load forecasting modeling and model construction, Burlington, 2010.

[39] F. Olivier, P. Aristidou, D. Ernst, and T. Van Cutsem, Active Management of Low-Voltage Networks for Mitigating Overvoltages Due to Photovoltaic Units, IEEE transactions on smart grid, vol. 7, no. 2, 2016.

[40] MicroSCADA X DMS600 4.6 System Overview, Hitachi Energy, 2022.

[41]MicroSCADA X DMS600 4.6, DMS600 Integration with SYS600, Hitachi Energy, 2022.

[42]Standard ASCII format (SAF) Technical Description, Enoro Oy, 2016. Available (accessed 7.11.2023): https://palvelut.datahub.fi/api/documents/file/0-232742-1-295447

[43]Mittaustietojen-SAF-aineistokuvaus kaasudatahubiin, Gasgrid Finland, Available (accessed 7.11.2023): https://gasgrid.fi/wp-content/uploads/Mittaustietojen-SAF-aineistokuvaus-Kaasudatahubiin-ID-8263.pdf

[44]Retail market services, Gasgrid Finland, website. Available (accessed 6.11.2023): https://gasgrid.fi/en/our-services/retail-market-services/

[45]J. Martínez Torres, Functional Statistics Outliers Detection and Quality Control. Basel, Switzerland: MDPI - Multidisciplinary Digital Publishing Institute, 2020.

[46]P. H. Ramsey and P. P. Ramsey, Optimal Trimming and Outlier Elimination, Journal of modern applied statistical methods, vol. 6, no. 2, 2007.

[47]Tammikuun 2023 kuukausikatsaus, Ilmatieteen laitos, 2023, website. Available (accessed 17.11.2023): https://www.ilmastokatsaus.fi/2023/02/21/tammikuun-2023-kuukausikatsaus/

[48]Huhtikuun 2023 kuukausikatsaus, Ilmatieteen laitos, 2023, website. Available (accessed 17.11.2023): https://www.ilmastokatsaus.fi/2023/05/16/huhtikuun-2023-kuukausikatsaus /

[49]Heinäkuun 2023 kuukausikatsaus, Ilmatieteen laitos, 2023, website. Available (accessed 17.11.2023): https://www.ilmastokatsaus.fi/2023/08/22/heinakuun-2023-kuukausikatsaus /