Linear simulation of large scale regional electricity distribution networks and its applications

Towards a controllable electricity network

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 4 maart 2021 om 10:00

door

Werner H.P. van Westering

Faculteit voor Werktuigbouwkunde, Maritieme Techniek & Technische Materiaalwetenschappen,
Technische Universiteit Delft, Delft, Nederland,
geboren te Haarlem, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. J. Hellendoorn
copromotor: prof. dr. ir. J.G. Slootweg

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. T. van der Hagen,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. ir. B.H.K. de Schutter,	Technische Universiteit Delft
Prof. dr. P. Palensky,	Technische Universiteit Delft
Prof. dr. G. Deconinck,	Katholieke Universiteit Leuven
Dr. ir. R.I.J. Dobbe,	Technische Universiteit Delft
Prof. dr. ir. W. de Jong,	Technische Universiteit Delft

Keywords: Electricity distribution network, energy transition, linear load flow, community battery, numerical analysis, Bayesian state estimation, Gaussian mixture models

Printed by: Drukwerknodig.nl

Front & Back: Pattern created by artist 'LauraKick'

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

Summary ix

Samenvatting xi

List of abbreviations xiii

1 Introduction 1
- 1.1 The changing world of the distribution network operator 2
- 1.2 The linear power flow problem 2
  - 1.2.1 Simulating reactive power without imaginary numbers . 5
  - 1.2.2 Modelling MV/LV transformers 6
- 1.3 Voltage and overload problems 7
- 1.4 Outline of the thesis 7
- 1.5 Thesis context 9

References 9

2 Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation 11
- 2.1 Introduction 12
- 2.2 Related work and contributions 12
- 2.3 Methodology 14
  - 2.3.1 Low voltage network model 14
  - 2.3.2 Simulating reactive power without imaginary numbers . 17
  - 2.3.3 Motivation for linear modeling 17
  - 2.3.4 Formulating the battery control problem 18
- 2.4 Experimental setup part I: The community battery of Rijsenhout 21
  - 2.4.1 Optimization results 22
  - 2.4.2 Checking the linearity assumption 24
- 2.5 Community battery design specifications 27
- 2.6 Experimental setup part II: Dimensioning the community battery of Rijsenhout 29
- 2.7 Conclusion 30

References 31

3 Linear power flow method improved with numerical analysis techniques applied to a very large network 35
- 3.1 Introduction 36
- 3.2 The large linear power flow problem (LLPF) 37
  - 3.2.1 Solving in terms of only real numbers 39

3.2.2 Modelling MV/LV transformers . . . 40
3.3 Comparison between linear and nonlinear power flow problems . . . 41
3.3.1 Comparison to the Newton power flow method . . . 41
3.3.2 Comparison to a commercial power flow software Vision . . . 42
3.4 Case study of large Dutch power grid . . . 42
3.4.1 Data and assumptions . . . 44
3.4.2 Simulation results . . . 45
3.5 Application of numerical analysis techniques on the LLPF problem . . . 46
3.5.1 LLPF problem with real components . . . 46
3.5.2 LLPF problem with complex components . . . 49
3.6 Conclusion . . . 50
References . . . 52

4 Forecasting-based state estimation for three-Phase distribution systems with limited sensing 55
4.1 Introduction . . . 56
4.1.1 Motivation . . . 56
4.1.2 Previous work . . . 57
4.2 Methodology . . . 58
4.2.1 Sources of information and uncertainty for state estimation . . . 59
4.2.2 Introduction to minimum mean square estimation . . . 61
4.3 Forecasting . . . 62
4.4 Power flow modeling . . . 64
4.4.1 Single-phase power flow . . . 64
4.4.2 Three-phase power flow . . . 65
4.5 Real-time estimation . . . 66
4.5.1 Voltage as a function of net load . . . 67
4.5.2 Voltage forecast statistics . . . 68
4.5.3 Constructing the state estimator . . . 69
4.6 Results . . . 69
4.6.1 Synthetic experiments . . . 69
4.6.2 Validation experiment on a utility testbed . . . 70
4.7 Suggestions for observability analysis and sensor placement . . . 73
4.8 Conclusion . . . 75
References . . . 75

5 Analysis of energy transition impact on the low voltage network using stochastic load and generation models 79
5.1 Introduction . . . 80
5.1.1 Related work . . . 80
5.1.2 Contributions . . . 81
5.1.3 Chapter outline . . . 82

5.2 Stochastic profile modeling . . . 82
5.2.1 Methodology . . . 82
5.2.2 Residential load . . . 83
5.2.3 Photovoltaics . . . 87
5.2.4 Electric vehicles . . . 91
5.2.5 Heat pumps . . . 93
5.3 Evaluating a large low voltage electricity network . . . 95
5.4 Case study: Congestion in a large real-world low voltage power grid . . . 98
5.4.1 General approach and scenario description . . . 99
5.4.2 Simulation results . . . 99
5.5 Conclusion . . . 100
References . . . 103

6 Conclusion **107**
6.1 Improvements over literature . . . 108
6.2 Further applications of the presented research . . . 109
6.3 Suggestions for future research . . . 111
6.4 Final remarks . . . 112
References . . . 112

Acknowledgements **115**

A Reference networks **117**
References . . . 117

Curriculum Vitæ **119**

List of Publications **121**

Summary

The volatility of renewable energy sources pose a significant challenge for Distribution Network Operators (DNOs) as it makes planning and maintaining a reliable electricity grid more complex. An essential tool in dealing with the uncertain behavior of renewable energy resources is the load flow simulation, i.e., the standard electricity network simulation in network design and operation. There is, however, still much untapped potential of applying these kind of simulations.

The thesis presents improvements to the theory on linear load flow approximations. The resulting algorithms are then applied to various real world problems: control of a community battery, handling very large simulations, coping with low sensor coverage and evaluating strategic scenario's with high uncertainty.

Firstly, theory is presented for the control of a community battery. It is shown how such a battery can be used for grid congestion reduction, backed up by a live experiment. A charge path optimization problem is posed as a linear problem and subsequently solved by an Linear Programming (LP) algorithm. It was found that the voltages and currents can be controlled to a great degree, increasing the grid capacity significantly. Network design formulas are described with which a DNO can quickly estimate the potential (de)stabilizing effect caused by a community battery on the steady-state voltages and currents in the grid.

Next, load flow simulations are improved by applying numerical analysis techniques and the accuracy and efficiency of a linear load flow approach is investigated. The resulting fast load flow algorithm is then applied to a very large problem: integrally simulating the low and medium voltage network of Alliander DNO, a grid with over 22 million cable segments with a total combined length of over 88,000 km, built according to international standards. It is shown that this integral simulation can identify voltage problems much more accurately.

Next, Bayesian state estimation is considered. A mathematical model is proposed to complement a limited set of real-time measurements with voltage predictions from forecast models. This method relies on Bayesian estimation formulated as a linear least squares estimation problem. The model is then applied to an IEEE benchmark and on a real network test bed. An observability analysis suggests strategies for optimal sensor placement.

Next, theory is presented on coping with uncertain long-term scenarios for strategic simulations. A stochastic profile model is proposed based on copulas which can be calibrated by technology adoption data. Using a Monte Carlo approach, the stochastic profiles of all DNO assets are then simulated, identifying parts of the network with heavy loads.

Finally, the thesis concludes by demonstrating additional applications of the presented methods, such as fast network capacity checks and reducing losses via network reconfiguration. It concludes by giving suggestions for future research.

Samenvatting

Het wereldwijde energielandschap zal de komende decennia aanzienlijk veranderen door de overschakeling op duurzame energie. Dit is een grote uitdaging voor netbeheerders die verantwoordelijk zijn voor het onderhouden van een betrouwbaar en betaalbaar distributienetwerk voor elektriciteit. Een essentieel hulpmiddel bij het omgaan met de onzekerheid van hernieuwbare energie is de loadflow-simulatie, d.w.z. de standaard simulatie gebruikt bij netwerkontwerp. Dit proefschrift heeft als doel de theorie over lineaire benadering van de loadflow te verbeteren en past de resultaten vervolgens toe op diverse echte problemen: Aansturing van een buurtbatterij, omgaan met zeer grote simulaties, omgaan met lage sensordekking en het evalueren van strategische scenario's met hoge mate van onzekerheid.

Het hoofdstuk over buurtbatterij-aansturing biedt de nodige theorie om een buurtbatterij te gebruiken om netcongestie op te lossen, onderbouwd met experimentele resultaten. Het optimalisatieprobleem voor het aansturen van de batterij wordt opgelost door Linear Programming. Het blijkt dat de spanningen en stromen in hoge mate kunnen worden geregeld, waardoor de netcapaciteit aanzienlijk wordt verhoogd. Er worden formules beschreven waarmee een netbeheerder snel de potentiële (de)stabilisatie van een netwerk door een buurtbatterij in kan schatten.

Het hoofdstuk over het verbeteren van loadflow-simulaties onderzoekt de nauwkeurigheid en efficiëntie van een lineaire loadflowbenadering door middel van numerieke analysetechnieken. Het resulterende snelle load flow-algoritme wordt vervolgens toegepast op een zeer groot probleem: het integraal simuleren van het laag- en middenspanningsnetwerk van Alliander. Dit netwerk bestaat uit 22 miljoen kabelsegmenten met een totale lengte meer dan 88.000 km. Vervolgens wordt getoond dat deze simulatie de spanningsproblemen nauwkeuriger kan vinden.

Het hoofdstuk over toestandsschatting stelt een methode voor om een beperkte reeks realtime metingen aan te vullen met voorspelde spanningen. De methode is gebaseerd op Bayesiaanse schatting geformuleerd als een lineair kleinstekwadratenprobleem. De methode wordt toegepast op een IEEE-benchmark en op een echt testbed bij Alliander. Een analyse van de sensordekking suggereert strategieën voor optimale sensorplaatsing.

Het hoofdstuk over strategische simulaties richt zich op het omgaan met onzekere strategische scenario's. Een stochastisch profielmodel wordt voorgesteld op basis van copula's die kunnen worden gekalibreerd met behulp van gegevens over technologische adoptie. Door een Monte Carlo-benadering worden vervolgens de stochastische profielen van alle elektriciteitskabels gesimuleerd.

Het proefschrift concludeert met het demonstreren van aanvullende toepassingen van de gepresenteerde methodes, zoals snelle netwerkcapaciteitscontroles en het verminderen van verliezen via aanpassingen in de netwerkconfiguratie. Tot slot worden nog enkele kansen voor vervolgonderzoek beschreven.

List of abbreviations

The following abbreviations are used in this manuscript:

AMD	Approximate Minimum Degree
AMI	Advanced Metering Infrastructure
ARMA	Auto-Regressive Moving Average
ARMSE	Average Root Mean Square Error
BESS	Battery Energy Storage Systems
BFS	Backward-Forward Sweep
BiCGSTAB	Bi-Conjugate Gradient Stabilized
CEA	Climate and Energy Agreement
CDF	Cumulative Density Function
CRPS	Continuous Ranked Probability Score
CG	Conjugate Gradient
DC	Direct Current
DG	Distributed Generation
DNO	Distribution Network Operator
DNRP	Distribution Network Recon-figuration Problem
EV	Electric Vehicle
FDLF	Fast Decoupled Load Flow
GP	Gaussian Process
GMM	Gaussian Mixture Model
GMRES	Generalized Minimal RESidual method
G-S	Gauss Seidel
HP	Heat Pump
IC	Incomplete Cholesky
ILU	Incomplete LU decomposition
KVL	Kirchhoff's Voltage Laws
LLPF	Large Linear Power Flow
LP	Linear Programming
LPF	Linear Power Flow
LLSE	Linear Least Squares Estimation
LU	Lower and Upper triangular matrix decomposition
LV	Low Voltage
MAPE	Mean Absolute Percentage Error
MPC	Model Predictive Control
MMSE	Minimum Mean Squares Estimation

MV	Medium Voltage
NA	Numerical Analysis
NNZ	Number of Non-Zeros
NPF	Nonlinear Power Flow
N-R	Newton power flow
PCG	Preconditioned Conjugate Gradient
PLC	Programmable Logical Circuit
PV	Photo-Voltaic power generation (solar power)
PMU	Phasor Measurement Unit
RCM	Reverse Cuthill-McKee
RES	Renewable Energy sources
RHS	Right Hand Side (of equation)
RMSPE	Root Mean Square Percentage Error
SCADA	Supervisory Control And Data Acquisition
SE	State Estimation
SPD	Symmetric and Positive Definite
WLS	Weighted Least Squares
ZI	Combination of constant impedance Z and constant current I load models
ZIP	Combination of ZI and constant power P load models

1

Introduction

This chapter describes the context of this thesis, namely the increase in electricity consumption due to the energy transition and its effects on distribution networks. The chapter provides the situation in The Netherlands as an example, using data of the Distribution Network Operator (DNO) Alliander NV. It proceeds to describe the basic concepts of modeling an electricity network; network structure, basic formulas and the linear power flow problem. The chapter concludes by giving the general structure of the thesis and summarizes how the papers are interconnected.

1

1.1. The changing world of the distribution network operator

The volatility of renewable energy sources pose a significant challenge for Distribution Network Operators (DNOs), as it makes planning and maintaining a reliable and affordable electricity distribution grid more complex. Especially the rise of residential solar power is challenging, as solar installations can cause local voltage problems which can be cost- and labour-intensive to solve.

An essential tool in dealing with the uncertainty of renewable energy resources is the load flow simulation, i.e. the standard electricity network simulation in network design and operation. Load flows are used to answer many of today's DNO questions, such as assessing the impact of increased network loads or finding the best place to connect a new customer.

There is however still much untapped potential of applying these kind of simulations. This thesis therefore aims to improve the the theory on linear loadflow approximations and then applies the results to various real world problems: Control of a community battery, handling very large simulations, coping with low sensor coverage and evaluating strategic scenario's with high uncertainty.

To illustrate the practical potential of the theory presented in this thesis, the theory was applied to the network of Alliander DNO. Alliander DNO is the largest DNO of the Netherlands and operates the Low Voltage (LV) and Medium Voltage (MV) power grid. Subject of this thesis are both the MV and LV network, which consist of approximately 80,000 km of underground cable serving over 3 million customers. It covers over 1/3rd of the total Dutch power grid and is divided in 22 million cable segments. The low voltage network operates on 230V/400V and will be modeled as a single phase balanced network. A schematic overview of a general DNO network is displayed in Figure 1.1.

This chapter consists of a brief introduction in modeling electrical distribution networks. Then it presents the general outline of this thesis.

1.2. The linear power flow problem

For monitoring overheating due to large currents and meeting voltage regulations, it is generally sufficient to model on a time scale of several minutes. The standard way to model such an electricity grid on this time scale is the load flow model [1, 2].

This section explains how the Linear Power Flow (LPF) problem is obtained using a constant impedance load model assuming a balanced network load. In this thesis, most load flow problems are assumed to be single-phase and balanced. The main reason for this assumption is the lack of data regarding phase connections and phase imbalance within Alliander DNO.

The LV and MV network are both modeled using the same physical model. The electricity network can be modeled as a graph $G(\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ represents the network buses (nodes) and $\mathcal{E}$ the network cables (edges). The relation between the nodal currents $I_{\mathcal{N}}$ and voltages $V_{\mathcal{N}}$ can be defined by Kirchoff's Current Law:

$$I_{\mathcal{N}} = YV_{\mathcal{N}}, \tag{1.1}$$

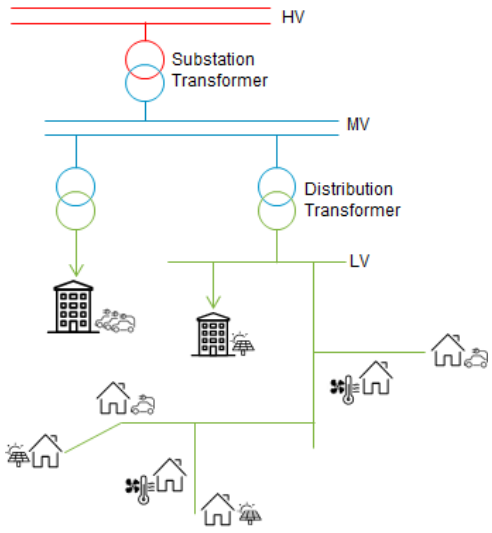


Figure 1.1: A schematic overview of the voltage levels of a DNO network. Alliander operates the Low Voltage (LV) and Medium Voltage (MV) networks. These networks operate on 400V/230V and 10kV respectively. The High Voltage (HV) network is not operated by Alliander DNO.

where Y is the so-called nodal admittance matrix. In (1.1), all variables are given in complex numbers as $V_{\mathcal{N}} = |V_{\mathcal{N}}|e^{\imath\phi} = V_{\mathcal{N}}^r + \imath V_{\mathcal{N}}^m$, $Y = G + \imath B$, and $I_{\mathcal{N}} = I_{\mathcal{N}}^r + \imath I_{\mathcal{N}}^m$. The admittance matrix can be directly obtained from the network lay-out using the following formula [3]:

$$Y = CZ_{\mathcal{E}}^{-1}C', \tag{1.2}$$

where C is a directional connection (branch-node incidence) matrix and $Z_{\mathcal{E}}$ is a square matrix with the corresponding impedance of each cable/edge ($\mathcal{E}$) on its main diagonal.

Generally, the load of the network is modeled as a combination of constant power, constant impedance, and constant current [2]. Even a balanced, single phase load flow problem is generally nonlinear, if a constant power load flow model is used. Adding constant power equations (in the form of $S = UI$ where S is the apparent customer power) to the system of (2.2) makes the problem non-linear. This in turn makes solving the necessary equations computationally expensive.

In this thesis, therefore only the constant impedance load model is used to keep the problem linear. All buses with customers are modeled as a constant impedance connected artificial ground buses as can be seen in Figure 1.2. Customer load is modeled by equivalent resistance R_{ng}^{eq} and reactance X_{ng}^{eq} which are defined by the

1

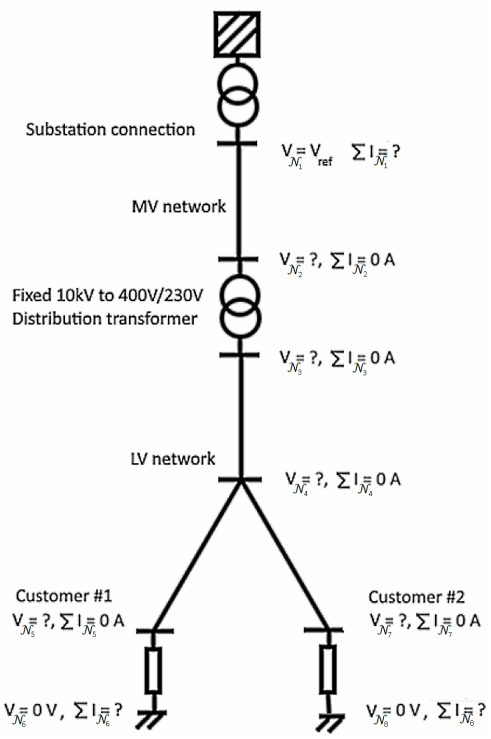


Figure 1.2: An example of a small MV/LV network connected to a substation transformer. While the MV network is always operated radially, the LV network is operated meshed in some areas. The reference voltage (V_{ref}) is defined at the secondary side of the substation transformer. If only the LV network is considered, the reference voltage (V_{ref}) is defined at the secondary side of the distribution transformer.

following formulae:

$$R_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2 P_n}{P_n^2 + Q_n^2} \tag{1.3}$$

$$X_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2 Q_n}{P_n^2 + Q_n^2} \tag{1.4}$$

where P_n and Q_n are active and reactive power consumption respectively of load bus n and V_{ref} is the nominal voltage. This assumption introduces a difference in outcome when compared to a non-linear load flow. The impact of this assumption is discussed extensively in Section 2.4.2 and is shown to be acceptable with voltage drops less than 10%.

Note that (1.1) cannot be solved directly, because not all elements are known in neither vector $I_{\mathcal{N}}$ and $V_{\mathcal{N}}$. To overcome this problem, we segment the problem in two equations which can be solved separately:

$$I_{\mathcal{N}} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, Y = \begin{bmatrix} Y_{11} & Y_{21}^T \\ Y_{21} & Y_{22} \end{bmatrix}, V_{\mathcal{N}} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{1.5}$$

We sort the rows of the matrices $I_{\mathcal{N}}$, Y, and $V_{\mathcal{N}}$ in such a way that all swing buses and ground buses are placed in V_1 and all unknown voltages of remaining buses are placed in V_2. In V_1, the voltage of the swing buses is set to the nominal level and ground buses are set to ground voltage level. Since the network is modeled as a set of voltage sources and resistances, Kirchoff's law dictates that $\Sigma I = 0$ on every bus in V_2. Therefore, I_2 is equal to 0 and the power flow equation now becomes:

$$\begin{bmatrix} I_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{21}^T \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{1.6}$$

Since V_1 is known, V_2 can be solved using the second row of (1.6) as:

$$Y_{22} V_2 = -Y_{21} V_1 \tag{1.7}$$

We now define a helper variable b, which will be of use later, and we solve for V_2:

$$b = -Y_{21} V_1, \tag{1.8}$$
$$Y_{22} V_2 = b, \tag{1.9}$$
$$V_2 = Y_{22}^{-1} b. \tag{1.10}$$

Then, we compute I_1 as:

$$I_1 = Y_{11} V_1 + Y_{21}^T V_2. \tag{1.11}$$

Finally, after computing the nodal voltages, the cable currents can be directly calculated by:

$$I_{\mathcal{E}} = Z_{\mathcal{E}} C' V_{\mathcal{N}}. \tag{1.12}$$

1.2.1. Simulating reactive power without imaginary numbers

It is possible to solve the equation in terms of only real numbers in order to ease the calculation or if the applied software does not support the combination of the complex variables and sparse matrices like the *R* programming language [4].

Neglecting imaginary parts

In LV networks customers use or produce very little reactive power on average. Additionally, due to the fact that the reactive power and reactance are generally an order lower than the active power and resistance, we can neglect the impact of reactive currents, reactive powers and cable reactance. Thus, the equivalent reactance X_{ng}^{eq} is removed and the equivalent resistance R_{ng}^{eq} becomes:

$$R_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2}{P_n}. \tag{1.13}$$

Furthermore, we know that $Y_{22} = G_{22} + \imath B_{22}$ and $b = b^r + \imath b^m$ in (1.10). By neglecting all imaginary parts from (1.10), we obtain the following linear equation:

$$|V_2| = G_{22}^{-1} b^r, \tag{1.14}$$

where $|V_2|$ is voltage magnitudes, G_{22} is the conductance of admittance matrix Y_{22} and $b^r = -G_{21}|V_1|$. This makes the power flow computation roughly 50% faster and is in some situations worth the modeling error introduced by this assumption. However, in low voltage and medium voltage networks this assumption is generally not valid.

Reformulating equations with complex numbers

Matrix equation (1.10) is given as:

$$V_2^r + \imath V_2^m = (G_{22} + \imath B_{22})^{-1}(b^r + \imath b^m). \tag{1.15}$$

Equation (1.15) can be reformulated into into the following matrix equation:

$$\begin{bmatrix} V_2^r \\ V_2^m \end{bmatrix} = \begin{bmatrix} G_{22} & -B_{22} \\ B_{22} & G_{22} \end{bmatrix}^{-1} \begin{bmatrix} b^r \\ b^m \end{bmatrix}. \tag{1.16}$$

After the computation, V_2 is calculated as $V_2 = V_2^r + \imath V_2^m$ using the computed real V_2^r and imaginary parts V_2^m in (1.16). In this case, we double the size of the equations but can avoid the complex numbering.

1.2.2. Modelling MV/LV transformers

To solve the entire network in a single simulation, the MV/LV transformers are modeled as an RL network [5] as displayed in Figure 1.3. Transformers are added to the impedance matrix $Z_\mathcal{E}$, as the link between the MV and LV network. In this

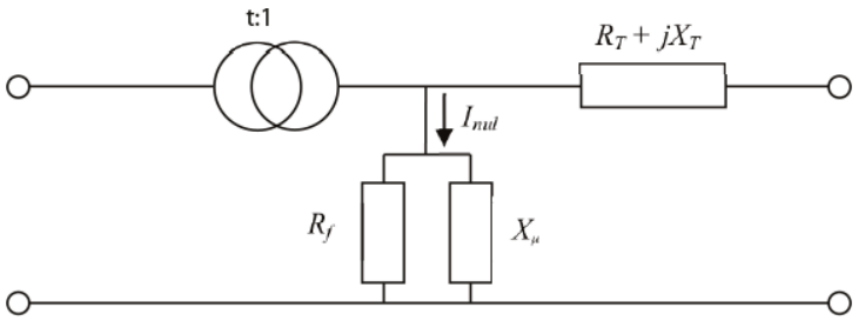


Figure 1.3: A schematic representation of the transformer model. The transformer is modeled as an RL network [5]. $R_T + jX_T$ denote the transformer's line impedance and R_f and X_μ denote the transformer's line-to-ground impedance.

model, an $\mathcal{E} \times \mathcal{E}$ diagonal scale matrix T is defined using the turns ratio t. For every cable behind the secondary side of the transformer, the corresponding value in T is the turns ratio of that transformer. For every cable in the medium voltage network, the corresponding value in the matrix T is 1. The impedance of the link is then scaled using $Z_p = T^2 Z_s$ where p, s denote the primary and secondary side of the transformer respectively. After the calculation, the turns ratio is used to re-scale the voltage and current to the LV-regime using $V_s = T^{-1} V_p$ and $I_s = T I_p$.

1.3. Voltage and overload problems

This thesis deals mainly with two types of problems which can occur in the electricity network, namely voltage problems and overload problems. While both of the problems are caused by high network load, their definitions are different. *Overload problems are defined as a cable current which exceeds the rated current of the cable.*

A voltage problem is defined as a voltage drop which exceeds the allowed voltage range. European law allows for a voltage which differs up to 10% from the nominal customer voltage. For example, a nominal single phase voltage of 230V (the European standard voltage in the low voltage network) allows for a customer voltage between 207V and 253V. However, the voltage problems in this thesis are defined according to the internal policy standard of Alliander DNO which has additional margin. *A voltage problem within Alliander DNO policy is defined as a voltage which differs no more than 9% of the nominal voltage.*

Furthermore, Alliander DNO policy dictates that on each 'voltage level' (i.e. MV and LV) the voltage drop is not allowed to be more than 4.5%. The historical reason for this decision is that the MV and LV network levels are designed separately. The impact of this design approach is discussed in Chapter 3.4.

1.4. Outline of the thesis

In this thesis, power flow simulation methods are applied to large scale problems in an attempt to solve various practical DNO problems. These problems include: real time control of a community battery, improving simulation speed for network design, Bayesian state estimation and strategic stochastic scenario evaluation.

Chapter 2

By installing a battery storage system in the power grid, Distribution Network Operators (DNOs) can solve congestion problems caused by decentralized renewable generation. Chapter 2 provides the necessary theory to use a community battery for grid congestion reduction, backed up by experimental results. The algorithms are applied and validated on a real-world community battery installation. The proposed control framework can be used to safeguard network constraints and is compatible with other battery control goals, such as energy trading or energy independence. Network design formulas are described with which a DNO can quickly estimate the potential (de)stabilizating impact of a community battery on the steady-state voltages and currents in the grid.

This chapter is based on the following publication: **W. van Westering and J. Hellendoorn**, *Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation*, International Journal of Electrical Power & Energy Systems **114**, 105349 (2019).

Chapter 3

In chapter 3 a fast linear power flow method is proposed that uses a constant impedance load model to simulate both the entire Low Voltage (LV) and Medium

Voltage (MV) networks in a single simulation. Accuracy and efficiency of this linear approach are validated by comparing it with the Newton power flow algorithm and a commercial network design tool Vision on various distribution networks including real network data.

This chapter is based on the following publication: **B. Sereeter, W. van Westering, C. Vuik and C. Witteveen**, *Linear Power Flow Method Improved With Numerical Analysis Techniques Applied to a Very Large Network*, Energies **12**, 4078 (2019), where I contributed to by co-authoring the numerical analysis simulation code, creating the comparison between integral MV/LV and LV simulations and performing the case study.

Chapter 4

State Estimation is an essential technique to provide observability in power systems. Chapter 4 proposes a method to complement a limited set of real-time measurements with voltage predictions from forecast models. The method differs from the classical weighted least-squares approach, and instead relies on Bayesian estimation formulated as a linear least squares estimation problem. A recently developed linear model for unbalanced 3-phase power flow is used to construct voltage predictions as a linear mapping of load predictions. The method is applied to an IEEE benchmark and on a real network test bed at the Dutch utility Alliander and an observability analysis suggests strategies for optimal sensor placement.

This chapter is based on the following publication: **R. Dobbe, W. van Westering, S. Liu, D. Arnold, D. Callaway and C. Tomlin**, *Linear Single- and Three-Phase Voltage Forecasting and Bayesian State Estimation with Limited Sensing*, IEEE Transactions on Power Systems (2020), where I contributed by applying the theory, setting up the Buurtbatterij experiment and helping performing the case study.

Chapter 5

The energy transition leads to an increased uncertainty in the development of future electric load profiles. To more accurately take into account these uncertainties in network planning and design, it is necessary to accurately model the possible impact of these developments on a large scale for many different scenarios. Chapter 5 proposes a modeling and assessment methodology which incorporates these needs. The resulting profiles are subsequently used in a fast linear network model, which is suitable to evaluate very large low voltage networks.

This chapter is based on the following publication: **R. Bernards, W. van Westering, J. Morrena, H. Slootweg**, *Analysis of Energy Transition Impact on the Low Voltage Network using Stochastic Load and Generation Models*, Energies (2020), Volume 13, where I contributed by expanding the network impact theory, obtaining the relevant data and performing the case study.

Chapter 6

Chapter 6 summarizes the improvements over existing methods found in literature. Furthermore, after presenting a brief overview of future research opportunities, this chapter considers multiple applications of the theory considered in this thesis.

1.5. Thesis context

This thesis was creating as a part-time project during my work within Alliander DNO. It is therefore focused on solving many day-to-day problems using a scientific approach. Because of the part-time nature of my PhD project, the estimated project length was six years. The project was formally started in February 2015 and nears its completion in March 2021, according to schedule.

References

[1] H. Le Nguyen, *Newton-raphson method in complex form [power system load flow analysis],* IEEE Transactions on Power Systems **12**, 1355 (1997).

[2] W. H. Kersting, *Distribution system modeling and analysis* (CRC press, 2001).

[3] J. Kirtley, *6.061 Introduction to Power Systems Class Notes Chapter 5 Introduction To Load Flow* (MIT Open Courseware, 2018).

[4] W. van Westering, B. Droste, and H. Hellendoorn, *Combined medium voltage and low voltage simulation to accurately determine the location of voltage problems in large grids,* in *2019 CIRED* (2019) p. 2.

[5] P. van Oirsouw, *Netten voor distributie van elektriciteit*, 2nd ed. (Phase to Phase B.V., 2011) Chap. 8.

2

Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation

By installing a battery storage system in the power grid, Distribution Network Operators (DNOs) can solve congestion problems caused by decentralized renewable generation. This chapter provides the necessary theory to use such a community battery for grid congestion reduction, backed up by experimental results. A simple network model was constructed by linearizing the load flow equations using a constant impedance load model. Using this model, an accurate estimate of voltage and overload problems is fed into a receding horizon charge path optimizer. The charge path optimization problem is posed as a linear problem and subsequently solved by an LP solver. The algorithms have been applied and validated on a real-world community battery installation. It was found that the voltages and currents can be controlled to a great degree, increasing the grid capacity significantly. The proposed control framework can be used to safeguard network constraints and is compatible with other battery control goals, such as energy trading or energy independence. Network design formulas are described with which a DNO can quickly estimate the potential (de)stabilizing effect of a community battery on the steady-state voltages and currents in the grid.

2.1. Introduction

The volatility of renewable energy sources pose a significant challenge for Distribution Network Operators (DNOs), as it makes planning and maintaining a reliable and affordable electricity distribution grid more complex. Especially the rise of residential solar power is challenging, as these installations can cause local voltage problems which can be cost intensive to solve.

A potential solution to this problem is congestion control using energy storage. By locally storing the energy generated by the solar power installations, the voltage and current in the low voltage network can be kept within the desired limits. The most common version of this solution is a home battery system. However, it is more efficient to use a community battery since a set of home batteries may adversely influence each other if they do not operate on a common objective [1, 2]. A community battery also requires less space and can be serviced more efficiently.

However, beside regulatory challenges, DNOs do generally not have the knowledge to design and employ a community battery, which results in both newly planned and currently installed storage capacity not being used for congestion control. This chapter provides the necessary theory to solve this problem, backed up by experimental results. With the principles developed in this chapter, a DNO can quickly estimate the potential (de)stabilization caused by a community battery on the steady-state voltages and currents in the grid. The control framework provided can be used to safeguard network constraints and is compatible with other battery control goals, such as energy trading or energy independence.

For experimentation purposes Alliander, the largest DNO of the Netherlands serving over three million customers, placed a community battery in Rijsenhout, a suburban village close to Amsterdam, the Netherlands. A schematic overview of the network of Alliander DNO is displayed in Figure 2.1. The battery is connected to the low voltage network and has a peak power of 55kW and a capacity of 126 kWh. The main goal of placing the battery was the broad goal of obtaining practical knowledge how a community battery can benefit the DNO.

This chapter reports on various aspects of DNO community battery utilization. It contains control strategies for using a community battery for LV network congestion management. It is the first study to combine a battery control system with a load flow model operating in real time. It also analyses the battery's (de)stabilization potential and provides design guidelines for new community batteries.

2.2. Related work and contributions

Using batteries in addition to a regular connection to the power grid is a relatively new phenomenon in Western Europe, because grid connections are reliable and relatively cheap. However, with the rise of decentralized renewable power generation, grid investment costs are expected to rise sharply [3–5]. This is a strong motivation for the DNOs to explore innovative solutions, such as battery storage.

Many studies try to find interesting new business cases for batteries. Using

Parts of this chapter have been published in International Journal of Electrical Power & Energy Systems **114**, 105349 (2019).

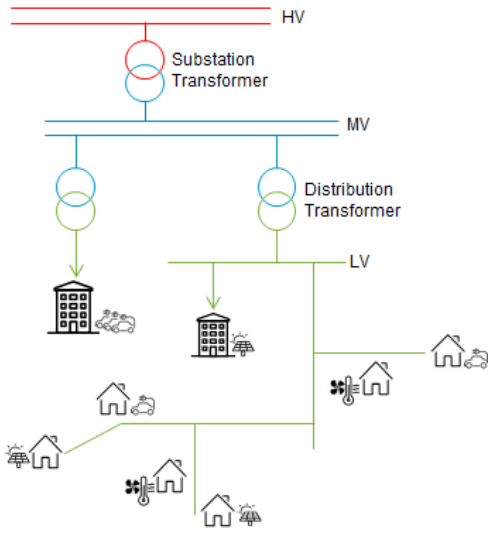


Figure 2.1: A schematic overview of the voltage levels of the network of an electricity distribution network. Alliander operates the Low Voltage (LV) and Medium Voltage (MV) networks. These networks operate on 400V/230V and 10kV respectively. The High Voltage (HV) network is not operated by Alliander DNO. The LV network is the main subject of this study.

electric vehicles for electricity storage purposes is not yet feasible [6]. Most work regarding Battery Energy Storage Systems (BESS) focuses on residential applications [7–11]. However, Parra [1, 2] calculates that a community battery is 56% cheaper than separate residential batteries for a 100-home community.

The field of battery charge path optimization has been studied in the literature quite extensively. For example, the battery scheduling problem for microgrid operation is investigated by [12–14]. Some of the studies also assume a connection to a larger grid [12] or mainly focus on the optimal size of a battery [15, 16].

Various receding horizon controllers for battery charge path optimization have been developed [17, 18]. Recently, battery controllers using Model Predictive Control (MPC) have been proposed [7, 12]. These controllers can be deployed both centralized or decentralized. However, these controllers generally do not guarantee a stable grid operation as the currents and voltages in the grid are not taken into account. This is also due to the fact that the non-linear load flow equations cannot be directly applied in a quadratic MPC controller. Furthermore, these MPC controllers have yet to be tested in a real world test bed. This chapter contributes to the literature by proposing and validating a network model which can be directly integrated into an MPC controller.

This is the first study that enhances the charge path optimization problem by adding a load flow model operating in real time, making it more interesting for a real world application. Most of the current algorithms work with predefined congestion points, i.e. network nodes which are expected to be most vulnerable to capacity or voltage problems. However, in practice the points are often hard to clearly define

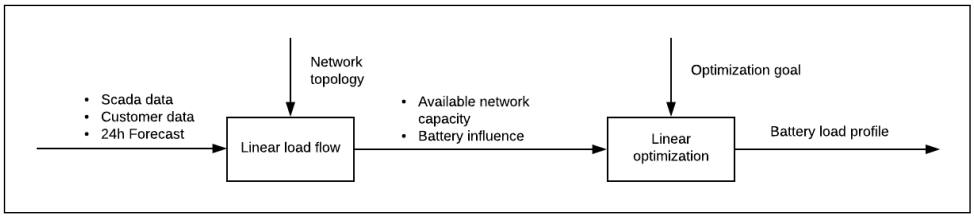


Figure 2.2: Schematic view of the Community Battery controller

for large networks in which the loads are constantly changing. Therefore adding a network model which is able to monitor all network nodes and lines simultaneously is a valuable addition to the current literature.

Furthermore, not much literature is available on applying and validating proposed algorithms on real world batteries. The community battery subject of this study is only the second ever in the Netherlands. The first one was placed by the Dutch DNO Enexis and has been used for validating a charge path optimization algorithm, reducing network losses and reducing transformer peak load [19–21]. However, since the battery was located next to the MV/LV transformer, the ability to influence the LV network was very limited in contrary to the community battery in this study.

Most DNOs have design rules regarding LV network design, but do not have policies available regarding electricity storage as it is a relatively new phenomenon in MV/LV grids. This chapter contributes by both providing a battery controller and describing community battery network design guidelines, specifically aimed at DNOs.

2.3. Methodology

Figure 2.2 contains a schematic of the battery controller. The remainder of this chapter is structured as follows: To calculate the characteristics of the LV network, a linear low voltage network model is constructed in Section 2.3.1 and 2.3.2. After the linearization is motivated in Section 2.3.3, the battery control problem is formulated in Section 2.3.4. The models are applied to the community battery of Rijsenhout and the results and accuracy of these models is investigated in Section 2.4. The results of the experiments are used to formulate battery design principles in Section 2.5 and are again applied in Section 2.6.

2.3.1. Low voltage network model

For monitoring overheating due to large currents and meeting voltage regulations, it is generally sufficient to model on a time scale of several minutes. The standard way to model such an electricity grid on this time scale is the load flow model [22, 23]. A load flow problem is generally nonlinear, due to its power constraints. This makes solving the necessary equations computationally expensive.

The standard approach for modeling DNO power grids is formulating a load flow problem and solving it using a Newton-Raphson methodology [22, 23]. Usually

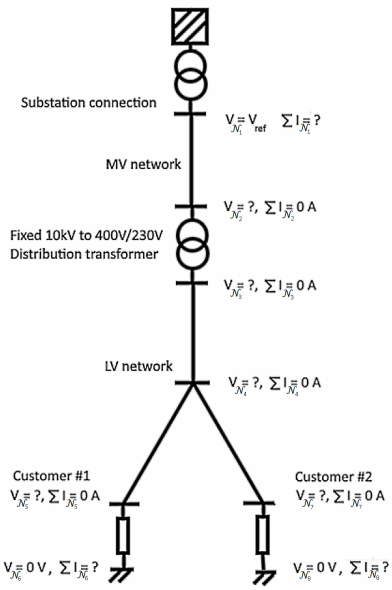


Figure 2.3: A small example MV/LV voltage network. The network has two customers modeled as resistors and a single connection to the medium voltage grid. The mathematical model described in this chapter starts at the secondary side of the MV/LV transformer, i.e. the MV network is out of scope.

the load is modeled as a combination of a constant power, constant impedance and constant current [23]. This chapter however proposes a simple linear load flow approach by only using a constant impedance load model and investigates its feasibility in a real world situation.

To create the constant impedance model as in Figure 2.3, it is necessary to convert the power use of a customer into an equivalent resistance. This can be done by the following formula:

$$Z_{\text{eq}} = U_{n,\text{ref}}^2 / P_{\text{user}} \quad \forall n \in \mathcal{N} \tag{2.1}$$

Here Z_{eq} is the equivalent resistance of the customer, P_{user} the real power consumption of the customer, n is a bus which represents a customer connection and $U_{n,\text{ref}}$ is the voltage at the customer location. Since the voltage at the customer is usually not known, the reference voltage is be assumed to be the nominal voltage.

From Figure 2.3 it can be observed that all nodes on the end of the network are now defined as swing buses, i.e., fixed voltage points. As the power constraints are replaced by resistances, the network now only consists of voltage sources, ground connections and resistors, resulting in a fully linear model.

The network is modeled as a graph. A standard way to define such a graph is by defining graph G as $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ are the nodes and $\mathcal{E}$ are the network edges. In case of an electricity network $\mathcal{N}$ represent the network buses and $\mathcal{E}$ are the network cables. The goal of the model is to determine the cable currents $I_{\mathcal{E}}$ and the nodal voltages $U_{\mathcal{N}}$.

The network voltages and currents can be obtained by using Ohm's law:

$$I_{\mathcal{N}} = \bar{Y} U_{\mathcal{N}} \tag{2.2}$$

Here $I_{\mathcal{N}}$ is the current entering a network bus and $\bar{Y}$ is the so-called admittance matrix. The admittance matrix can be directly obtained from the network lay-out using the following formula [24]:

$$\bar{Y} = A Z_{\mathcal{E}}^{-1} A' \tag{2.3}$$

Here A is a directional connection matrix. Every row corresponds to a network bus. Every column of A corresponds to a network cable. Each cable should have exactly one starting point denoted by a '1' and one end point denoted by '-1'. It does not matter which bus of A contains the minus sign, as the resulting admittance matrix $\bar{Y}$ will stay the same. $Z_{\mathcal{E}}$ is a square matrix and has the corresponding impedance of each cable and the equivalent resistance of the customers (Z_{eq}) on its diagonal. Since the matrix is diagonal, its inverse can be easily calculated by taking the inverse of every diagonal element.

However, (2.2) cannot be solved directly, because not all elements are known in neither vector $I_{\mathcal{E}}$ and $U_{\mathcal{N}}$. To overcome this problem, it is practical to segment the problem in two equations which can be solved separately. This can be done by sorting the rows of the matrices $I_{\mathcal{N}}$, $\bar{Y}$ and $U_{\mathcal{N}}$ in such a way that all swing buses are $\in U_1$. The segments are then defined as:

$$I_{\mathcal{N}} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, \bar{Y} = \begin{bmatrix} K & L \\ L' & M \end{bmatrix}, U_{\mathcal{N}} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{2.4}$$

Since the network is modeled as a set of voltage sources and resistances, Kirchoff's law dictates that $\Sigma I = 0$ on every bus in U_2. Therefore I_2 is equal to $\bar{0}$. All the voltages on the end nodes, represented by U_1 are known. The voltages in U_1 are zero, except for the transformer voltage. The load flow equations now become:

$$\begin{bmatrix} I_1 \\ \bar{0} \end{bmatrix} = \begin{bmatrix} K & L \\ L' & M \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{2.5}$$

A natural way to solve for U_2 is:

$$U_2 = -M^{-1}(L' U_1) \tag{2.6}$$

However, matrix M is usually too large and too costly to invert. Fortunately, it is not necessary to compute M^{-1}. Instead, it is more practical to solve:

$$L' U_1 = -M U_2 \tag{2.7}$$

Since this equation is in the form $Ax = B$ it can be solved in many practical ways e.g. a sparse QR decomposition. Finally after computing the voltages, the cable currents can be directly calculated by:

$$I_{\mathcal{E}} = Z_{\mathcal{E}} A' U_{\mathcal{N}} \tag{2.8}$$

2.3.2. Simulating reactive power without imaginary numbers

Not all simulation environments can solve complex numbers. For example, the programming language *R* and the Matlab PLC compiler have out-of-the-box support for solving matrices which are both sparse and complex. In these cases, it is beneficial to simulate the imaginary parts of the load flow simulation in terms of only real numbers.

To add reactive power to the load flow simulation, the cable reactances are added to $Z_\mathcal{E}$, such that elements of $Z_\mathcal{E}$, Y, U and $I \in \mathbb{C}$. To include these efficiently in (2.2), it can be expanded [25, 26] to:

$$\begin{bmatrix} I_\mathbb{R} \\ I_\mathbb{C} \end{bmatrix} = \begin{bmatrix} Y_\mathbb{R} & -Y_\mathbb{C} \\ Y_\mathbb{C} & Y_\mathbb{R} \end{bmatrix} \begin{bmatrix} U_\mathbb{R} \\ U_\mathbb{C} \end{bmatrix} \tag{2.9}$$

where the subscripts $\mathbb{R}, \mathbb{C}$ are used to indicate respectively the real and imaginary part of the matrix. Thus, $Y_\mathbb{R} = \mathrm{Re}(Y) = A\,\mathrm{Re}(Z^{-1})A'$, and correspondingly $Y_\mathbb{C} = \mathrm{Im}(Y)$. Using the same method as before this can be simplified to:

$$\begin{bmatrix} M_\mathbb{R} & -M_\mathbb{C} \\ M_\mathbb{C} & M_\mathbb{R} \end{bmatrix} \begin{bmatrix} U_{\mathbb{R},2} \\ U_{\mathbb{C},2} \end{bmatrix} = -\begin{bmatrix} L_\mathbb{R} & -L_\mathbb{C} \\ L_\mathbb{C} & L_\mathbb{R} \end{bmatrix} \begin{bmatrix} U_{\mathbb{R},1} \\ U_{\mathbb{C},1} \end{bmatrix} \tag{2.10}$$

which is the complex variant of the equation $MU_2 = LU_1$. By solving this equation, the voltages can be determined. Equation (2.9) can be used to find the currents through the cable-segments. Then $U_\mathcal{N}$ and $I_\mathcal{N}$ can be found by:

$$\begin{aligned} U_\mathcal{N} &= \sqrt{U_\mathbb{R}^2 + U_\mathbb{C}^2} \\ I_\mathcal{N} &= \sqrt{I_\mathbb{R}^2 + I_\mathbb{C}^2} \end{aligned} \tag{2.11}$$

2.3.3. Motivation for linear modeling

The LV networks are generally well conditioned for linear simulations. Compared to the MV network, they consist of relatively short cables with a low X/R ratio. As can be derived from 2.1, the difference between the non linear constant power model and the constant impedance model is caused by the voltage drop i.e. the difference of the estimated voltage $U_{n,\mathrm{ref}}$ and the actual voltage U_n. (As described in Section 1.3.)

According to Alliander DNO policies derived from European law, the voltage drop in the LV network is not allowed to be more than 4.5%. In a network with a voltage drop of 4.5% the absolute difference between a constant impedance model and a constant power model is less then 1V. However, the linearization quickly loses its accuracy as the voltage drop gets higher and is only to be applied on networks with a 'relative small' voltage drop.

On a side note, it can be argued that modeling the customers as a constant impedance load model is not necessarily less accurate as a constant power load model. In reality, customers will have a mix of devices which require a constant power load, such as home computers and TVs, and devices which are in reality a constant impedance load, such as boilers and heaters.

The main reasons for linearizing the load flow model are the improvement in speed and stability regarding a non linear model. Since the load flow equations can be solved without iterative methods, they can be solved for large networks in a very short time span [25]. This makes it viable for control purposes, as it can be used to evaluate many different control strategies. Computational power is often an expensive resource in a control environment. For example, the local controller of the community battery has a clock speed of 500MHz and 64MB RAM, which is very slow compared to a modern PC.

Regarding model stability, a linear network model it is not prone to finding unfeasible solutions or numerical difficulties, which can occur in normal load flows [23]. Given that stability of a controller is essential, this property makes the linear method more suitable for control.

While in most cases the LV network is radial, this is not necessarily always the case. There are low voltage networks which are operated in a non-radial manner and sometimes span over a thousand kilometer of cable and supply tens of thousands of households. It happens that the load flow equations as formulated in this chapter are directly applicable to these large LV grids, while maintaining good performance [5].

A final advantage of a linear network model is its linear additive property, which means that each network load configuration can be simulated independently. In practice this means that all loads can be simulated separately and the resulting voltage drop and cable currents can be obtained by simply taking the sum of all solutions. This property will be exploited in next section to efficiently determine the maximum power the battery supply to or draw from the network.

A small downside of the constant impedance model is that it is prone to computational errors when a customer's power consumption is very close to zero. As can be seen in (2.1), if the power consumption is zero, the equivalent resistance is infinite. In practice, this problem can be easily solved by ensuring that the power consumption of each customer is always a few watts, which has a negligible influence on the outcome of the simulation.

2.3.4. Formulating the battery control problem

Given the model of the LV network and the framework of Figure 2.2, the next step is to formulate a battery controller which safeguards the voltage and current constraints of the network while being compatible with other control goals, such as day-trading. Furthermore the algorithm has to be stable and operable in real-time.

The controlled variable is the battery power at each time step P_t. The final goal of the controller is to keep the battery at a certain given charge level $E_{t,\text{ref}}$. This desired charge level is given by another entity, like a day-trader who is using the battery for energy trading. The optimization function is now posed as a discrete receding horizon problem. The objective function becomes:

$$\underset{P_t}{\text{minimize}} \quad \sum_{t=1}^{T} |E_t - E_{t,\text{ref}}| \tag{2.12}$$

The power the battery can inject in or draw from the network is limited by its rated power P_{bat}, but also by the network voltages at each customer bus ($U_c \ \forall c \in \mathcal{N}$) and cable currents I_e at each cable (e). These constraints have to be satisfied at every time step t in prediction horizon T. E_t is the energy in the battery at time step t. The first step is to apply the load flow model to this optimization problem.

Because the constructed load flow model is linear, network states can be evaluated independently using the principle of superposition. This is useful, as the impact of the network load on the voltages and currents can be calculated separately from the impact of the battery power. The maximum and minimum battery power can therefore be obtained by dividing the available voltage drop by the voltage drop caused by the battery at 1 W. Since the network is only as strong as its weakest connection; the weakest cable or bus determines the boundaries of the battery. These boundaries can be calculated with the following formulas:

$$\begin{aligned} P_{U,\text{max},t} &= \min \frac{U_{\text{max}} - U_{n,t}}{\Delta U_{P_{\text{W},n}}} \quad \forall n \in \mathcal{N}, t \\ P_{U,\text{min},t} &= \max \frac{U_{\text{min}} - U_{n,t}}{\Delta U_{P_{\text{W},n}}} \quad \forall n \in \mathcal{N}, t \end{aligned} \tag{2.13}$$

Here $P_{U,\text{max},t}$ and $P_{U,\text{min},t}$ are the maximum and minimum power the battery is allowed to inject into the network at time t without violating any voltage limits, U_{max} and U_{min} are the maximum and minimum allowed voltage at each customer by law. $U_{n,t}$ is the voltage at each customer which can be calculated by solving (2.2).

$\Delta U_{P_{\text{W},n}}$ is the voltage drop by applying 1 W of battery power to the grid in V/W. This variable is time-invariant and only depends on the network properties. The number can be obtained by solving by setting the battery power to 1 W, setting the customer power to a low but nonzero value and solving (2.2). The customer load cannot be set to zero as it would result in an infinite equivalent resistance, which can cause numerical issues with most numerical matrix solvers. $\Delta U_{P_{\text{W},n}}$ is network-dependent and does not vary over time.

For monitoring currents similar formulas exists:

$$\begin{aligned} P_{I,\text{max},t} &= \min \frac{I_{\text{max},e} - I_{P_t,e,t}}{I_{\Delta P_{\text{W},n}}} \quad \forall e \in \mathcal{E} \\ P_{I,\text{min},t} &= \max \frac{I_{\text{min},e} - I_{P_t,e,t}}{I_{\Delta P_{\text{W},e}}} \quad \forall e \in \mathcal{E} \end{aligned} \tag{2.14}$$

Here $P_{I,\text{max},t}$ and $P_{I,\text{min},t}$ are the maximum and minimum power the battery is allowed to inject into the network at time t without violating any current limits. $I_{\text{max},e}$ and $I_{\text{min},e}$ are the maximum and minimum allowed currents at cable e. $I_{P_t,e,t}$ is the current at each cable which can be calculated by solving (2.2) and applying (2.8). $I_{\Delta P_{\text{W},e}}$ is the current change per 1 W of battery power applied to the grid in A/W. Just like the $\Delta U_{P_{\text{W},c}}$ this variable is time-invariant and only depends on the network properties.

The battery has also a maximum rated power. The full network-related constraints are now defined as:

$$\begin{aligned} P_{\text{max},t} &= \min(P_{\text{max},t,U}, P_{I,\text{max},t}, P_{\text{bat,max},t}) \\ P_{\text{min},t} &= \max(P_{\text{min},t,U}, P_{\text{min},t,I}, P_{\text{bat,min},t}) \end{aligned} \tag{2.15}$$

The optimization problem with added constraints becomes:

$$\begin{aligned} \underset{P_t}{\text{minimize}} \quad & \sum_{t=1}^{T} |E_t - E_{t,\text{ref}}| \\ \text{subject to} \quad & P_{\text{min},t} \leq P_t \leq P_{\text{max},t} \quad \forall t \in T \\ & 0 \leq E_t \leq E_{\text{max},t} \end{aligned} \tag{2.16}$$

Here P_t is the real power the battery supplies to the electricity network at timestep t. E is the energy stored in the battery, which cannot exceed E_{max}. The first constraint corresponds to the network-related power limit. The second constraint ensures that the battery will not discharge when it is empty and not charge when it is full. If all the currents and voltages are within their boundaries, the battery does not need to act. However, if an undesired value is found, the battery will try to mitigate the problem.

However, the formulated problem can not directly be put in a linear solver in its current form. To solve the problem using a linear solver, it is necessary to incorporate the absolute term of the objective function into the constraint function. This is achieved by adding an extra dummy variable $\bar{E}$.

Also, the required voltage and current boundaries may be unattainable, because of the practical limitations of the battery. In such a situation, the linear solver will not find a feasible solution and the battery will be inactive. A more desirable behaviour is in a practical case to meet the required voltage and current constraints as much as possible.

To this end, a barrier function has been implemented. The barrier function gives a large penalty for violating the voltage and current boundaries, barely influencing the regular optimization. In case of unattainable requirements, the solver will still find a solution which violates the constraints as little as possible. The variable corresponding to the barrier function is P_{over}. This function is given a large weight c, where $c >> \bar{E}$.

The definitive optimization problem now becomes:

$$\begin{aligned} \underset{E_t, \bar{E}_t, P_{\text{over}}}{\text{minimize}} \quad & \bar{E} + c \cdot P_{\text{over}} \\ \text{subject to} \quad & \bar{E} + E_t \geq E_{\text{ref}} \qquad \forall t \in T \\ & \bar{E} - E_t \leq E_{\text{ref}} \\ & 0 \leq E_t \leq E_{\text{max},t} \\ & E_{t+1} - E_t - P_{\text{over},t} \leq P_{\text{max},t} \\ & E_{t+1} - E_t + P_{\text{over},t} \geq P_{\text{min},t} \\ & E_1 = E_{\text{start}} \end{aligned} \tag{2.17}$$

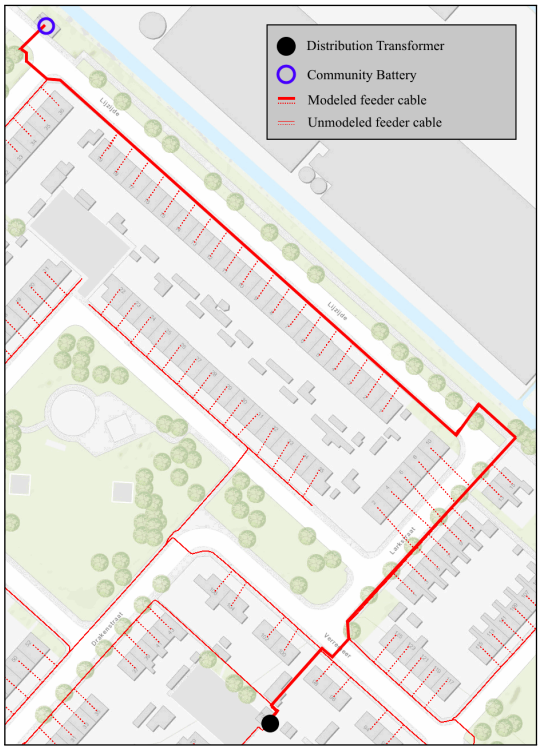


Figure 2.4: GIS view of Liander's low voltage network of Rijsenhout. [27] The outlined modeled network is the feeder that is considered for the LV model. The unmodeled cables are not physically connected to the modeled network, except for a connection in de MV/LV transformer.

This problem can be directly solved by a linear optimization solver. Since the problem is linear, the solution will be optimal if it is found. If the constraints do not conflict with each other, the solution always exists.

Since the battery controller is designed with a horizon of several days, the energy lost by the self-discharging of the battery is neglected. Furthermore, the load cycle efficiency of the battery is also neglected, as it is known to be over 90% in normal operating ranges. To mitigate the inaccuracies caused by these assumptions, E_{start} has to updated at every optimization step using the measurement of the state of charge of the battery provided by the battery management system.

2.4. Experimental setup part I: The community battery of Rijsenhout

Alliander, the largest DNO of the Netherlands serving over three million customers, placed a community battery in Rijsenhout, a suburban village close to Amsterdam, the Netherlands. The battery is connected to the low voltage power grid as can be seen in Figure 2.4. The community battery has a usable energy rating of 126 kWh

and a 55 kW peak power rating. The battery itself is capable of a higher power output, but 55 kW was the expected limit of the power that could be connected to the network without causing voltage problems.

Using the conventional load flow software and modeling assumptions of Alliander DNO, an analysis of the network of Rijsenhout showed that the network was expected to have no voltage or capacity problems. However, during the experiments it became clear that the conventional modeling assumptions were incorrect and the network was subject to voltages which were too high according to regulations. Sensor data proved that the voltage problems were caused by fluctuations of the voltage on the medium voltage grid, which exceeded the modeling assumptions. However, this situation provided an excellent opportunity to prove that the battery could also mitigate the voltage problems. In the Netherlands, inverters are required to automatically switch off in the event the voltage is above 250 Volts to mitigate over-voltages. This threshold was exceeded on a regular basis in the Rijsenhout network.

During the experiment, only active power was considered because very little reactive power could be expected to be present in this LV network. The customers in this network are regular households, which are known to consume little reactive power. Also, the X/R ratio of the cables is very low (around 0.1), making the phase angle nearly constant in the entire network. There is also a practical reason for neglecting reactive power as the installed sensors only logged real power.

To make the experiment broader than just the DNO perspective, an additional control objective was formulated. Most of the customers have their own PV installation, and by aggregating their consumption and defining it as E_{ref}, the customers can 'live on their own solar energy' as much as possible. This is also of interest to the DNO as it mitigates peak loads from other network areas.

2.4.1. Optimization results

At the distribution transformer and the community battery both power and voltage are measured. The modeled part of the network consists of 34 customers. At 12 households, the power was measured. For privacy reasons, their exact location could not be displayed, but they are almost uniformly distributed along the cable. The data which is displayed in this section is averaged on the time scale of one minute.

Figure 2.5 shows the result of the attempt to make the LV network self-sufficient by applying the optimization algorithm from (2.17). The community battery did most work in August, nearly doubling the self-consumption of the generated solar energy within the LV grid. Still 77 kWh per household could not be stored in the battery because of capacity limits in this months. August and September are the two last months of the summer in the Netherlands. In October and November, there was much more power consumption on average and less solar power generation. It can be observed from Figure 2.5 that in October and November almost no power was delivered to the grid for this reason. Battery losses (and other transportation losses in general) are not part of Figure 2.5, because they have to be compensated by the DNO.

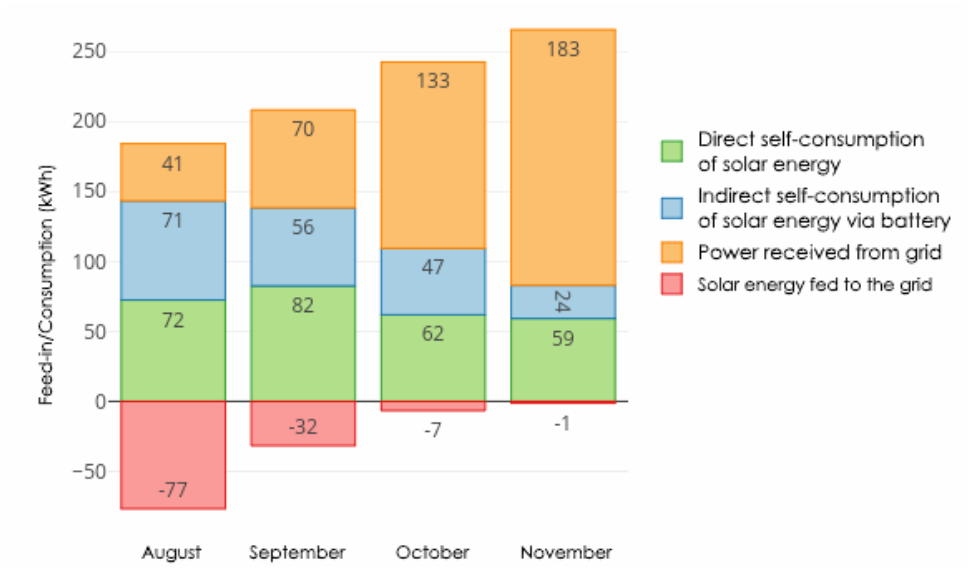


Figure 2.5: Source of the electricity of a single average customer in the community battery LV grid from August 2017 to November 2017.

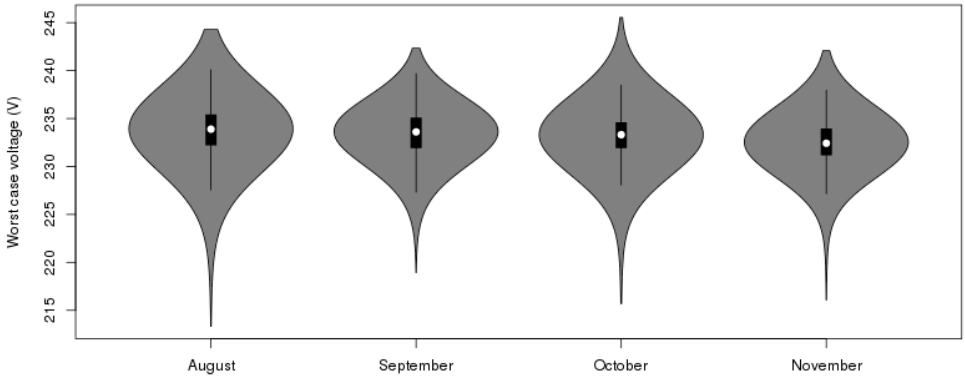


Figure 2.6: Measured voltages at the community battery in the network from August 2017 to November 2017. The LV voltage never exceeds the bounds of 245 Volt and 215 Volt as required. The grey area depicts the density of the measurement points in the specific month. The white dot depicts the month's mean voltage. The black bar denotes the interquartile range, i.e. the values between the first and second standard deviation.

Figure 2.6 shows the maximum measured voltage at the community battery. The battery controller keeps all voltage within the set bounds of 245 Volt and 215 Volt. In September and November, the community battery did not need to act to keep the voltages within the required bounds. Since the battery is located at the very end of the LV network, it is reasonable to assume that this is close to the extreme voltage in the network. During the months the community battery and its controller were active, the voltage was lowered from previously observed voltage peaks of 250 Volt to voltage peaks of 245 Volt, mitigating the voltage problems.

For solving (2.17), the optimizer depends on a prediction of the power consumption and solar power generation. This prediction is obtained by training a regression model using historical data and was provided by an external party. The model has a Mean Absolute Percentage Error (MAPE) of 5% for predicting household load and 10% for PV power 24 hours ahead. During the experiment it was discovered that the accuracy of the energy consumption predictions is relatively unimportant. The controller anticipates on high demand or load by reserving capacity of the battery. Using this available energy and the available real-time measurements, the controller then reacts to the voltage/current problems once they actually arise. It turns out that the most important prediction feature is the required amount of energy to mitigate voltage/current problems, not the exact peak loads.

2.4.2. Checking the linearity assumption

This section investigates how well the linear constant impedance load flow model applies to low voltage networks. The LV model constructed in the previous sections relies on a main assumption: The load model is assumed to behave as a constant impedance.

The theoretical difference between the two models can be observed in Figure 2.7. Figure 2.7 is only the result of simulating a single cable, but the results are generalize-able, since the difference between load models only depends on the voltage drop according to (2.1).

It can be observed that the results are very close for low voltage drops, but quickly diverge if the voltage drops become larger. For a voltage drop in the network of 4.5% (using constant power), the difference of the two load models is 0.44%, which is 1.0V at a reference voltage of 230V. At 10%, the difference is 2.4%, which is 5.5V at a reference voltage of 230V.

If this difference is acceptable depends on the application. In most cases, the differences are still well within the uncertainty caused of other imprecise properties, such as unknown loads and imperfect cable data. In the case of studying an LV network, the voltage drop is never to be expected to be larger than 4.5 % and in most cases an error of 1V is acceptable.

With a voltage drop over 4.5%, however, one needs to be careful. For network design purposes, usually a maximum voltage drop of 10% is used. The difference between load models causes a 2.5% error in this range, which can be significant. However, the error is very predictable and can be corrected for by either adjusting the maximum allowed voltages or by iterating the model with the newly simulated voltages at the customer.

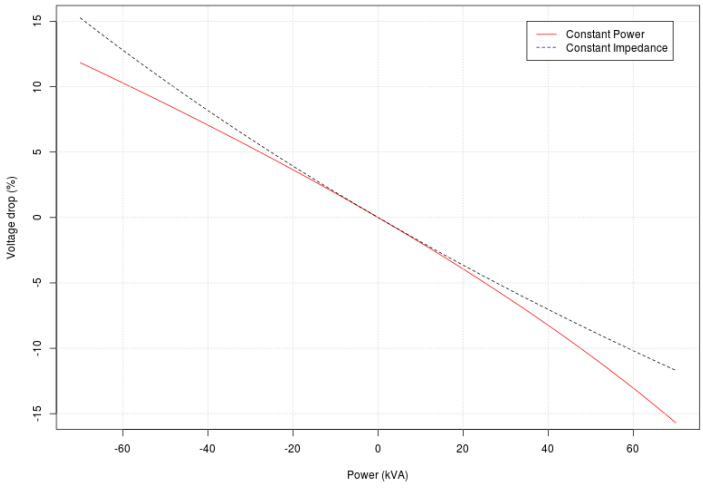


Figure 2.7: The difference between voltage drop of a constant impedance load and a constant power load. The voltage drops were obtained by simulating a single power line (Z = 0.1 Ohm) with a single load.

A final note is that neither model is perfectly accurate. Most household loads are a mixture of the two load models. Hence this section also validates the linear model using empirical results.

To determine if the constant-impedance model is indeed accurate on the voltage range of the LV network, a short experiment was performed. As can be observed in Figure 2.8, the battery was given a significant 'saw tooth' shaped load profile as a reference. The charging experiment was performed in a few hours around noon, which is the time with the least power consumption during the day because of the presence of solar panels. The customer power consumption is significantly less than the battery power.

As can be seen in Figure 2.8, the battery ramped up and down from 50kW, its maximum rated power. It can also be concluded that the battery can control the voltage at the end of the LV network either 12 Volt up or down, covers the entire range of the allowed 4.5% voltage drop on LV networks. To determine the exact relation between battery power and voltage drop, the plot in Figure 2.9 was constructed. From this figure, it can be concluded that the relation between battery power and voltage drop can indeed be approached by a linear function within the operating range of the battery.

The deviations around zero battery power, which can be observed in Figure 2.8, are caused by the imperfect inverter. The battery inverter cannot behave linearly at very low battery power levels. It can also be observed in Figure 2.9 that the linear fit has a slight additive bias. This is caused by the small residual load which also can be observed in Figure 2.8.

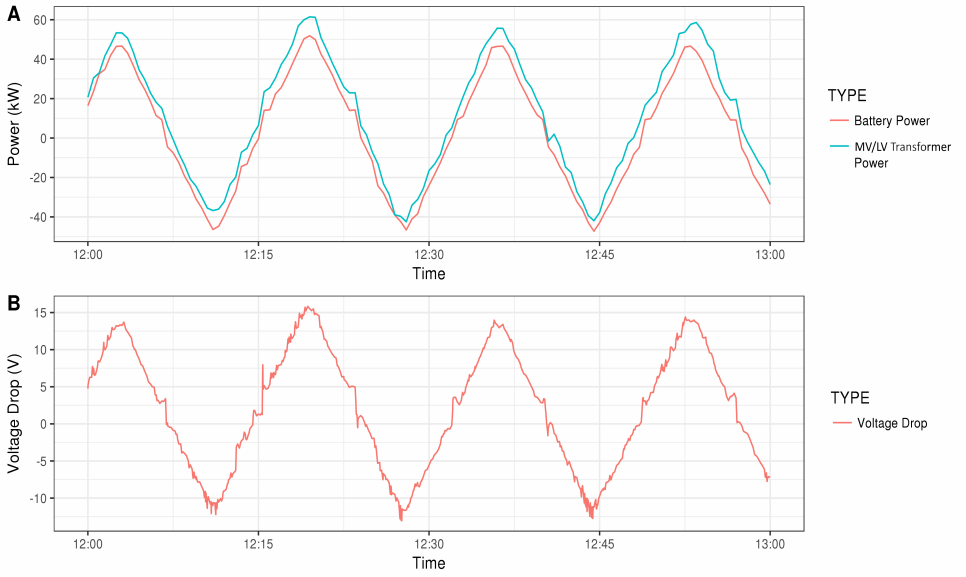


Figure 2.8: To determine the characteristics of the voltage drop the battery was given a 'saw tooth' shaped charging profile. It can be observed that there is an approximate linear relation between the battery power and the voltage drop between the transformer and the battery.

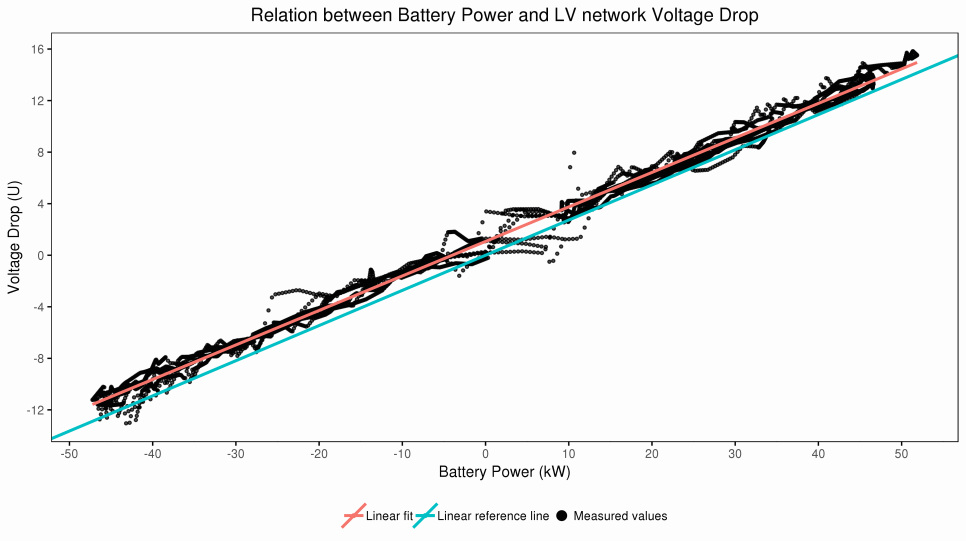


Figure 2.9: The relation between the Battery Power and the measured voltage drop from the MV/LV transformer to the Battery. The time span of this figure is the same as in Figure 2.8. It can be observed that the relation between battery power and voltage drop can be approximated by a linear function.

2.5. Community battery design specifications

The following section contains design principles to quickly determine the key properties of a community battery for network congestion reduction purposes, in new or existing grids. Using the observation of the previous sections, generalized rules have been established. These rules have been designed to be used by network planners and have been kept as simple as possible. It has been assumed that a standard load flow simulation is unavailable to maximize the simplicity of the network analysis. A drawing of the low voltage network is sufficient to apply the proposed rules, once the problem and its size are known.

There are two main motivations considered to place a community battery for network congestion reduction: to control the community voltage and to control the community currents. While it is theoretically possible to also control the network power factor to some extent, this is currently not a priority for distribution network operators, because of its rare occurrence.

A simple but realistic situation is assumed. The network has a relatively simple radial structure and its cable locations and properties are known. It is also assumed that the location and size of the voltage problems are roughly known. These either have been determined using smart meter data or direct (temporary) measurements.

As will be motivated in the following paragraphs, the most important properties of a community battery are:

- The location of the connection of the battery to the low voltage power grid
- The battery power rating
- The battery energy capacity

The first step in designing a community battery is determining its location. The size of a large battery and its control installation is significant, which strongly limits the number of available placement locations. For example, in the case of the community battery of Rijsenhout the battery size is half a standard shipping container and only a single placement location was available.

Given is the network model of Figure 2.10 and the linear relation between battery power and voltage drop as seen in Figure 2.9. If one assumes that the customer load is not significantly influenced by the voltage drop, simple approximate formulas can be constructed for battery placement.

In the previous section it was shown that there is a linear relation between battery power and voltage level, which motivates the next formula. Once the location is determined, one can determine the minimal required power rating of the battery with the following formula:

$$P_{\min,U} = \alpha \frac{\Delta U}{l} \tag{2.18}$$

Here $P_{\min,U}$ is the minimal battery power required to solve the voltage problem (W), ΔU is the size of the largest voltage problem (V) i.e. the node which has the largest

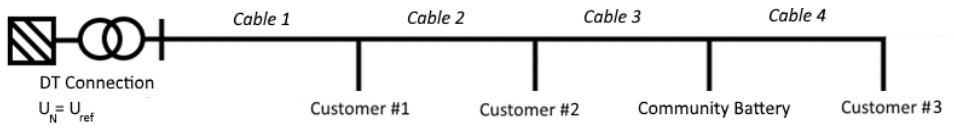


Figure 2.10: A battery placement example. If a battery is placed at between customer 2 and 3, it can control the currents and voltage drops in cable 1,2 and 3.

gap between the allowed voltage (U_{max}) and its maximum measured voltage. l is the length of the part of the cable which is shared by both the customer with the largest voltage problem and the community battery (m). For example, if in Figure 2.10 ΔU is located at customer 2, then $l = l_{\text{Cable 1}} + l_{\text{Cable 2}}$. α is a cable dependent factor ($\frac{\text{m}}{\text{W}\cdot\text{V}}$) which represents the average cable resistance per meter. It can be obtained by the formula: $\alpha = U_{\text{ref}}/\rho$. Here U_{ref} is the reference voltage (V) and ρ is the cable specific resistance (Ω/m). α is used instead of adding U_{ref} and ρ to (2.18), making it easier to construct a look-up table for engineers. As a general guideline, a community battery should be connected with the longest possible path to the MV/LV transformer. Or to put in another way, the community battery should share as much meters of cable with the customers as possible.

To determine the minimum power required to mitigate current problems and given the network model's assumptions, the following formula is proposed, assuming that the power will mainly flow between the MV/LV transformer and the battery:

$$P_{\text{min},I} = U_{\text{ref}} \cdot \Delta I \tag{2.19}$$

Here ΔI is the size of the largest current problem, i.e. the difference between the actual current and the maximum cable capacity. It is essential that this current problem is situated on the cable between the battery and the distribution transformer or else the battery power will have negligible effect on this particular current. For example, if in Figure 2.10 ΔI is located at customer 3, then the community battery cannot solve this problem.

The minimal required battery power now becomes:

$$P_{\text{min}} = \max P_{\text{min},U}, P_{\text{min},I} \tag{2.20}$$

It can be concluded from (2.18) and its underlying equations, that the location of the battery is its most critical aspect. (2.20) shows that the location has great impact on the required amount of battery power regarding voltage problems. Because of the linear property of (2.18), a battery placed twice as far away from the MV/LV connection generally needs only half the power rating. Current problems cannot be influenced at all if the battery is not in the correct position, as (2.19) requires the overcurrent to be between the transformer and the battery itself.

Once the minimal required battery power has been determined, one can determine the required battery storage capacity. The storage capacity should meet two criteria; it should be sufficiently large to provide the requested power and solve the voltage or current problem. These criteria are displayed in (2.21) and (2.22) respectively.

$$E_{\text{min}} = |P_{\text{min}}| \cdot C \tag{2.21}$$

In this formula E_{min} is the minimal required storage capacity (kWh). C is the so-called C-value [28], a factor which expresses the relation between battery power and capacity (kWh/kW). This factor is mostly technology dependent and can be as high as 3 for lithium-ion batteries [29]. The second criterion is:

$$E_{\text{min}} = \frac{1}{t_s} \sum_{i=1}^{n} P_{\text{min},i} \tag{2.22}$$

Here t_s is the sampling period, n the number of samples and $P_{\text{min},i}$ the required battery power at time step i. The formula should be applied for the duration of the voltage/current largest problem. Since community batteries are generally not meant for seasonal storage, it is sufficient to sample several days.

Using the newly determined properties of the community battery, a cost-benefit analysis can be easily obtained because location, power rating and capacity are also the most important factors for battery costs. The exact break down of the operational and capital costs of the battery and its comparison to conventional network strengthening methods are beyond the scope of this chapter.

2.6. Experimental setup part II: Dimensioning the community battery of Rijsenhout

In this section, the previously proposed formulas are applied on the community battery of Rijsenhout as an example. While studying the network of Rijsenhout, it was discovered that the largest voltage problem is 5 Volt. The location of this problem is at the customer closest to the community battery. Applying (2.18) results in a required battery power of 15 kW. This is much less than the battery's rated power of 55 kW. No current problems were measured in the network of Rijsenhout, so (2.19) does not need to be applied.

At its worst, the voltage problem was present for several hours. Determining the required capacity by applying (2.22) yielded a required capacity of 35 kWh. However, given a C-value of 3 and using (2.21), the minimal required capacity turns out to be at least 45 kWh. This is also much less than the rated capacity of 125 kWh. It turns out the community battery of Rijsenhout could have been approximately 50% smaller.

It is interesting to note that the optimal location for stabilizing the battery is the inverse of the optimal location for day trading. (2.18) shows that a battery has a larger influence on the voltage if it is further away from the transformer. While this is a desirable trait if the battery is used for network stabilization, a battery used primarily for day trading should be placed as close to the transformer as possible to minimize the impact on the grid. However, regarding voltage problems the negative impact of day trading can be partly mitigated by using the reactive power control capabilities of a battery inverter. These capabilities were not available in this field test and are therefore beyond the scope of this chapter.

Furthermore, (2.18) can also be used to determine the potential de-stabilization of the grid by a large battery. For example: Given the fact that the community battery of Rijsenhout can control the voltage 12 Volt and that the maximal measured voltage in the LV network is already 250 Volt, the community battery can easily cause voltages of over the legal limit of 253 Volt. To be able to estimate the occurrence of these kinds of problems is of great use for DNOs and it shows that it is risky to not monitor installed batteries with power limits in the order of 10kW.

2.7. Conclusion

This chapter provides a solid foundation for integrating residential and community-level storage in the existing LV network. A fast linear LV model was developed and applied to the LV feeder of the community-battery in Rijsenhout. The model was proven to be sufficiently accurate for network stabilization purposes.

The battery control theory was formulated as a linear optimization problem. A receding horizon controller was developed to be used in a continuous way. The controller is suited very well to be integrated with other battery control goals, while still securing the voltages and currents within the network. It has been shown that a community battery is able to stabilize and control the loads in a real world low voltage network to a large extend.

A step-by-step method was proposed for quickly modeling the impact of new and existing batteries on the LV grid. Both the stabilizing and destabilizing potential regarding steady-state voltages and currents of the battery can be quickly estimated. By unlocking the potential of battery storage on a DNO level, a fast and secure energy transition is one step closer.

Future research

Being able to control such a large battery to freely test control algorithms provides many opportunities for future research. Next steps will include the application of state estimation algorithms to optimize the estimates of the voltages and currents and also properly account for uncertainties in network measurements and properties.

Furthermore, the exact low voltage network load in generally not known. It is the author's ambition to create an accurate load model by applying system identification methods, using the community battery as a means to 'excite' the network to find the voltage and current dependability of the power loads. This could result in a general method for DNOs to identify the load types of their networks.

As additional validation, the theoretical network design formulas proposed in this chapter will be tested on more test beds. A test bed of 50 residential batteries is currently in development. Researchers are welcome to contact the corresponding author of this chapter to see if their own algorithms can be tested at our facility.

Acknowledgment

The authors give special thanks to the community battery project team at Alliander, without this research would not have been possible. The team provided the actual

experimental setup from design to realization, a process which took over a year. The project team consisted of: Jan Willem Eising, Hans Beckers, Rob van Olst, Sander Schoot Uiterkamp, Paul Albers, Dieneke van der Berg, Olaf de Leeuw, Maikel Juriaan Kryzewski, Mike van der Heijden, Mathijs de Groot, Maurice van Duijnhoven, Sidoeri Dekker, Peter Wiebes, Gilbert Smink and Harold Veldkamp. Also thanks to Roel Dobbe and Jacco Heres who provided much valuable feedback. Thanks to Barbera Droste for providing the framework regarding reactive power modeling.

References

[1] D. Parra, S. A. Norman, G. S. Walker, and M. Gillott, *Optimum community energy storage system for demand load shifting,* Applied Energy **174**, 130 (2016).

[2] D. Parra, S. A. Norman, G. S. Walker, and M. Gillott, *Optimum community energy storage for renewable energy and demand load management,* Applied Energy **200**, 358 (2017).

[3] P. Mallet, P. O. Granstrom, P. Hallberg, G. Lorenz, and P. Mandatova, *Power to the people!: European perspectives on the future of electric distribution,* IEEE Power and Energy Magazine **12**, 51 (2014).

[4] E. Veldman, M. Gibescu, H. J. Slootweg, and W. L. Kling, *Scenario-based modelling of future residential electricity demands and assessing their impact on distribution grids,* Energy Policy **56**, 233 (2013).

[5] W. van Westering, A. Zondervan, A. Bakkeren, F. Mijnhardt, and J. van der Els, *Assessing and mitigating the impact of the energy demand in 2030 on the dutch regional power distribution grid,* in *Networking, Sensing, and Control (ICNSC), 2016 IEEE 13th International Conference on* (IEEE, 2016) pp. 1–6.

[6] C. Heymans, S. B. Walker, S. B. Young, and M. Fowler, *Economic analysis of second use electric vehicle batteries for residential energy storage and load-levelling,* Energy Policy **71**, 22 (2014).

[7] K. Worthmann, C. M. Kellett, P. Braun, L. Grüne, and S. R. Weller, *Distributed and decentralized control of residential energy systems incorporating battery storage,* IEEE Transactions on Smart Grid **6**, 1914 (2015).

[8] F. M. Vieira, P. S. Moura, and A. T. de Almeida, *Energy storage system for self-consumption of photovoltaic energy in residential zero energy buildings,* Renewable Energy **103**, 308 (2017).

[9] S. Kumar, S. Chandiwala, V. R. Khare, and M. Pant, *Integrating energy efficiency with renewables and energy storage for a smarter and greener residential solution,* in *ISGW 2017: Compendium of Technical Papers*, edited by R. K. Pillai, G. Ghatikar, R. Seethapathy, V. L. Sonavane, S. A. Khaparde, P. K. Yemula, S. Chaudhuri, and A. Venkateswaran (Springer Singapore, Singapore, 2018) pp. 231–235.

[10] R. K. Lam, D. H. Tran, and H. G. Yeh, *Economics of residential energy arbitrage in california using a pv system with directly connected energy storage,* in *2015 IEEE Green Energy and Systems Conference (IGESC)* (2015) pp. 67–79.

[11] M. Kabir, Y. Mishra, G. Ledwich, Z. Y. Dong, and K. P. Wong, *Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder,* IEEE Transactions on industrial Informatics **10**, 967 (2014).

[12] A. Parisio, E. Rikos, and L. Glielmo, *A model predictive control approach to microgrid operation optimization,* IEEE Transactions on Control Systems Technology **22**, 1813 (2014).

[13] S. Koohi-Kamali, N. Rahim, and H. Mokhlis, *Smart power management algorithm in microgrid consisting of photovoltaic, diesel, and battery storage plants considering variations in sunlight, temperature, and load,* Energy Conversion and Management **84**, 562 (2014).

[14] A. Das and Z. Ni, *A computationally efficient optimization approach for battery systems in islanded microgrid,* IEEE Transactions on Smart Grid (2017).

[15] J. Abushnaf and A. Rassau, *Impact of energy management system on the sizing of a grid-connected pv/battery system,* The Electricity Journal **31**, 58 (2018).

[16] F. Keck, M. Lenzen, A. Vassallo, and M. Li, *The impact of battery energy storage for renewable energy power grids in australia,* Energy (2019), https://doi.org/10.1016/j.energy.2019.02.053.

[17] P. Wolfs and G. S. Reddy, *A receding predictive horizon approach to the periodic optimization of community battery energy storage systems,* in *Power Engineering Conference (AUPEC), 2012 22nd Australasian Universities* (IEEE, 2012) pp. 1–6.

[18] E. L. Ratnam and S. R. Weller, *Receding horizon optimization-based approaches to managing supply voltages and power flows in a distribution grid with battery storage co-located with solar pv,* Applied Energy **210**, 1017 (2018).

[19] R. J. de Groot, B. M. Vonk, H. J. Beckers, and J. G. Slootweg, *Development of a charge path optimization controller block for a battery energy storage system,* IFAC Proceedings Volumes **47**, 8583 (2014).

[20] R. De Groot, F. Van Overbeeke, S. Schouwenaar, and H. Slootweg, *Smart storage in the enexis lv distribution grid,* 22nd International Conference and Exhibition on Electricity Distribution, CIRED 2013 (2013).

[21] J. J. C. M. van Dun, R. J. W. de Groot, J. Morren, and J. G. Slootweg, *Control of a battery energy storage system connected to a low voltage grid,* in *2015 IEEE Eindhoven PowerTech* (2015) pp. 1–6.

[22] H. Le Nguyen, *Newton-raphson method in complex form [power system load flow analysis],* IEEE Transactions on Power Systems **12**, 1355 (1997).

[23] W. H. Kersting, *Distribution system modeling and analysis* (CRC press, 2001).

[24] J. Kirtley, *6.061 Introduction to Power Systems Class Notes Chapter 5 Introduction To Load Flow* (MIT Open Courseware, 2018).

[25] W. van Westering, B. Droste, and H. Hellendoorn, *Combined medium voltage and low voltage simulation to accurately determine the location of voltage problems in large grids,* in *2019 CIRED* (2019) p. 2.

[26] R. Military and I. Popa, *On the numerical solving of complex linear systems,* International Journal of Pure and Applied Mathematics **76**, 113 (2012).

[27] R. Dobbe, W. van Westering, S. Liu, D. Arnold, D. Callaway, and C. Tomlin, *Linear single- and three-phase voltage forecasting and bayesian state estimation with limited sensing,* IEEE Transactions on Power Systems **35**, 1674 (2020).

[28] MIT Electric Vehicle Team, *A Guide to Understanding Battery Specifications* (MIT, 2018).

[29] N. M. L. Tan, T. Abe, and H. Akagi, *A 6-kw, 2-kwh lithium-ion battery energy storage system using a bidirectional isolated dc-dc converter,* in *The 2010 International Power Electronics Conference - ECCE ASIA -* (2010) pp. 46–52.

3

Linear power flow method improved with numerical analysis techniques applied to a very large network

In this chapter, we propose a fast linear power flow method using a constant impedance load model to simulate both the entire Low Voltage (LV) and Medium Voltage (MV) networks in a complete network simulation. Accuracy and efficiency of this linear approach are validated by comparing it with the Newton power flow algorithm and a commercial network design tool Vision on various distribution networks including real network data. Results show that our method can be as accurate as classical Nonlinear Power Flow (NPF) methods using a constant power load model and additionally, it is much faster than NPF computations. In our research, it is shown that voltage problems can be identified more accurately when MV and LV are integrally evaluated. Moreover, Numerical Analysis (NA) techniques are applied to the Large Linear Power Flow (LLPF) problem with 27 million nonzeros in order to improve the computation time by studying the properties of the linear system. Finally, the original computation times of LLPF problems with real and complex components are reduced by 2.8 times and 5.7 times, respectively.

Parts of this chapter have been published in Energies **12**, 4078 (2019). Co-author: Baljinnyam Sereeter

3.1. Introduction

Due to the volatility of renewable energy sources (RES), Distribution Network Operators (DNOs) have a great need for faster power flow calculations for simulating different scenarios for network design. DNOs traditionally treat LV and MV networks as two separate entities where both voltage levels have their own set of assumptions and design policies. However, simulating both the LV and MV networks in a single power flow computation has many advantages. It can result in more effective grid management and opens possibility for integral network design [1]. A fast model, as described in this chapter, also allows large scale live network load modeling, a feature required for many 'smart' solutions such as real time congestion management. Furthermore, it is shown in this chapter that voltage problems can be determined more accurately. On the other hand, this will increase the size and complexity of the power flow problem.

The power flow, or load flow problem, is computed by grid operators to determine whether the power system can function properly for the given generation and consumption. Traditionally, the power flow problem is formulated as a nonlinear system of equations. Therefore, iterative type of methods such as the Gauss-Seidel (G-S), Newton power flow (N-R) and Fast Decoupled Load Flow (FDLF) [2–4] are widely used to solve the so-called Nonlinear Power Flow (NPF) problem for transmission networks. However, these conventional power flow methods do not always converge [5] when they are applied to the distribution power flow problem due to some special features of the distribution network, such as radial or weakly meshed structure, high R/X ratio, line's length and unbalanced loads. Many methods [6–9] have been developed on distribution power flow analysis and the most of them are based on the Backward-Forward Sweep (BFS) algorithm. Several reviews on distribution power flow solution methods can be found in [10–12].

Iterative NPF solution methods solve linearized sub-problems at each iteration step. This means that the computational time of NPF computations can be improved by studying the properties of the linear system solved in every iteration and applying Numerical Analysis (NA) techniques such as different reordering schemes, various direct solvers and numerous Krylov subspace methods. It has been shown that iterative linear solvers can sometimes have faster solution times over sparse direct solvers for very large power flow problems [13–15], making it worth improving them.

Another way to ease the calculation and to speed up the computational time is to linearize NPF equations using some approximations and assumptions in order to obtain the Linear Power Flow (LPF) equations. After the linearization, the resulting LPF equations can be directly solved without requiring iterations. Therefore, LPF computations are generally faster than NPF computations and are more suitable to be applied on very large networks with millions of cables for real time simulation. The best-known example of the LPF problem is the DC load flow [16] where linear relations are determined between the active power injections P and the voltage angles δ, and the reactive power injections Q and the deviations of the unknown voltage magnitudes $\Delta|V|$. Furthermore, the linear power flow formulation is obtained based on a voltage dependent (ZI) load model and some numerical

approximations on the imaginary part of the nodal voltages in [17]. Another linear power flow model based on Taylor's series expansion was proposed in [18]. A direct method taking an advantage of special structure of distribution systems is also developed in [19].

Alternatively, the LPF equations are achieved using only a constant impedance (Z) load model and are used for both strategic studies in [1] and network battery storage control in [20]. This linear model allows for computations to efficiently simulate very large networks, as 22 million buses are simulated in 60 seconds [1]. However, the linearization quickly loses its accuracy as the voltage drop gets higher and is only to be applied on networks with a voltage drop of less than 10%.

In this chapter, we propose a fast LPF algorithm improved with NA techniques to solve the Large LPF (LLPF) problem with 27 million nonzeros simulating both the entire LV and MV networks in a single simulation. We investigate the efficiency and accuracy of the LPF computation using a constant impedance (Z) load model by comparing it with the Newton power flow algorithm and a commercial network design tool Vision on various distribution networks. Furthermore, the network of Alliander DNO is used in our linear power flow computations, which contains both the LV and MV power grids and consists of approximately 80.000 km of cable serving over three million customers.

The main motivation for using a linear model is that the MV/LV network of Alliander DNO cannot be solved by our NPF solver (Vision). Additionally, MV/LV networks are generally very well conditioned for linear power flow calculations because according to Dutch law and Alliander DNO policies, the voltage drop in the MV/LV network is not allowed to be more than 9% in total (as described in Section 1.3). On the LV network, the reactive power and reactance are generally an order lower than the active power and resistance of the network respectively. For this reason, if one simulates the LV grid only and starts from the secondary side of the distribution transformer, it is generally sufficient to only simulate the real part of the network [20]. Therefore, for the application of NA techniques on the LLPF problem, we consider the LLPF problem with first real and then complex components.

This chapter is structured as follows. Section 3.2 introduces the mathematical modeling of the LPF problem using a constant impedance load model. In Section 3.3, the LPF computation is compared with the NPF computation in order to validate the accuracy and efficiency of this linear approach. The case study of the large network of Alliander DNO is given in Section 3.4. Section 3.5 describes the application of NA techniques on LLPF problems using real and complex components. Finally, the conclusion is given in Section 3.6.

3.2. The large linear power flow problem (LLPF)

This section explains how the LPF problem is obtained using a constant impedance load model assuming a balanced network load. The MV and LV network are both modeled using the same physical model. The electricity network can be modeled as a graph $G(\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ represents the network buses and $\mathcal{E}$ the network cables. The relation between the nodal currents $I_{\mathcal{N}}$ and voltages $V_{\mathcal{N}}$ can be defined

by Kirchoff's Current Law:

$$I_{\mathcal{N}} = YV_{\mathcal{N}}, \tag{3.1}$$

where Y is the so-called nodal admittance matrix. In equation (3.1), all variables are given in complex numbers as $V_{\mathcal{N}} = |V_{\mathcal{N}}|e^{\imath\phi} = V_{\mathcal{N}}^r + \imath V_{\mathcal{N}}^m$, $Y = G + \imath B$, and $I_{\mathcal{N}} = I_{\mathcal{N}}^r + \imath I_{\mathcal{N}}^m$. The admittance matrix can be directly obtained from the network lay-out using the following formula [21]:

$$Y = CZ_{\mathcal{E}}^{-1}C', \tag{3.2}$$

where C is a directional connection (branch-node incidence) matrix and $Z_{\mathcal{E}}$ is a square matrix with the corresponding impedance of each cable/edge ($\mathcal{E}$) on its main diagonal.

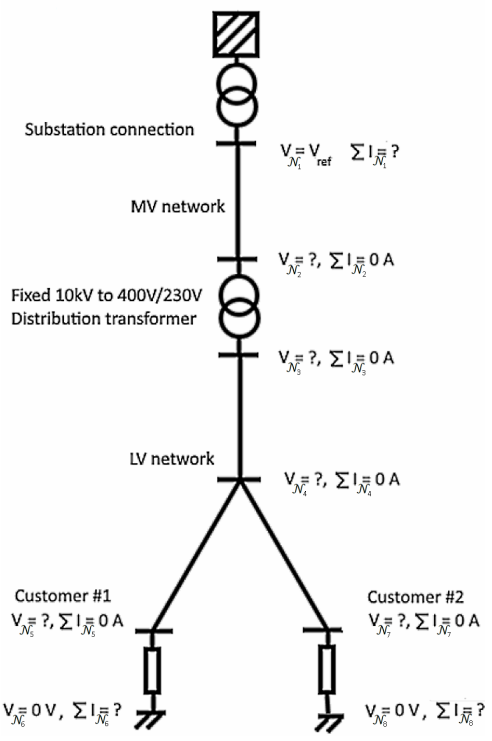


Figure 3.1: An example of a small MV/LV network connected to a substation transformer. While the MV network is always operated radially, the LV network is operated meshed in some areas. The reference voltage (V_{ref}) is defined at the secondary side of the substation transformer. If only the LV network is considered, the reference voltage (V_{ref}) is defined at the secondary side of the distribution transformer.

Generally, the load of the network is modeled as a combination of constant impedance (Z), constant current (I) and constant power (P), also known as the ZIP model [22]. In this chapter, we use only a constant impedance load model in order to obtain LPF equations, so only the Z part of the model is considered. Customers

are modeled by an equivalent resistance R_{ng}^{eq} and reactance X_{ng}^{eq} which are defined by the following formula:

$$R_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2 P_n}{P_n^2 + Q_n^2} \tag{3.3}$$

$$X_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2 Q_n}{P_n^2 + Q_n^2} \tag{3.4}$$

where P_n and Q_n are active and reactive power consumption of load bus n and V_{ref} is the nominal voltage. Note that equation (3.1) cannot be solved directly, because not all elements are known in neither vector $I_{\mathcal{N}}$ and $V_{\mathcal{N}}$. To overcome this problem, we segment the problem in two equations which can be solved separately as:

$$I_{\mathcal{N}} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, Y = \begin{bmatrix} Y_{11} & Y_{21}^T \\ Y_{21} & Y_{22} \end{bmatrix}, V_{\mathcal{N}} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{3.5}$$

We sort the rows of the matrices $I_{\mathcal{N}}$, Y, and $V_{\mathcal{N}}$ in such a way that all swing buses and ground buses are placed in V_1 and all unknown voltages of remaining buses are placed in V_2. In V_1, the voltage of the swing buses is set to the nominal level and ground buses are set to ground voltage level. In the integral MV/LV simulation presented in this chapter, the secondary side of each HV/MV transformer is a swing bus. Since the network is modeled as a set of voltage sources and resistances, Kirchoff's law dictates that $\Sigma I = 0$ on every bus in V_2. Therefore, I_2 is equal to 0 and the power flow equations now become:

$$\begin{bmatrix} I_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{21}^T \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{3.6}$$

Since V_1 is known, V_2 can be solved using the second row of (3.6) as:

$$Y_{22} V_2 = -Y_{21} V_1, \tag{3.7}$$

$$Y_{22} V_2 = b, \tag{3.8}$$

$$V_2 = Y_{22}^{-1} b. \tag{3.9}$$

Then, we compute I_1 as:

$$I_1 = Y_{11} V_1 + Y_{21}^T V_2. \tag{3.10}$$

Finally, after computing the nodal voltages, the cable currents can be directly calculated by:

$$I_{\mathcal{E}} = Z_{\mathcal{E}} C' V_{\mathcal{N}}. \tag{3.11}$$

3.2.1. Solving in terms of only real numbers

It is possible to solve the equation in terms of only real numbers in order to ease the calculation or if your software does not support the combination of the complex variables and sparse matrices like the *R* programming language [1].

Neglecting imaginary parts

In LV networks customers use or produce very little reactive power on average. Additionally, due to the fact that the reactive power and reactance are generally an order lower than the active power and resistance, we can neglect the impact of reactive currents, reactive powers and cable reactance. Thus, the equivalent reactance X_{ng}^{eq} is removed and the equivalent resistance R_{ng}^{eq} becomes as:

$$R_{ng}^{\text{eq}} = \frac{V_{\text{ref}}^2}{P_n}. \tag{3.12}$$

Furthermore, we know that $Y_{22} = G_{22} + \imath B_{22}$ and $b = b^r + \imath b^m$ in Equation (3.9). By neglecting all imaginary parts from Equation (3.9), we obtain the following linear equations as follows:

$$|V_2| = G_{22}^{-1} b^r, \tag{3.13}$$

where $|V_2|$ is voltage magnitudes, G_{22} is the conductance of admittance matrix Y_{22} and $b^r = -G_{21}|V_1|$. This makes the power flow computation roughly 50% faster and it might be worth the modeling error introduced by this assumption.

Reformulating equations with complex numbers

Matrix equation (3.9) is given as:

$$V_2^r + \imath V_2^m = (G_{22} + \imath B_{22})^{-1}(b^r + \imath b^m). \tag{3.14}$$

Equation (3.14) can be reformulated into into the following matrix equation:

$$\begin{bmatrix} V_2^r \\ V_2^m \end{bmatrix} = \begin{bmatrix} G_{22} & -B_{22} \\ B_{22} & G_{22} \end{bmatrix}^{-1} \begin{bmatrix} b^r \\ b^m \end{bmatrix}. \tag{3.15}$$

After the computation, original V_2 is calculated as $V_2 = V_2^r + \imath V_2^m$ using the computed real V_2^r and imaginary parts V_2^m in (3.15). In this case, we double the size of the equations but can avoid the complex numbering.

3.2.2. Modelling MV/LV transformers

To solve the entire network in a single simulation, the MV/LV transformers are modeled as an RL network [23] as displayed in Figure 3.2. Transformers are added to the impedance matrix $Z_\mathcal{E}$, as the link between the MV and LV network. In this model, an $\mathcal{E} \times \mathcal{E}$ diagonal scale matrix T is defined using the turns ratio t. For every cable behind the secondary side of the transformer, the corresponding value in T is the turns ratio of that transformer. For every cable in the medium voltage network, the corresponding value in the matrix T is 1. The impedance of the link is then scaled using $Z_p = T^2 Z_s$ where p, s denote the primary and secondary side of the transformer respectively. After the calculation, the turns ratio is used to re-scale the voltage and current to the LV-regime using $V_s = T^{-1} V_p$ and $I_s = T I_p$.

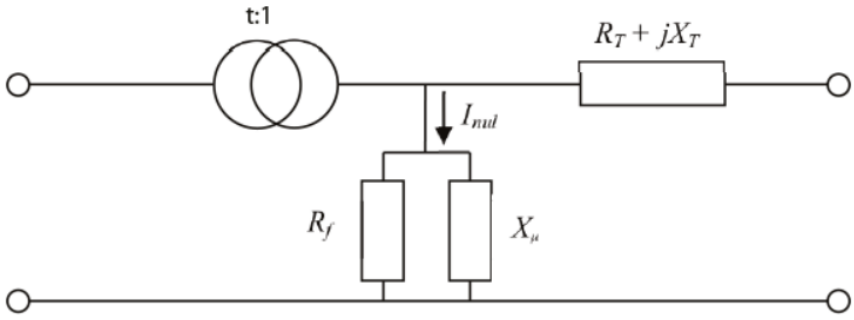


Figure 3.2: A schematic representation of the transformer model. The transformer is modeled as an RL network [23]. $R_T + jX_T$ denote the transformer's line impedance and R_f and X_μ denote the transformer's line-to-ground impedance.

3.3. Comparison between linear and nonlinear power flow problems

In this section, we compare the LPF approach using a constant impedance load model to the NPF methods using a constant power load model in terms of the accuracy and speed. The Newton power flow method developed in [24, 25] and a commercial network design software Vision [26], are used for NPF computations.

Schematic Figures of the studied networks can be found in Appendix A. Two small balanced distribution networks (Case33 and Case69: network details can be found in [27]) and a large balanced distribution network of Alliander DNO (case991) are considered for the comparison. Network Case991 is the MV network behind the substation Rauwerd, which supplies a significant part of Friesland, a province of The Netherlands. The MV network consists of 991 buses of which 237 are load buses. The network also has significant renewable generation, mostly in the form of wind generators. The same load model is used for the wind generators as for the regular loads, resulting in a negative constant power load in the case of the NPF and a negative constant impedance load in the LPF. The total peak load of this network is 10.3 MVA and the total feed-in peak power is 1.3 MVA. The nominal substation voltage (V_{ref}) is 10.5 kV.

3.3.1. Comparison to the Newton power flow method

The network currents and voltages have been calculated by applying the LPF and NPF methods on the same network with same properties such as reference voltage and network loads. Table 3.1 shows the total CPU time of both NPF and LPF computations and relative difference between nodal voltages where V^N and V^L are the computed nodal voltages of NPF and LPF computations respectively. The CPU time also includes data processing time. As shown in the table, the LPF computation is between 7 to 8 times faster than the NPF computation. In the last column of Table 3.1, the relative difference $\frac{||V^N - V^L||_2}{||V^N||_2}$ between LPF and NPF solutions, is given and

Table 3.1: The CPU time of NPF and LPF computations (Average of 10 computations) and relative difference

Test cases	CPU time (sec)		$\frac{\text{NPF (time)}}{\text{LPF(time)}}$	$\frac{\|\|V^N - V^L\|\|_2}{\|\|V^N\|\|_2}$
	LPF	NPF & iteration		
Case33	0.0005	0.0039 & 3 it	8.3946	5.87×10^{-4}
Case69	0.0006	0.0047 & 4 it	8.3585	0.0011
Case991	0.0016	0.0117 & 3 it	7.2025	7.91×10^{-6}

as we can see, the difference is very small. Figure 3.3 shows the voltage profile of two test cases and compares the results of NPF and LPF computations. The figure visually confirms that the LPF method can be as accurate as the NPF methods.

3.3.2. Comparison to a commercial power flow software Vision

To further validate the accuracy of the linear modeling approach, the algorithm is compared to a commercial network design tool Vision [26] which is the main software of all Dutch DNOs to dimension their MV networks and uses a constant power load model.

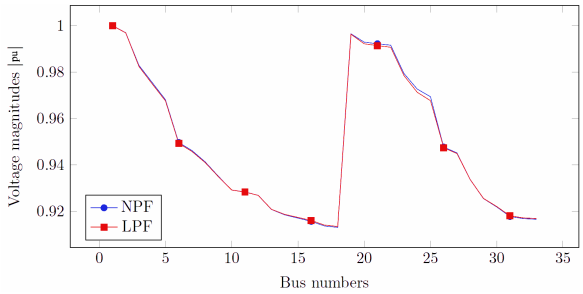
The results of the comparison can be found in Figure 3.4. The figure shows two scenarios; a high load scenario with 100% peak load and 0% feed-in and a high feed-in scenario with 25% peak load and 100% feed-in. The network is relatively heavy loaded because the edge of the network has a voltage drop of around 4%, where 4.5% is the allowed maximum as can be seen in the top-left histogram. It can be observed in middle column that the difference between the load assumptions is not more than 0.6%, which is close enough for network design and control purposes.

From both Figure 3.3 and Figure 3.4, we can see that the LPF approach using a 'constant impedance' load model can be as accurate as classical NPF methods with a 'constant power' load model. In addition, the LPF approach is much faster than NPF approach as we have seen in Table 3.1.

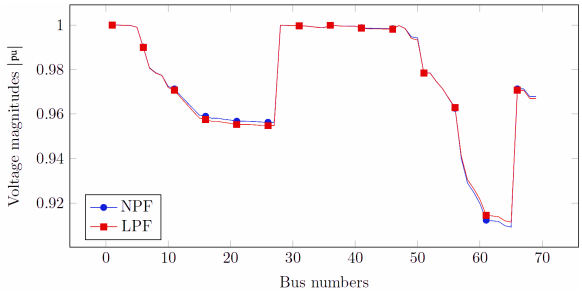
Therefore, this linear power flow approach can be a very powerful tool for electrical grid operators to control the very large networks in real time.

3.4. Case study of large Dutch power grid

In order to demonstrate the impact of integrally simulating the MV/LV grid, a case study has been assembled. The case study focuses on voltage problems. Since the voltage end points are the same for both the MV/LV and LV simulations, the results are easy to compare. The goal of the model is to support large-scale investment policy decisions such as: 'how many transformers will be overloaded the next 30 years?' or 'In which area of the country should more engineers be recruited for cable replacement?'. The model was created for techniques such as time series analysis and agent based modeling which all require evaluating many different load configurations. Several real world MV/LV networks have been studied in literature, some of which have in the order of 100,000 buses [28, 29]. However, the networks



(a) Case33



(b) Case69

Figure 3.3: Comparison of NPF and LPF computations for voltage magnitudes of all buses in two test networks Case33 and Case69.

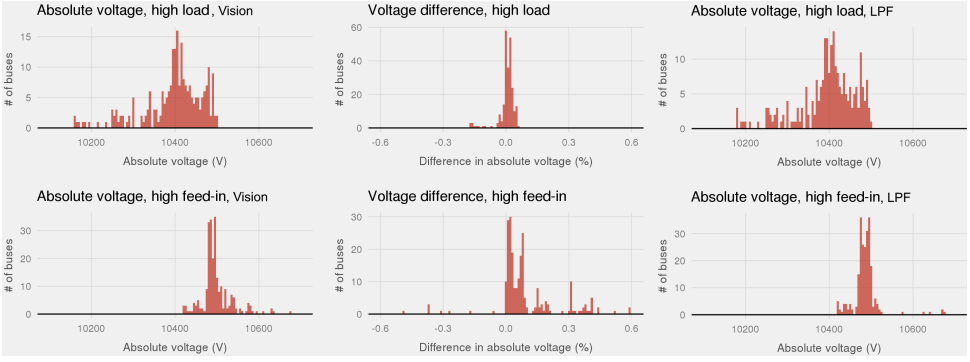


Figure 3.4: Histograms of the comparison between the LPF model and the commercial Vision network design tool of the network of substation Rauwerd. The left column shows a histogram of the absolute network voltages as calculated by the commercial design tool Vision., showing the severity of the voltage drop in the network compared to the nominal voltage of 10,500 Volt. The right column shows the same calculation performed with the LPF model. The middle histogram shows the difference between the two simulations.

from these studies are still several orders smaller than the network of this study which has over 22 million buses.

3.4.1. Data and assumptions

The network of Alliander DNO is used in our linear power flow computations, which contains both the LV and MV distribution networks and consists of approximately 80,000 km of cable serving over three million customers as shown in Figure 3.5. It covers over 1/3rd of the total Dutch power grid. The MV network of Alliander DNO consists of 100,000 cable segments whereas the LV network consists of over 22 million buses, three million customers (load nodes), several thousands of generators and around 250 HV/MV substations. Data sets consist of all cable segments, connectivity, and impedance. The connectivity, voltage ratio and impedance of all transformers are used in the power flow computation. The network is mostly radial, but some LV networks are strongly meshed and can serve over 100,000 customers.

In line with Alliander DNO policy, a voltage problem is defined as a voltage drop of over 9% in the MV/LV network from the secondary side of the HV/MV substation transformer to the customer, taking into account that both networks operate on a different voltage level. If only the LV network is considered, the allowed voltage drop is 4.5% from the secondary side of the distribution transformer to the customer. In the MV/LV simulation the nominal voltage (V_{ref}) is assumed to be 10.5 kV and in the LV simulation the nominal voltage (V_{ref}) is assumed to be 400 V.

Given that the linear model is only valid for relatively low voltage drops, it is important to note that voltage drops of 4.5% on the LV network and over 9% on the MV/LV network are always specified as a 'voltage problem' by the DNO. The exact height of the voltage problem is of less importance, as it needs to be solved anyway. The goal of the case study is therefore only to find the problems and not necessarily determining the problem severity.

The current presence of decentralized renewable energy generation is relatively low, around 4% of the total power generation. Generator buses are also modeled as load buses using the constant impedance model. While this is not accurate for the few buses controlling reactive power, it is in line with current Alliander DNO modeling practices.

The network is modeled as a single-phase balanced network, because no data is available of which customer is connected to which phase. While this is a best-case assumption, it is still a good starting point for finding voltage problems in the network and is only an issue for small LV networks with a little number of customers where the law of large numbers does not apply.

To run the linear power flow computation, all three million end users have been given a load of 1.1 kVA with a power factor of 0.95. The power consumption of 1.1 kW is the peak design power for regular households for LV grids containing over forty households.[1]

[1] While more detailed data is available within Alliander, it could not be used for publication purposes because of privacy issues. However, the 1.1 kVA assumption yields comparable results on locations with a sufficient number of customers.

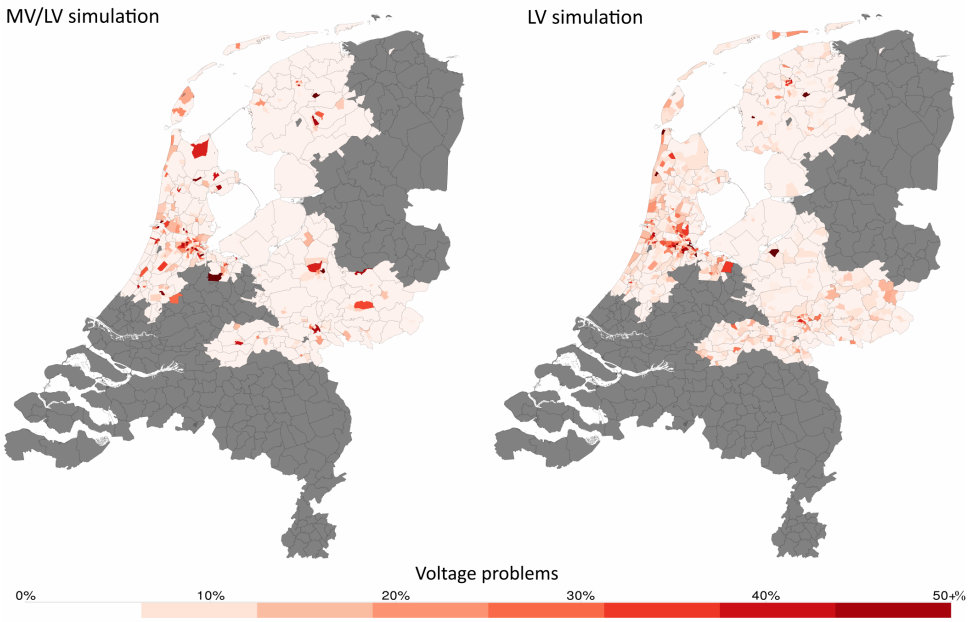


Figure 3.5: The geographical distribution of customers with voltage problems in the Alliander service area on a postal code level. The area depicted is the entirety of The Netherlands of which Alliander services the non-grey area. This figure has been obtained by computing the power flow of the entire Alliander network with an average load of 1.1 kVA per customer with a power factor of 0.95.

3.4.2. Simulation results

For this large power flow simulations on LV and MV/LV networks, we use a custom implementation of the LPF approach described in Section 3.2. The LLPF problem with complex components (3.9) is solved in 58 seconds on a single processor core using the *R* programming language. If only active power is considered, the problem is solved in 29 seconds. All linear algebra is implemented by the authors, using the methods from the 'Matrix' package (which is a C wrapper for the BLAS and LAPACK matrix computation libraries).

The resulting geographical distribution of voltage problems is displayed in Figure 3.5. Moreover, the absolute number of voltage problems in the MV/LV simulation is 150 thousand, 5% of the total number of customers. The absolute number of voltage problems in the LV simulation is 180 thousand, 6% of the total number of customers. While these percentages are low, they are still very significant as voltage problems can be quite costly to solve.

While the number of voltage problems are in the same order of magnitude between simulations, the locations of the problems are vastly different. The voltage problems overlapped only 20% between the two simulations as can also be observed in Figure 3.5. The lack of overlap has a severe implication, namely that searching for congestion by only simulating LV networks yields the wrong voltage problem locations. Therefore, it is clear that an identical load configuration will re-

sult in a very different layout of voltage problems if the MV/LV network is simulated integrally or only the LV network is taken into account.

It can also be observed from Figure 3.5 that the problems in the MV/LV simulation are more concentrated compared to the LV simulation due to the fact that a LV network with high loads influences neighboring networks via the MV network. This is very useful information for a DNO, since it also implies that multiple LV voltage problems can be solved by tactically strengthening the MV network.

A subject for future research is a comparison of the calculated problems in this case study with reported problems reported to the DNO. This is not trivial as a good comparison data set is not available. Voltage problems are an emerging issue and currently only very few voltage problems are actually detected by the DNO. This problem is also not easily solved using smart meter data. The smart meter only saves voltage and consumption data from the past 10 days, which is very little information to obtain a good peak consumption pattern. Furthermore, privacy rules and bandwidth limitations also do not allow for constant customer voltage monitoring.

Between the two simulations paradigms (LV only or MV/LV integrally), the integral MV/LV network simulation gives a more accurate estimate of the voltage problem locations as differences in the MV network are taken into account. This conclusion calls for network design using integral MV/LV simulations; a new design paradigm for the DNOs.

3.5. Application of numerical analysis techniques on the LLPF problem

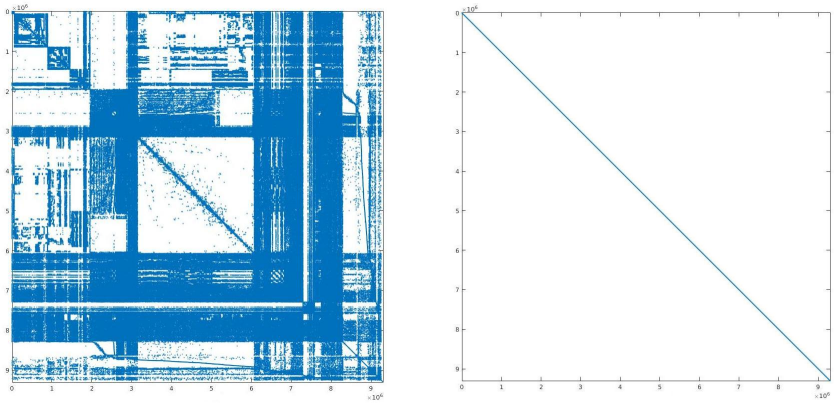
In this section, we show how NA techniques can be used to improve the CPU time of the LLPF computation introduced in Section 3.4. For this purpose, all solution methods are re-implemented in Matlab. We consider both LLPF problems with complex numbers (3.9) and without imaginary parts (3.13). For the numerical experiments, all computations are done on Intel computer i5-6500 3.2 GHz CPU with four cores and 64 GB memory.

3.5.1. LLPF problem with real components

Let us consider the LLPF problem with real components (3.13) where the size of matrix G_{22} is $9,300,775 \times 9,300,775$ and the number of nonzeros is $27,867,547$. Due to the large dimension of the matrix, it is very costly to compute the inverse of the matrix G_{22}^{-1}. Therefore, we study the properties of the matrix G_{22} and seek the fastest way to solve equation (3.13).

By analyzing matrix G_{22}, we observe that G_{22} is a sparse and Symmetric and Positive Definite (SPD) matrix. Due to its SPD properties, we can use NA techniques that are developed for this type of matrices such as the Cholesky decomposition, Incomplete Cholesky (IC), and the Conjugate Gradient (CG) iterative method on G_{22}. In addition, some reordering techniques such as Reverse Cuthill-McKee (RCM) and Approximate Minimum Degree (AMD) permutations could improve the properties of G_{22} as well. Figure 3.6 shows the sparsity structure of G_{22} and reordered G_{22} using

RCM. From the figure, it is clear that the sparseness properties of the matrix G_{22}



(a) Original G_{22} (b) Reordered G_{22}

Figure 3.6: Sparsity of matrix G_{22} and reordered G_{22} using RCM.

are improved by using RCM reordering.

For the direct solver, the Cholesky decomposition with RCM reordering could solve the linear system (3.13) fast. Since G_{22} is a SPD matrix, the best iterative method for matrix equation (3.13) is CG. Furthermore, the convergence rate of CG depends on eigenvalues λ_k of G_{22}. Table 3.2 shows the largest $\lambda_{\max}$ and smallest $\lambda_{\min}$ magnitude eigenvalues and the condition number $\kappa_2(\cdot) = \frac{\lambda_{\max}}{\lambda_{\min}}$ of G_{22} and preconditioned G_{22} as $M^{-1}G_{22}$. From the first row of Table 3.2, we see that the condition number of the matrix G_{22} is very large which means that G_{22} is ill-conditioned. Therefore, using CG without any preconditioner on the linear system (3.13) cannot improve the computational time since many iterations are required for CG. Thus, the Preconditioned Conjugate Gradient (PCG) method is a proper choice to use instead of CG. In PCG, we solve the transformed system as:

Table 3.2: The largest and smallest magnitude eigenvalues and the condition numbers for matrix G_{22} and preconditioned G_{22}

Matrix	$\lambda_{\max}$	$\lambda_{\min}$	$\kappa_2(\cdot)$
G_{22}	$9.34 * 10^{8}$	$1.38 * 10^{-3}$	$6.73 * 10^{12}$
$M^{-1}G_{22}$, $L = \text{Chol}$	1	1	1
$L = \text{IC}(10^{-5})$	1.70	0.01	123.43
$L = \text{IC}(10^{-6})$	1.28	0.27	4.72
$L = \text{IC}(10^{-7})$	1.15	0.83	1.38
$L = \text{IC}(10^{-8})$	1.002	0.994	1.008

$$M^{-1}G_{22}|V_2| = M^{-1}b^r \tag{3.16}$$

where M is called a preconditioner and is a SPD matrix. The eigenvalues of $M^{-1}G_{22}$ should be clustered around one, resulting in a faster convergence for PCG. Generally, M is obtained as $M = LL'$ where L is a lower triangular matrix. We can compute L using Cholesky or Incomplete Cholesky decompositions on G_{22} or on reordered G_{22}. The eigenvalues of $M^{-1}G_{22}$ can be improved by choosing a right preconditioner M for G_{22}.

In the second row of Table 3.2, the Cholesky decomposition is used for L and results in eigenvalues equal to one for the preconditioned G_{22}. Therefore, PCG with the Choleksy decomposition is expected to converge after one iteration for equation (3.13). However, using the full Cholesky decomposition for L is computationally expensive and the solution time can be larger than using a direct method. In order to decrease the computation time of constructing the lower triangular matrix L, we can use the Incomplete Cholesky decomposition instead of the full Cholesky.

In rows 3 - 6 of Table 3.2, we see how the eigenvalues and condition number of $M^{-1}G_{22}$ are improved by changing the drop tolerance of IC. Moreover, we can conclude that preconditioner M using IC(10^{-8}) or IC with a drop tolerance smaller than 10^{-8} for L can be a good preconditioner for matrix G_{22} in terms of the computational time and number of iterations for PCG.

Table 3.3 shows the comparison between various linear solvers on equation (3.13) in terms of the CPU time, number of iterations and the number of non-zeros (NNZ). All results are averaged over 10 computations. For PCG, the maximum iteration and relative tolerance are set to 100 and 10^{-5} respectively. Most of these methods have been applied to isolated non-linear PF problems. Table 3.3 contributes by comparing the methods both to the LPF and to each other.

Table 3.3: Comparison between various NA techniques on the LLPF problem with real components (3.13)

Algorithms	Time & Iter	$\frac{\|\|V_2^l - V_2^d\|\|_2}{\|\|V_2^d\|\|_2}$	NNZ
$-G_{22}\backslash b^r$	14.32 sec	$8.12 * 10^{-11}$	$27,867,547$
$G_{22}\backslash b^r$	7.12 sec	0	
+ RCM	6.94 sec	$6.69 * 10^{-12}$	
Cholesky	152.2 sec	$7.31 * 10^{-12}$	$257,293,316$
+ RCM	**5.01 sec**	$\mathbf{9.51 * 10^{-12}}$	$\mathbf{20,726,961}$
PCG(IC(0)) + RCM	NA	NA	$18,584,161$
PCG(Cholesky) + RCM	6.24 sec & 1 it	$9.51 * 10^{-12}$	$20,726,961$
PCG(IC(10^{-5})) + RCM	6.65 sec & 4 it	0.007	$19,722,635$
PCG(IC(10^{-8})) + RCM	**4.96 sec & 1 it**	$\mathbf{2.42 * 10^{-4}}$	$\mathbf{20,314,280}$

The first and second rows of Table 3.3 are the results of direct solvers using Matlab's backslash \ operator without any additional techniques. It is necessary to mention that the CPU time of the first row doubles the CPU time of second row due to the positioning of the minus sign in equation (3.7). In addition, if we write the minus sign on the left side of the equation (3.7), $-G_{22}$ is not a positive definite matrix which results in large computational time. Therefore, it is better to put the minus sign on the right side of equation (3.7) and to keep it inside the vector b.

For the direct solver, the Cholesky decomposition with RCM reordering results in the fastest computational time for matrix equation (3.13) as we can see from Table 3.3. Furthermore, as we expected, IC(10^{-8}) with RCM reordering is the best preconditioner for G_{22} that results in only one iteration in 4.96 seconds for PCG. However, when IC(10^{-8}) is used for the preconditioner, the relative difference between the direct and iterative solutions $\frac{||V_2^i - V_2^d||_2}{||V_2^d||_2}$ is high compared to other options. Therefore, we also solve the problem (3.13) with various tolerances for PCG and drop tolerances for IC. Numerical results are given in Table 3.4.

Table 3.4: Numerical results of PCG with various tolerances for both IC and PCG

Time & Iter & $\frac{\|V_2^i - V_2^d\|_2}{\|V_2^d\|_2}$		Relative tolerance for PCG	
		10^{-7}	10^{-8}
Drop tolerance for IC	10^{-9}	4.96 sec & 1 it & $2.40 * 10^{-5}$	4.96 sec & 1 it & $2.40 * 10^{-5}$
	10^{-10}	4.96 sec & 1 it & $2.31 * 10^{-6}$	4.96 sec & 1 it & $2.31 * 10^{-6}$

From Table 3.4, we see that the relative difference $\frac{||V_2^i - V_2^d||_2}{||V_2^d||_2}$ can be improved by decreasing the drop tolerance (10^{-9}, 10^{-10}, $\cdots$) of IC for the preconditioner $M = LL'$ while keeping PCG still converge after 1 iteration. Additionally, applying IC gives us smaller NNZ compared to full Cholesky and direct solvers.

Finally, the original computation time (14.32 sec) of LLPF problems with real components (3.13) is improved by 2.8 times (4.96 sec) using NA techniques.

3.5.2. LLPF problem with complex components

In this subsection, we consider the LLPF problem with complex components (3.9) and (3.15). For simplicity, let us denote the matrix

$$\begin{bmatrix} G_{22} & -B_{22} \\ B_{22} & G_{22} \end{bmatrix}$$

in equation (3.15) by A. Matrices Y_{22} and A are not positive definite unlike G_{22}. Moreover, matrix Y_{22} is symmetric and matrix A is non-symmetric. Therefore, the Cholesky decomposition and CG are not suitable for these types of matrices. Instead, the LU decomposition, GMRES and BiCGSTAB methods are more convenient to use on matrices Y_{22} and A. For iterative solvers, GMRES and BiCGSTAB, the maximum iteration and relative tolerance are set to 20 and 10^{-6} respectively.

Table 3.5 shows the comparison between various NA techniques on the LLPF problem (3.9) in terms of the CPU time, number of iterations and the relative difference between the direct and iterative solutions. In Alliander DNO, equation (3.15) is used to solve the LLPF problem because the *R* programming language does not support complex numbers. Furthermore, from the first and second rows of Table 3.5, we can see that using equation (3.9) to solve the LLPF problem with complex

components is almost 2.5 times faster than using equation (3.15) when Matlab's backslash \ operator is used without any additional techniques. Therefore, we use equation (3.9) for further experiments.

Table 3.5: Comparison between numerous NA techniques on the LLPF problem with complex components (3.9)

Algorithms	Time & Iter	$\frac{\|\|V_2^i - V_2^d\|\|_2}{\|\|V_2^d\|\|_2}$	NNZ
Eq. (3.15)	42.6 sec	0	111,470,118
Eq. (3.9): $Y_{22}\backslash b$	17.23 sec	$3.03 * 10^{-11}$	27,867,547
+ RCM	15.58 sec	$1.90 * 10^{-11}$	
LU + RCM	**7.41 sec**	$\mathbf{5.84 * 10^{-11}}$	**32,284,123**
GMRES(ilu(0)) + RCM	177.86 sec & 20 it	0.3427	27,867,547
BiCGSTAB(ilu(0)) + RCM	56.21 sec & 20 it	0.2503	
GMRES(ilu(10^{-8})) + RCM	18.75 sec & 2 it	$7.23 * 10^{-08}$	31,629,906
GMRES(ilu(10^{-11})) + RCM	13.78 sec & 1 it	$9.82 * 10^{-08}$	32,031,268
GMRES(ilu(10^{-14})) + RCM	14.27 sec & 1 it	$9.60 * 10^{-11}$	32,244,575
BiCGSTAB(ilu(10^{-10})) + RCM	10.57 sec & 0.5 it	$1.12 * 10^{-06}$	31,920,611
BiCGSTAB(ilu(10^{-12})) + RCM	10.77 sec & 0.5 it	$8.73 * 10^{-09}$	32,119,629
BiCGSTAB(ilu(10^{-14})) + RCM	10.92 sec & 0.5 it	$9.61 * 10^{-11}$	32,244,575

The same RCM reordering is applied to matrix Y_{22} in order to improve the structure of the matrix. The best computational time (7.41 sec) is achieved by the direct solver LU decomposition on the reordered Y_{22} using RCM as can be seen from Table 3.5. For the iterative methods, the best computation time with the smallest relative difference is obtained by BiCGSTAB with ILU(10^{-14}) as a preconditioner and RCM reordering. However, the best CPU time of the iterative method is still larger than the best CPU time of the direct solver due to the fact that ILU, GMRES, and BiCGSTAB are not implemented in the optimal way in Matlab. Furthermore, both LU and ILU decompositions provide relatively similar NNZ for the LLPF problem with complex components.

As a result of the application of NA techniques, the original computation time (42.6 sec) of LLPF problems with complex components (3.9) is improved by 5.7 times (7.41 sec).

3.6. Conclusion

In this chapter, we propose a fast LPF method using a constant impedance load model to simulate both the entire LV and MV networks in a single simulation. Mathematical modeling of power systems and transformers is given and the algorithm of the LPF approach is explained in detail. We validate the performance of this LPF approach by comparing it with the Newton power flow method and a commercial network design tool (Vision) on various distribution networks. Our results show that this LPF method can be as accurate as classical NPF methods using a constant power load model and additionally it is much faster than NPF computations. For the

largest test network, the entire LV and MV networks of Alliander DNO are used in our linear power flow computations. In our research, it is shown that voltage problems can be identified more accurately when MV and LV are integrally evaluated.

Moreover, NA techniques are applied to the LLPF problem in order to improve the computation time by studying the properties of the linear system. In the numerical analysis, reordering techniques (RCM) and numerous direct solvers (Cholesky, IC, LU, and ILU) and various Krylov subspace methods (CG, PCG, GMRES, and BiCGSTAB) are chosen and applied to the LLPF problem with both real and complex components. Finally, the original computation times of LLPF problems with real and complex components are reduced by 2.8 times and 5.7 times respectively as a result of the application of NA techniques.

The algorithms in this chapter are being applied within Alliander DNO. These applications include: large scale strategic modeling, automatic network design and automatic outage-recovery plans. For the subsequent research, this linear approach will be extended to unbalanced networks and applied to large scale state estimation using thousands of sensors.

Acknowledgment

The authors would like to thank Barbera Droste for providing the theory regarding adding the reactive power and transformer model to the LLPF. Thanks to Jaap van Wijck in performing the comparison of the LLPF to the commercial power flow solver.

References

[1] W. van Westering, B. Droste, and H. Hellendoorn, *Combined medium voltage and low voltage simulation to accurately determine the location of voltage problems in large grids,* in *2019 CIRED* (2019) p. 2.

[2] W. D. Stevenson, *Elements of power system analysis,* (1975).

[3] W. F. Tinney and C. E. Hart, *Power flow solution by Newton's method,* IEEE Transactions on Power Apparatus and systems , 1449 (1967).

[4] B. Stott and O. Alsaç, *Fast decoupled load flow,* IEEE transactions on power apparatus and systems , 859 (1974).

[5] S. C. Tripathy, G. D. Prasad, O. P. Malik, and G. S. Hope, *Load-flow solutions for ill-conditioned power systems by a newton-like method,* IEEE Transactions on Power Apparatus and Systems **PAS-101**, 3648 (1982).

[6] C. S. Cheng and D. Shirmohammadi, *A three-phase power flow method for real-time distribution system analysis,* Power Systems, IEEE Transactions on **10**, 671 (1995).

[7] M. Haque, *A general load flow method for distribution systems,* Electric Power Systems Research **54**, 47 (2000).

[8] V. M. da Costa, N. Martins, and J. L. R. Pereira, *Developments in the newton raphson power flow formulation based on current injections,* IEEE Transactions on Power Systems **14**, 1320 (1999).

[9] G. Chang, S. Chu, and H. Wang, *An improved backward/forward sweep load flow algorithm for radial distribution systems,* IEEE Transactions on power systems **22**, 882 (2007).

[10] J. A. Martinez and J. Mahseredjian, *Load flow calculations in distribution systems with distributed resources. A review,* in *Power and Energy Society General Meeting, 2011 IEEE* (IEEE, 2011) pp. 1–8.

[11] K. Balamurugan and D. Srinivasan, *Review of power flow studies on distribution network with distributed generation,* in *Power Electronics and Drive Systems (PEDS), 2011 IEEE Ninth International Conference on* (IEEE, 2011) pp. 411–417.

[12] U. Eminoglu and M. H. Hocaoglu, *Distribution systems forward/backward sweep-based power flow algorithms: A review and comparison study,* Electric Power Components and Systems **37**, 91 (2008).

[13] Y.-S. Zhang and H.-D. Chiang, *Fast Newton-FGMRES solver for large-scale power flow study,* IEEE Transactions on Power Systems **25**, 769 (2010).

[14] R. Idema, D. J. Lahaye, C. Vuik, and L. van der Sluis, *Scalable Newton-Krylov solver for very large power flow problems,* IEEE Transactions on Power Systems **27**, 390 (2012).

[15] D. Lahaye and K. Vuik, *Globalized Newton–Krylov–Schwarz AC load flow methods for future power systems,* in *Intelligent Integrated Energy Systems* (Springer, 2019) pp. 79–98.

[16] P. Schavemaker and L. van der Sluis, *Electrical Power System Essentials* (Wiley, 2008).

[17] J. R. Martí, H. Ahmadi, and L. Bashualdo, *Linear power-flow formulation based on a voltage-dependent load model,* IEEE Transactions on Power Delivery **28**, 1682 (2013).

[18] O. D. Montoya, L. Grisales-Noreña, D. González-Montoya, C. Ramos-Paja, and A. Garces, *Linear power flow formulation for low-voltage dc power grids,* Electric Power Systems Research **163**, 375 (2018).

[19] J.-H. Teng, *A direct approach for distribution system load flow solutions,* Power Delivery, IEEE Transactions on **18**, 882 (2003).

[20] W. van Westering and H. Hellendoorn, *Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation,* International Journal of Electrical Power & Energy Systems **114**, 105349 (2020).

[21] J. Kirtley, *6.061 Introduction to Power Systems Class Notes Chapter 5 Introduction To Load Flow* (MIT Open Courseware, 2018).

[22] W. H. Kersting, *Distribution system modeling and analysis* (CRC press, 2001).

[23] P. van Oirsouw, *Netten voor distributie van elektriciteit*, 2nd ed. (Phase to Phase B.V., 2011) Chap. 8.

[24] B. Sereeter, K. Vuik, and C. Witteveen, *Newton power flow methods for unbalanced three-phase distribution networks,* Energies **10**, 1658 (2017).

[25] B. Sereeter, C. Vuik, and C. Witteveen, *On a comparison of Newton-Raphson solvers for power flow problems,* Journal of Computational and Applied Mathematics **360**, 157 (2019).

[26] *Vision Network Analysis [application],* `http://www.phasetophase.nl/en_products/vision_network_analysis`, accessed: 07-12-2016.

[27] M. E. Baran and F. F. Wu, *Network reconfiguration in distribution systems for loss reduction and load balancing,* IEEE Transactions on Power Delivery **4**, 1401 (1989).

[28] A. B. Birchfield, T. Xu, and T. J. Overbye, *Power flow convergence and reactive power planning in the creation of large synthetic grids,* IEEE Transactions on Power Systems **33**, 6667 (2018).

[29] R. Bolgaryn, A. Scheidler, and M. Braun, *Combined planning of medium and low voltage grids,* 13th IEEE PowerTech Conference (2019).

4

Forecasting-based state estimation for three-Phase distribution systems with limited sensing

State Estimation is an essential technique to provide observability in power systems. Traditionally developed for high-voltage transmission networks, state estimation requires equipping networks with many real-time sensors, which remains a challenge at the scale of distribution networks. This chapter proposes a method to complement a limited set of real-time measurements with voltage predictions from forecast models. The method differs from the classical weighted least-squares approach, and instead relies on Bayesian estimation formulated as a linear least squares estimation problem. We integrate recently developed linear models for unbalanced 3-phase power flow to construct voltage predictions as a linear mapping of load predictions. The estimation step is a linear computation allowing high resolution state estimate updates, for instance by exploiting a small set of phasor measurement units. Uncertainties can be determined a priori and smoothed a posteriori, making the method useful for both planning, operation and post hoc analysis. The method is applied to an IEEE benchmark and on a real network testbed at the Dutch utility Alliander. An observability analysis suggests strategies for optimal sensor placement.

Parts of this chapter have been published in IEEE Transactions on Power Systems (2020). Co-author: Roel Dobbe

4.1. Introduction

4.1.1. Motivation

The operation of electric distribution networks is faced with new challenges due to the rapid adoption of distributed generation (DG) and electrification of our society, such as in driving and heating. The inherent intermittency of renewable generation combined with the diversification of demand make power flow more variable and harder to predict, leading to new protection issues, such as unintended islanding or tripping , and economic burden due to accelerated wear [1, 2]. To understand and mitigate these risks, many Distribution System Operators (DSOs) are building a stronger information layer on top of their physical infrastructure exploiting recent advances in sensing and communication. Firstly, by gathering historical data from SCADA and AMI systems to enable *forecasting* of demand, flow and voltage variables. Unfortunately, the increasing variability of power yields probability distributions with long tails, which cause forecasting methods to do poorly in situations when observability is most needed; when extreme and potentially dangerous events happen. Secondly, by applying traditional *state estimation* methods to distribution systems. This is conceptually similar, but technically and economically a more challenging task. Power system state estimation (SE) is the process of leveraging measurement from a subset of states in an electric network to estimate states that are not measured in real-time. In transmission systems, the need for system reliability and economies of scale have long motivated the development of state estimation methods [3, 4]. In the traditional setting, a state estimator relies on an *overdetermined* formulation for the unknown/unmeasured variables to be *observable*. This means that the number of available measurements must be greater than or equal to the number of unknowns (to be estimated). Ensuring observability requires DSOs to equip most buses in a network with real-time sensors and communication infrastructure, leading to steep investments that are hard to implement across all territories.

To overcome issues of scalability and prediction error, we combine forecasting based on historical load information with estimation using a limited number of real-time sensors. Our method focuses on estimating the voltage phasor at all buses with no sensor. Voltage forecasting will be done over a slower timescale using load data collected historically and modeling three-phase power flow with novel linear approximations [5]. We then use linear least squares estimation (LLSE) to update the forecasts in real-time using a small number of sensors with high temporal resolution. This is motivated by the recent introduction of synchrophasors for distribution systems, which allow the real-time assessment of dynamics [6, 7]. The updated forecasts will then constitute estimated voltage quantities at nodes without sensors at the same high temporal resolution. The method provides DSOs with a critical functionality to correct forecasted values when the system is deviating from the expected scenario, allowing detection and mitigation of risky scenarios. The LLSE method requires a new analysis of network observability and sensor placement, for which suggestions are given.

4.1.2. Previous work

In conventional state estimation, the measurements $z \in \mathbb{R}^{N_m}$ are expressed as a function of the quantities that are estimated $x \in \mathbb{R}^{N_n}$, by using power flow modeling:

$$z = h(x)\,. \tag{4.1}$$

The state estimation problem is then solved using a weighted least squares (WLS) problem:

$$x^* = \arg\min_x (z - h(x))^\top W (z - h(x))\,. \tag{4.2}$$

For the WLS problem to yield a meaningful result, Equation (4.1) needs to be *overdetermined*, which means that the number of measured variables needs to be greater than the number of estimated variables; $N_m > N_n$. A key challenge in distribution grids is to estimate an N_n-dimensional state vector in scenarios where only a limited set of sensors is available, i.e. $N_m < N_n$, which does not satisfy the requirements for conventional state estimation. As a result, the standard estimation problem is *underdetermined* and hence ill-posed from a computational point of view. In practice, this means that the state vector x is not *observable*. To overcome this inherent challenge, *pseudo-measurements* are typically used to augment real-time measurements in a weighted least squares (WLS) estimation algorithm. Pseudo-measurements are often calculated using load forecasts or historical data that tend to be less accurate than real-time measurements. Initial efforts considered augmenting an already fully observed measurement vector with extra load forecasts [8–10]. Later efforts tried to use a more limited number of real-time measurements with forecasts from Gaussian Mixture Models [11] or Artificial Neural Networks [12]. There have been made many contributions made to enable Distribution System State Estimation based on traditional WLS, and we refer the reader to [13] for a rigorous overview.

Unfortunately, the WLS method requires extensive tuning and is rather sensitive to errors and bad data [14]. Göl and Abur address this challenge through combining WLS with a least absolute value method that is more robust to error, yielding a *hybrid* estimator that combines a limited number of phasor measurement units (PMUs) with a high refresh rate (at the order of 30 Hz), and a fully observed ($M \geq n$) set of SCADA measurements at slower refresh rate (order of 5 to 15 min). The weighted least absolute value method used helps to robustify the the estimate between each SCADA update, but does not address a scenario where a limited set of measurements is available. Furthermore, the estimator is designed for transmission systems that can rely on robust communication networks. SCADA in distribution systems often lack a reliable communication infrastructure, which can lead to packet failure and unreliable state estimates. Other work by Schenato [15] and Weng [16] study the use of Bayesian estimation for SE, using load statistics to determine a *prior* probabilistic forecast $\hat{x}$ of state variables, which can be updated based on a limited set of real-time measurements. These papers shows their accuracy is comparable to that of conventional WLS estimators, and estimation error confidence intervals can be computed off-line, allowing for engineering trade-offs between number of sensors and estimation accuracy.

Contributions

This work forms a bridge between forecasting and full state estimation in three-phase distribution systems by embracing a Bayesian approach. Inspired by the development of linear approximations for unbalanced three-phase power flow [5], we derive a *closed-form analytic state estimator* that takes as its inputs load forecasting information, a network model and real-time measurements from a limited set of sensors. Our method estimates *voltage phasor differences* rather than on the absolute phasor. This reduces modeling and forecast errors and solves the issue of having to stop a numerical algorithm to solve power flow equations as done in [15]. This also circumvents the need for a reference voltage (typically the feeder head as in [15]) and allows the algorithm to be implemented in a distributed fashion for different parts of the network. In addition, the estimator in [15] models the voltage and current phasor in rectangular form, i.e. the complex voltage at node i is formulated as $\mathbb{V}_i = \text{Re}(\mathbb{V}) + j\text{Im}(\mathbb{V})$, where $j := \sqrt{-1}$. Since available synchrophasors typically report in polar form, i.e. $\mathbb{V}_i = V_i \angle \delta_i$, a nonlinear transformation is needed to adjust sensor readings before feeding these to the estimator: $\text{Re}(\mathbb{V}) = V_i \cos \delta_i$ and $\text{Im}(\mathbb{V}) = V_i \sin \delta_i$. This can lead to undesirable magnification of measurement errors in the voltage angle δ_i, which can be problematic for applications where angle estimates are used, for instance to close switches [5]. Our method formulates the problem in polar form, bypassing this source of error. Applied and assessed on a specific IEEE test feeder, we show that the method reduces average the error of forecasts by an average 60%, with more dramatic improvements for specific buses where forecasts are not able to perform appropriately. Lastly, we implement the full method on a utility testbed showcasing its applicability in real world circumstances.

Notation

We use $\|\cdot\|$ to denote the ℓ_2-norm, $(\cdot)^*$ to denote an optimal value, and $^\top$ stands for the transpose operator. $\mathfrak{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Throughout this work, we use the symbol $\circ$ to represent the Hadamard Product of two matrices (or vectors) of the same dimension, also known of the element-wise product, such that:

$$C = A \circ B = B \circ A \Rightarrow C_{ij} = A_{ij} B_{ij} = B_{ij} A_{ij}$$

where i indicates the row and j indicates the column of the vector or matrix.

4.2. Methodology

This chapter proposes a data-driven approach to do state estimation, that relies on minimum mean squares estimation (MMSE) [17]. MMSE is related to weighted least squares, but grounded in Bayesian principles and does not require an overdetermined measurement equation. Instead, our method relies on linear power flow models that enable us to express *voltage differences* (both magnitude and angle) throughout a network as a function of nodal load and generation. By expressing both the measured differences and the estimated differences as a function of the load, we are able to set up a linear least squares estimation (LLSE) problem, the

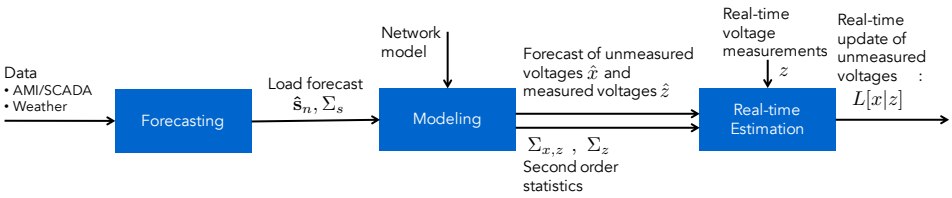


Figure 4.1: Overview of the forecasting and state estimation methodology.

linear version of MMSE. The LLSE has an analytical solution that can update the forecast of non-measured voltage differences by comparing the measured voltage differences against their forecasted values. As such, the method reminds of the Kalman Filter [18], which is a repeated execution of LLSE problems taking into account the dynamic evolution of the state variables. The approach comes with a trade-off, as the quality of the updates depends on the number of sensors and their placement in the network. This is discussed in Section 4.7.

The MMSE approach enables an end-to-end pipeline from historical load and network data to voltage forecasting to updating these forecasts in real-time using a limited set of sensors. The methodology is depicted in Figure 4.1. The three main steps of forecasting, modeling and real-time estimation are developed in Sections 4.3, 4.4.2 and 4.5.3 respectively. Before we dissect these steps we first cover the sources of uncertainty in state estimation and we introduce MMSE.

4.2.1. Sources of information and uncertainty for state estimation

The practical reality of DSOs is that they face many sources of uncertainty in trying to construct a state estimator, outlined in Figure 4.2. Following the proposed construction of the state estimator as depicted in Figure 4.1, the overall accuracy of the available information for state estimation depends on three sources of uncertainty: accuracy of the forecasted quantities, of the modeling procedure and of the quantities measured in real-time.

The forecast may be based on a DSO's historical data, which can include SCADA data of network variables, advanced metering infrastructure (AMI) readings of household consumption (or an anonymized/aggregated version of these), data of distributed generation and storage, public weather data (temperature, humidity, solar irradiance). These data sources typically do not form a perfect representation to forecast all the necessary quantities in the network. Certain nodes may not be recorded, the recordings may be noisy or miss certain data points, and the sampling rate of the recordings may be lower than the anticipated rate for updating the state estimator. In the modeling step, inaccuracy arises from parameter data that is outdated or measured with noise or the use of approximations (such as linear power flow) to enable efficient computation. Lastly, for the actual estimation, we rely on a limited number of sensors, which may be subject to measurement noise and could have various sampling rates.

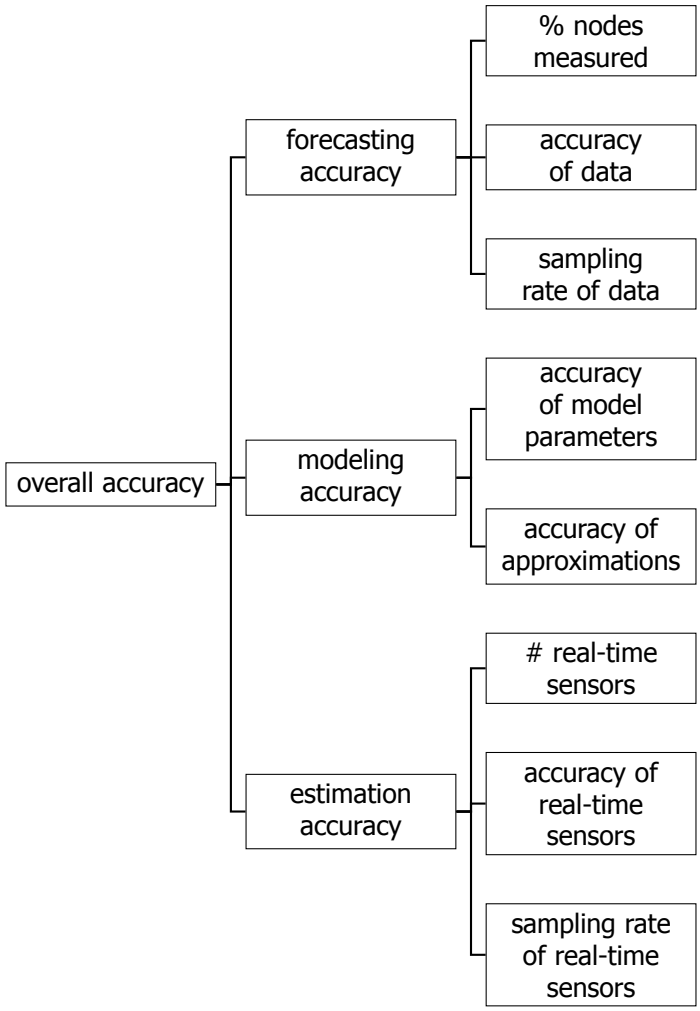


Figure 4.2: Sources of uncertainty affecting the accuracy of the state estimator

4.2.2. Introduction to minimum mean square estimation

Consider the context of having a set of voltage phasor measurements $Z \in \mathbb{R}^{N_m}$ and an unobserved random variable $X \in \mathbb{R}^{N_n}$, representing all non-measured voltage phasors. We aim to determine an estimate of X based on Z that is close to X in some sense. Assume we are given a joint distribution of (X, Z). We want to find an estimator $\hat{X} = g(Z)$ that minimizes the mean square error $E[\|X - \hat{X}\|^2]$. One can show that the minimum mean squares estimate (MMSE) of X given Z is equivalent to the conditional expectation $\hat{X} = E[X|Z]$, [17].

We consider the case in which both the estimator and the measurements are linear in a shared set of variables for which distributions are available, in our case in the form of load statistics. Let (X, Z) be vectors of random variables on some probability space. It turns out that the estimator minimizing the mean square error is also linear in the measurements, i.e. the *linear least squares estimator (LLSE)* denoted $L[X|Z] = a + BZ$ is a linear function of Z, with $a \in \mathbb{C}^{N_n}$ and $B \in \mathbb{C}^{N_n \times N_m}$ attained by the following optimization problem:

$$\min_{a,B} E\left[\|X - a - BZ\|^2\right]. \tag{4.3}$$

Following [17], it can be shown that

$$a^* = E[X] - \Sigma_{X,Z}\Sigma_Z^{-1}E[Z]\ ,\ B^* = \Sigma_{X,Z}\Sigma_Z^{-1}, \tag{4.4}$$

where $\Sigma_{X,Z} \in \mathbb{R}^{N_n \times N_m}$ and $\Sigma_Z \in \mathbb{R}^{N_m \times N_m}$ denote the cross-covariance matrix of X and Z and covariance matrix of Z. Hence, if Σ_Z is nonsingular, the LLSE is given by

$$L[X|Z] = E[X] + \Sigma_{X,Z}\Sigma_Z^{-1}(Z - E[Z])\,. \tag{4.5}$$

Lastly, if Σ_Z is singular, then Σ_Z^{-1} can be replaced by the left pseudo-inverse $\Sigma_Z^{\dagger}$. Interpreting (4.5), $(Z - E[Z])$ represents a deviation of the actual measurement Z from its expected value $E[Z]$, which is called an *innovation*. This innovation triggers the Bayesian estimator $L[X|Z]$ to propose an update of the forecast $E[X]$ by a linear scaling through the covariance matrices. Alternatively, $L[X|Z] = g(Z)$ can be interpreted as a *projection* of X onto the set of affine functions of Z.

The LLSE has a number of important benefits. Firstly, it has an analytical closed-form solution that can be used to neatly integrate real-time measurements Z and forecast information (as we will see in Section 4.3). Secondly, it is not necessary to explicitly calculate the Bayesian posterior probability density function over X, because $L[X|Z]$ only depends on the first two moments of X and Z. Thirdly, it works for many distributions $(X, Z) \sim \mathfrak{D}$, as long as $\mathfrak{D}$ has well defined first and second moments [17]. Lastly, the number of measurements N_m does not need to be larger than the number of to-be-estimated states N_n, which is the most significant difference with other ubiquitous estimation schemes such as weighted least squares and Gauss-Markov estimation that do not work for $N_m < N_n$. The main challenge of any MMSE approach is understanding what information is lost in the projection that happens in Equation (4.5) through the mapping $\Sigma_{X,Z}\Sigma_Z^{-1}$. For our state estimation method this requires revisiting the notion of network observability, typically defined for situations where $N_m > N_n$, which we will do in Section 4.7, based on the full formulation of the estimator.

4.3. Forecasting

We consider the design of a machine learning model to forecast the mean μ_s and covariance matrix Σ_s of the load $\boldsymbol{s}$, which are then used to forecast the mean and covariance of the voltage magnitude and phase. In practical contexts, DSOs may not have access to voltage or load readings from AMI in real-time, but it is possible that historical readings are used, in combination with other predictive covariates, to predict load values for a future time.

Machine learning models have been used in a variety of ways to predict load values [19]. Two relevant examples are auto-regressive moving average (ARMA) models for short term load forecasting and data-driven modeling of physical systems that utilizes regression trees to predict loads, with notable benefits to both. An ARMA model is able to capture trends in previous datapoints [20]. ARMA models are often not practical in distribution operation, since the AMI data is mostly not available in real-time, preventing the use of recent load values. A regression tree model is able to cluster data based on certain characteristics, such as day of the week, temperature, and humidity [21]. Its interpretability makes it useful in contexts where operators need to make decisions based on a model's predictions.

In our setting, the MMSE estimator defined in Section 4.2.2 necessitates the input of a point estimate of the load and its covariance matrix. This requirement motivates the use of Gaussian Processes (GPs), which offer both mean and variance information [22]. A GP is are also flexible in that they can have continuous ARMA features as well as dicrete features as its inputs. GPs have previously been used in similar applications for short term load forecasting to predict maximum daily loads [23]. Using GPs does introduce some bias, as load distributions tend to be non-Gaussian, though typically near-unimodal. In our setting, this bias can be compensated by the estimation step. Using a more sophisticated method to retrieve first and second moment information from historical data is left as future work.

Let $\mathcal{N}$ denote a set of buses Nodes are indexed by $n = 0, 1, \dots, N-1$, where N is the order (number of nodes) of the distribution feeder, and node 0 denotes the feeder head (or substation). For each node $n \in \mathcal{N}$, we start with a data set of historical readings of inputs $X_n = \{x_n[t] \in \mathcal{X}_n\}_{t=1}^{T}$ and load values $S_n = \{\boldsymbol{s}_n[t] \in \mathcal{Y}_n\}_{t=1}^{T}$. The inputs consist of real-valued and discrete-valued features. We consider the following real-valued features at time t:

$$\begin{bmatrix} l_n[t-1] \cdots l_n[t-k] & d_n[t-1] \cdots d_n[t-k+1] & \theta[t] & \eta[t] \end{bmatrix}, \tag{4.6}$$

where $l_n[t]$ denotes the load value for bus n at time t, $d_n[t]$ the difference in load between time $t-1$ and t, and θ_t and η_t are the temperature and humidity at time t. Note that a typical distribution feeder SCADA system often does not have access to load measurement, and hence the features l and d may only be available historically or in real-time for only a subset of the buses. Hence, we also consider discrete-valued features representing date and time:

$$\begin{bmatrix} DST & MOY & BD & DOW & HOD & MOH \end{bmatrix}, \tag{4.7}$$

which respectively denote an indicator for daylight saving time, month of year, an indicator for business day, day of week, hour of day and minute of hour.

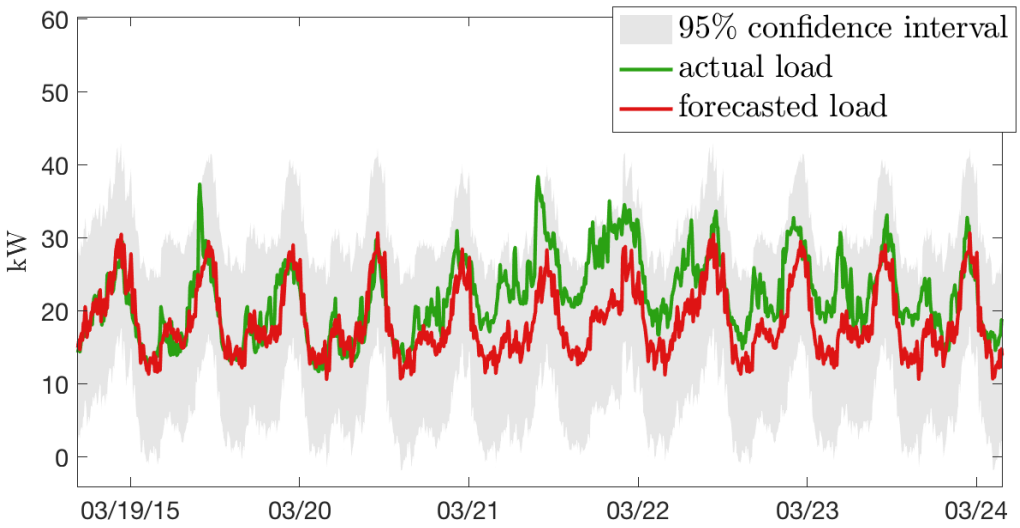


Figure 4.3: Forecast of an aggregate load using a Gaussian Process model with only discrete-valued time features. Only 10% of the loads in the aggregate were recorded in historical data. The other 90% of loads were imputed with the average load profile. Poor forecast performance, such as on March 21st, motivates the use of Bayesian estimation.

We now want to train a function $f_n : \mathcal{X}_n \to \mathcal{Y}_n$ with data that best predicts $\boldsymbol{s}_n[t]$ at some time t based on an input with accessible inputs $x_n[t]$. A GP defined on an input space $\mathcal{X}_n$ can be formulated as

$$f_n(x) = \boldsymbol{\phi}_n(x_n)^\top \beta_n + g_n(x)\,, \tag{4.8}$$

where $g_n(x)$ is a zero-mean GP represented as $\mathcal{GP}(0, k_n(x_n, x_n))$, with kernel $k_n(x_n, x_n)$ modeling the covariance across the input space $\mathcal{X}_n$. $\boldsymbol{\phi}_n(x_n)^\top \beta_n$ determines the translation of the GP from the origin, with $\boldsymbol{\phi}_n(x_n)$ a feature basis for the output given the input vector x_n, β_n are learned coefficients or weights for the basis features [22, Section 2.2]. Given this framework, we can model the distribution of an output at a certain input x_n^*:

$$f(x_n^*) \mid x_n^*, X_n, S_n \sim \mathfrak{N}(\boldsymbol{\phi}_n(x_n^*)^\top \beta_n + g_n(x_n^*), \sigma^2)\,, \tag{4.9}$$

The primary assumption under GPs is that it models a collection of random variables, any finite number of which have a joint Gaussian distribution. Notice that there are two different variances in the system – $k_n(x_n, x_n)$ and σ^2. The first variance, $k_n(x_n, x_n)$ is the variance on the estimate induced by the covariance of the input features as defined by a covariance function. σ^2 is the noise variance of the data as a whole. To challenge the method, in Section 4.6, we consider a GP model that is based on a poor historical data set and no access to real-valued features. Figure 4.3 exemplifies the resulting forecast accuracy, motivating the use of Bayesian estimation to account for forecast errors such as those experienced on March 21st.

4.4. Power flow modeling

Let $\mathcal{T} = (\mathcal{N}, \mathcal{E})$ denote a graph representing a radial distribution feeder, where $\mathcal{N}$ is the set of nodes of the feeder and $\mathcal{E}$ is the set of line segments. Nodes are indexed by $n = 0, 1, \ldots, N-1$, where N is the order (number of nodes) of the distribution feeder, and node 0 denotes the feeder head (or substation). We treat node 0 as an infinite bus, decoupling interactions in the downstream distribution system from the rest of the grid. We also consider a set of nodes equipped with sensors $\mathcal{M} \subset \mathcal{N}$.

4.4.1. Single-phase power flow

In our prior work [24], we constructed the state estimator for single-phase power flow, using the well studied DistFlow equations [25]:

$$\begin{aligned} P_m &= p_m^c - p_m^g + \sum_{n:(m,n)\in\mathcal{E}} P_n + r_{mn}\ell_{mn}\,, \\ Q_m &= q_m^c - q_m^g + \sum_{n:(m,n)\in\mathcal{E}} Q_n + x_{mn}\ell_{mn}\,, \\ V_m^2 - V_n^2 &= 2(r_{mn}P_n + x_{mn}Q_n) + (r_{mn}^2 + x_{mn}^2)\ell_{mn}\,, \end{aligned} \tag{4.10}$$

where r_{mn} and x_{mn} are the resistance and reactance of the line between nodes m and n, $P_m, Q_m \in \mathbb{R}$ denote the real and reactive power flowing on a branch into node m, $p_m^c, q_m^c \in \mathbb{R}$ are the real and reactive power consumption at node m, $p_m^g, q_m^g \in \mathbb{R}$ the real and reactive generation at node m, v_m is the voltage magnitude at node m, and ℓ_{mn} is the squared magnitude of the current on branch (m,n). In our setting we might also have access to the voltage angle δ_m. The nonlinear equation for voltage angle difference over a branch is given by

$$\sin(\delta_m - \delta_n) = \frac{x_{mn}P_n - r_{mn}Q_n}{v_m v_n}\,. \tag{4.11}$$

We propose the approximation that assumes that $\ell_{mn} \approx 0$, the voltage magnitudes $v_m \approx 1$ p.u. and $|\delta_m - \delta_n| \ll 1$. If we also define $Y_m \triangleq V_m^2$, we end up with the linear equations

$$\begin{aligned} P_m &\approx p_m^c - p_m^g + \sum_{n:(m,n)\in\mathcal{E}} P_n\,, \\ Q_m &\approx q_m^c - q_m^g + \sum_{n:(m,n)\in\mathcal{E}} Q_n\,, \\ Y_m - Y_n &\approx 2r_{mn}P_n + 2x_{mn}Q_n\,, \\ \delta_m - \delta_n &\approx x_{mn}P_n - r_{mn}Q_n\,. \end{aligned} \tag{4.12}$$

Notice that these are the linearized DistFlow equations [25, 26], augmented with an update equation for voltage angle differences. The latter equation was studied recently motivated by the development of PMUs for distribution grids [7].

4.4.2. Three-phase power flow

More recently, a linear approximation power flow in unbalanced three phase distribution networks was developed [5, 27, 28]. This model can be thought of as an extension of the *DistFlow* model [25] to unbalanced circuits, and was coined the *Dist3Flow* model. In this setting, each node and line segment can have up to three phases, labeled a, b, and c. Phases are referred to by the variables ϕ and ψ, where $\phi \in \{a, b, c\}$, $\psi \in \{a, b, c\}$. If line (m, n) exists, its phases must be a subset of the phases present at both node m and node n.

The current/voltage relationship for a three phase line (m, n) between adjacent nodes m and n is captured by Kirchhoff's Voltage Laws (KVL) in its full (4.13), and compact form (4.14):

$$\begin{bmatrix} V_m^a \\ V_m^b \\ V_m^c \end{bmatrix} = \begin{bmatrix} V_n^a \\ V_n^b \\ V_n^c \end{bmatrix} + \begin{bmatrix} Z_{mn}^{aa} & Z_{mn}^{ab} & Z_{mn}^{ac} \\ Z_{mn}^{ba} & Z_{mn}^{bb} & Z_{mn}^{bc} \\ Z_{mn}^{ca} & Z_{mn}^{cb} & Z_{mn}^{cc} \end{bmatrix} \begin{bmatrix} I_{mn}^a \\ I_{mn}^b \\ I_{mn}^c \end{bmatrix} \tag{4.13}$$

$$\mathbf{V}_m = \mathbf{V}_n + \mathbf{Z}_{mn}\mathbf{I}_{mn} \tag{4.14}$$

Here, $Z_{mn}^{\phi\psi} = r_{mn}^{\phi\psi} + jx_{mn}^{\phi\psi}$ denotes the complex impedance of line (m, n) across phases ϕ and ψ.

Next, we define the per phase complex power as $\mathbf{S}_{mn}^{\phi} = \mathbf{V}_n^{\phi}\left(\mathbf{I}_{mn}^{\phi}\right)^*$, and the 3×1 vector of complex power phasors $\mathbf{S}_{mn} = \mathbf{V}_n \circ \mathbf{I}_{mn}^*$ where $\mathbf{S}_{mn}$ is the power from node m to node n at node n.

$$\sum_{l:(l,m)\in\mathcal{E}} \mathbf{S}_{lm} = \mathbf{s}_m + \sum_{n:(m,n)\in\mathcal{E}} \mathbf{S}_{mn} + \mathbf{L}_{mn} \tag{4.15}$$

The term $\mathbf{L}_{mn} \in \mathbf{C}^{3\times 1}$ is a nonlinear and non-convex loss term. As in [29] and [25], we assume that losses are negligible compared to line flows, so that $\left|L_{mn}^{\phi}\right| \ll \left|S_{mn}^{\phi}\right| \ \forall (m, n) \in \mathcal{E}$. Thus, we neglect line losses, linearizing (4.15) into (4.16).

$$\sum_{l:(l,m)\in\mathcal{E}} \mathbf{S}_{lm} \approx \mathbf{s}_m + \sum_{n:(m,n)\in\mathcal{E}} \mathbf{S}_{mn} \tag{4.16}$$

Now, we define the real scalar $y_m^{\phi} = \left|V_m^{\phi}\right|^2 = V_m^{\phi}(V_m^{\phi})^*$, the 3×1 real vector $\mathbf{y}_m = \left[y_m^a, y_m^b, y_m^c\right]^T = \mathbf{V}_m \circ \mathbf{V}_m^*$. With these definitions, [5] derives the following equations that govern the relationship between squared voltage magnitudes and complex power flow across line (m, n):

$$\begin{aligned} &\mathbf{y}_m = \mathbf{y}_n + 2\mathbf{M}_{mn}\mathbf{P}_{mn} - 2\mathbf{N}_{mn}\mathbf{Q}_{mn} + \mathbf{H}_{mn} \\ &\mathbf{M}_{mn} = \Re\left\{\Gamma_n \circ \mathbf{Z}_{mn}^*\right\}, \mathbf{N}_{mn} = \Im\left\{\Gamma_n \circ \mathbf{Z}_{mn}^*\right\}, \end{aligned} \tag{4.17}$$

where $\Gamma_n = \mathbf{V}_n\left(1/\mathbf{V}_n\right)^T \in \mathbb{C}^{3\times 3}$ represents a matrix with voltage balance ratios across all phases at node n. Hence, we have that $\Gamma_n(\phi, \phi) = 1$ and $\Gamma_n(\phi, \psi) =$

$V_n^{\phi}/V_n^{\psi} \triangleq \gamma_n^{\phi\psi}$. Futhermore, $\mathbf{H}_{mn} = (\mathbf{Z}_{mn}\mathbf{I}_{mn})\circ(\mathbf{Z}_{mn}\mathbf{I}_{mn})^* = (\mathbf{V}_m - \mathbf{V}_n)\circ(\mathbf{V}_m - \mathbf{V}_n)^*$ is a 3×1 real-valued vector representing higher-order terms. Notice that we have separated the complex power vector into its active and reactive components, $\mathbf{S}_{mn} = \mathbf{P}_{mn} + j\mathbf{Q}_{mn}$.

This nonlinear and nonconvex system is difficult to incorporate into a state estimation or optimization formulation without the use of convex relaxations. Following the analysis in [29], we apply two approximations. The first is that the higher order term $\mathbf{H}_{mn}$, which is the change in voltage associated with losses, is negligible, such that $\mathbf{H}_{mn} \approx [0,\ 0,\ 0]^T\ \forall(m,n) \in \mathcal{E}$. The second assumes that node voltages are "nearly balanced" (i.e. approximately equal in magnitude and 120° apart). This is only applied to Γ_n in the RHS of (4.17), such that $\gamma_n^{ab} = \gamma_n^{bc} = \gamma_n^{ca} \approx \alpha$, and $\gamma_n^{ac} = \gamma_n^{ba} = \gamma_n^{cb} \approx \alpha^2$ for all $n \in \mathcal{N}$. Under these assumptions, we retrieve

$$\Gamma_n = \begin{bmatrix} 1 & \gamma_n^{ab} & \gamma_n^{ab} \\ \gamma_n^{ba} & 1 & \gamma_n^{bc} \\ \gamma_n^{ca} & \gamma_n^{cb} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha^2 & 1 & \alpha \\ \alpha & \alpha^2 & 1 \end{bmatrix} \forall n \in \mathcal{N}\,, \tag{4.18}$$

where $\alpha = 1\angle 120° = \frac{1}{2}(-1 + j\sqrt{3})$ and $\alpha^{-1} = \alpha^2 = \alpha^* = 1\angle 240° = \frac{1}{2}(-1 - j\sqrt{3})$. Note that we make the "nearly balanced" assumption in the process of the formal derivation as in [29, 30], but that does not imply that node voltages need to actually be perfectly balanced for the linearizion to be valid. Applying the approximations for $\mathbf{H}_{mn}$ and Γ_n to (4.17), we arrive at a linear system of equations:

$$\mathbf{y}_m \approx \mathbf{y}_n + 2\mathbf{M}_{mn}\mathbf{P}_{mn} - 2\mathbf{N}_{mn}\mathbf{Q}_{mn}\,,\ \text{with} \tag{4.19}$$

$$2\mathbf{M}_{mn} = \ldots$$
$$\begin{bmatrix} 2r_{mn}^{aa} & -r_{mn}^{ab} + \sqrt{3}x_{mn}^{ab} & -r_{mn}^{ac} - \sqrt{3}x_{mn}^{ac} \\ -r_{mn}^{ba} - \sqrt{3}x_{mn}^{ba} & 2r_{mn}^{bb} & -r_{mn}^{bc} + \sqrt{3}x_{mn}^{bc} \\ -r_{mn}^{ca} + \sqrt{3}x_{mn}^{ca} & -r_{mn}^{cb} - \sqrt{3}x_{mn}^{cb} & 2r_{mn}^{cc} \end{bmatrix}, \tag{4.20}$$

$$2\mathbf{N}_{mn} = \ldots$$
$$\begin{bmatrix} -2x_{mn}^{aa} & x_{mn}^{ab} + \sqrt{3}r_{mn}^{ab} & x_{mn}^{ac} - \sqrt{3}r_{mn}^{ac} \\ x_{mn}^{ba} - \sqrt{3}r_{mn}^{ba} & -2x_{mn}^{bb} & x_{bc} + \sqrt{3}r_{mn}^{bc} \\ x_{mn}^{ca} + \sqrt{3}r_{mn}^{ca} & x_{mn}^{cb} - \sqrt{3}r_{mn}^{cb} & -2x_{mn}^{cc} \end{bmatrix}. \tag{4.21}$$

The linear approximation in [5] also enables a linear mapping for voltage angles, similar to Equation (4.19). In the rest of this chapter, we will focus on voltage magnitude and leave the extension to voltage angles as an exercise.

4.5. Real-time estimation

In this Section, we construct the state estimator based on linear least squares estimation. This method takes in a prior distribution on measured and unmeasured voltage variables, and updates this in real-time with a limited set of measurements. To do so, we require the prior statistics of the voltage based on load forecasts (Section 4.3) and power flow modeling (Section 4.4.2). We first express measured

and unmeasured voltage variables as a linear function of the net load. We can then construct the necessary matrices to express the voltage forecast as function of load statistics.

4.5.1. Voltage as a function of net load

Consider the vector with all the differences in squared voltage magnitude stacked with the differences in voltage angles over all the branches (i.e. for every set of adjacent nodes) in the network,

$$\Delta \boldsymbol{y} \triangleq \begin{bmatrix} \mathbf{y}_0 - \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{N-1} - \mathbf{y}_N \end{bmatrix} \in \mathbb{R}^{3N}. \tag{4.22}$$

With Equation (4.19), we can build a model for all the voltage differences over wires throughout the network

$$\Delta \boldsymbol{y} \quad = 2 \left[\text{blkdiag}(\mathbf{M}_{ij}) \ \text{blkdiag}(\mathbf{N}_{ij}) \right] \begin{bmatrix} \text{vec}(\mathbf{P}_{ij}) \\ \text{vec}(\mathbf{Q}_{ij}) \end{bmatrix} = \boldsymbol{Z}_b \boldsymbol{S} \ , \tag{4.23}$$

where $\boldsymbol{S} \in \mathbb{R}^{6N}$ is the vector with real and reactive branch flows stacked vertically, and $\boldsymbol{Z}_b \in \mathbb{R}^{3N \times 6N}$ is a horizontal stack of two block diagonal matrices with the corresponding 3-by-3 matrices from respectively (4.20) and (4.21). With Equation (4.16), we can express the branch flows $\boldsymbol{S}$ in terms of the nodal net loads, which yields$\boldsymbol{S} = \mathcal{P}_b \boldsymbol{s}$, with $\boldsymbol{s} \in \mathbb{R}^{6N}$ a vector with the nodal net loads, real and reactive power p_n, q_n, $n \in \mathcal{N}$ stacked vertically, and $\mathcal{P}_b \in \mathbb{R}^{6N \times 6N}$ a binary matrix in which a row represents a branch with 1s selecting the nodes downstream of the branch. We have now expressed the differences in voltage magnitude over all N lines in terms of the nodal load vector,

$$\Delta \boldsymbol{y} = \boldsymbol{Z}_b \mathcal{P}_b \boldsymbol{s} \triangleq \boldsymbol{Z}_n \boldsymbol{s}, \tag{4.24}$$

where $\boldsymbol{Z}_n \triangleq \boldsymbol{Z}_b \mathcal{P}_b \in \mathbb{R}^{3N \times 6N}$.

Measured quantities

In our actual setting, we do not directly measure voltage differences over all individual wires. Instead, we place the sensors over a distance spanning multiple branches and buses. The voltage difference over the path can be rewritten as the sum of the individual differences of the branches lying on the path. For example, for a later network with 3 buses, the squared voltage magnitude difference over the path from bus 0 through bus 1 to bus 2 can be expressed as

$$\begin{aligned} \mathbf{y}_0 - \mathbf{y}_2 &= (\mathbf{y}_0 - \mathbf{y}_1) + (\mathbf{y}_1 - \mathbf{y}_2) \\ &= 2(\mathbf{M}_{01}\mathbf{P}_1 + \mathbf{N}_{01}\mathbf{Q}_1) + 2(\mathbf{M}_{12}\mathbf{P}_2 + \mathbf{N}_{12}\mathbf{Q}_2) \\ &= 2\mathbf{M}_{01}\mathbf{p}_1 + 2\mathbf{N}_{01}\mathbf{q}_1 + 2(\mathbf{M}_{01} + \\ &\quad \mathbf{M}_{12})\mathbf{p}_2 + 2(\mathbf{N}_{01} + \mathbf{N}_{12})\mathbf{q}_2 \end{aligned}$$

Notice that we can form this equation by doing a row operation on Equation (4.24), i.e. adding up rows 1 and 2 of Z_n. Imagine that we have a set of differences that refer to the placement of our sensors,

$$\Delta \boldsymbol{y}_m \triangleq \begin{bmatrix} \mathbf{y}_{m_2} - \mathbf{y}_{m_1} \\ \vdots \\ \mathbf{y}_{m_M} - \mathbf{y}_{m_{M-1}} \end{bmatrix} \in \mathbb{R}^{3(M-1)}, \tag{4.25}$$

with $m_1, \ldots, m_M \in \mathcal{M}$. We can now formulate the equations, by adding up the differences of all individual lines in between the sensors. I.e., we can formulate a permutation matrix such that $\Delta \boldsymbol{y}_m = \mathcal{P}_m \Delta \boldsymbol{y}$, and hence

$$\Delta \boldsymbol{y}_m = \mathcal{P}_m \boldsymbol{Z}_n \boldsymbol{s} = \boldsymbol{Z}_m \boldsymbol{s}, \tag{4.26}$$

where $\boldsymbol{Z}_m \triangleq \mathcal{P}_m \boldsymbol{Z}_n = \mathcal{P}_m \boldsymbol{Z}_b \mathcal{P}_b \in \mathbb{R}^{3(M-1)\times 6N}$. This gives us an expression for the measured quantities as a function of the nodal load vector.

Non-measured quantities - voltage estimation

We are interested to estimate voltage magnitude and angle at all the $N - M$ buses in the network that are not equipped with a sensor. We aim to do this given a measurement of the voltage phasor at a limited number of M buses in the network, and forecast statistics on the nodal load vector $\boldsymbol{s}$. We consider the differences in voltage between a location we want to estimate and a nearby sensor location. These differences are collected in a vector $\Delta \boldsymbol{Y}_e$ to be estimated as a function of the load vector $\boldsymbol{s}$, similar to the construction of the measurement equation:

$$\Delta \boldsymbol{y}_e = \boldsymbol{Z}_e \boldsymbol{s} \in \mathbb{R}^{3(N+1-M)}, \tag{4.27}$$

where $\boldsymbol{Z}_e \triangleq \mathcal{P}_m \boldsymbol{Z}_n = \mathcal{P}_e \boldsymbol{Z}_b \mathcal{P}_b \in \mathbb{R}^{3(N+1-M)\times 6N}$ is constructed in the same way as $\boldsymbol{Z}_m$ in (4.26). In order to retrieve an estimate of the absolute voltage value, we can simply take the nearest sensor reading and add/subtract the estimated difference between the location and that sensor location.

4.5.2. Voltage forecast statistics

We now have that our measurements are voltage phasor differences, i.e. $z = \Delta \boldsymbol{y}_m$ and the estimation quantities are other voltage phasor difference, i.e. $x = \Delta \boldsymbol{y}_e$. Given the linear relationships with the load vector $\boldsymbol{s}$, we can now derive the statistics on z. The mean of z is

$$\begin{aligned} \mu_z(t) &= E(\Delta \boldsymbol{y}_m), \\ &= E(\boldsymbol{Z}_m \boldsymbol{s}), \\ &= \boldsymbol{Z}_m E(\boldsymbol{s}), \\ &= \boldsymbol{Z}_m \mu_s(t). \end{aligned} \tag{4.28}$$

Similarly, we have that $\mu_x(t) = E(\Delta \boldsymbol{y}_e) = Z_e \mu_s(t)$. The covariance of z is

$$\begin{aligned} \Sigma_z(t) &= E((z - \mu_z)(z - \mu_z)^\top), \\ &= E((\boldsymbol{Z}_m \boldsymbol{s} - \boldsymbol{Z}_m \mu_s(t))(\boldsymbol{Z}_m \boldsymbol{s} - \boldsymbol{Z}_m \mu_s(t))^\top), \\ &= \boldsymbol{Z}_m E((\boldsymbol{s} - \mu_s(t))(\boldsymbol{s} - \mu_s)^\top) \boldsymbol{Z}_m^\top, \\ &= \boldsymbol{Z}_m \Sigma_s(t) \boldsymbol{Z}_m^\top. \end{aligned} \tag{4.29}$$

Similarly, we have that the cross-covariance of x and z is $\Sigma_{x,z}(t) = \boldsymbol{Z}_e \Sigma_s(t) \boldsymbol{Z}_m^\top$. This yields all the statistics we need to construct the distribution grid state estimator.

4.5.3. Constructing the state estimator

We can now analytically derive the LLSE of $\boldsymbol{y}_e$ given measurements $\boldsymbol{y}_m$, as a specific form of Equation (4.5) presented in Section 4.2.2. For our voltage estimation setting this yields

$$\begin{aligned} L[\Delta \boldsymbol{y}_e | \Delta \boldsymbol{y}_m] \quad &= E(\Delta \boldsymbol{y}_e) + \dots \\ &\quad \Sigma_{\Delta \boldsymbol{y}_e, \Delta \boldsymbol{y}_m} \Sigma_{\Delta \boldsymbol{y}_m}^{-1} (\Delta \boldsymbol{y}_m - E(\Delta \boldsymbol{y}_m)), \\ &= \boldsymbol{Z}_e \mu_s + \dots \\ &\quad \boldsymbol{Z}_e \Sigma_s \boldsymbol{Z}_m^\top \left(\boldsymbol{Z}_m \Sigma_s \boldsymbol{Z}_m^\top \right)^{-1} (\Delta \boldsymbol{Y}_m - \boldsymbol{Z}_m \mu_s), \end{aligned} \tag{4.30}$$

where we dropped the time index for brevity. With $L[\Delta \boldsymbol{y}_e | \Delta \boldsymbol{y}_m]$, the voltage estimates can be retrieved as

$$\hat{\boldsymbol{V}}_e = \sqrt{\boldsymbol{y}_{\text{near}} + L[\Delta \boldsymbol{y}_e | \Delta \boldsymbol{y}_m]}, \tag{4.31}$$

where $\boldsymbol{V}_e$ denotes a stacked vector with voltages for all buses without measurement, and $\boldsymbol{y}_{\text{near}}$ are the squared voltages at the nearest measured bus for each estimated bus.

Notice that (4.30) is written in the form $\Delta \hat{\boldsymbol{y}}_e = f(\Delta \boldsymbol{y}_m)$, as all of the other information needed to evaluate the estimator are forecast statistics, which are known a priori. As such, (4.30) uses the statistical information of the net loads $\boldsymbol{s}$, in combination with the topology and impedance information of the network, to present a closed-form analytical estimator of voltage differences throughout the network. This function is linear in the measurements and efficiently computed in real-time.

4.6. Results

4.6.1. Synthetic experiments

Earlier work implemented the distribution grid state estimator on a single-phase radial network [24]. To validate the estimator on a three-phase network, we used a modified version of the IEEE 37 bus distribution feeder model [31], depicted in Figure 4.4. The feeder voltage and power ratings were left unchanged (4.8 kV and 2.5 MVA), as were line segment configuration assignments. We ignored the transformer at node 775 and the voltage regulator at the feeder head. We assumed all loads were constant power. The data used in this experiment are from datasets provided by Pecan Street for educational use [32]. The raw data contained 15-minute-interval data sampled from July 1, 2013 to September 26, 2016. We aggregated different household time series from the Pecan Street data set such that the aggregated time series data had a spot load marginally less than the 3-phase real and reactive spot loads defined by the IEEE feeder model [31]. The aggregated time series were then used to build a Gaussian Process forecast model for real and reactive power at each bus, as outlined in Section 4.3. Voltage sensors were placed at nine different buses, indicated by red hexagons in Figure 4.4. Figure 4.5 shows the result for one time instance. To assess the overall performance, we

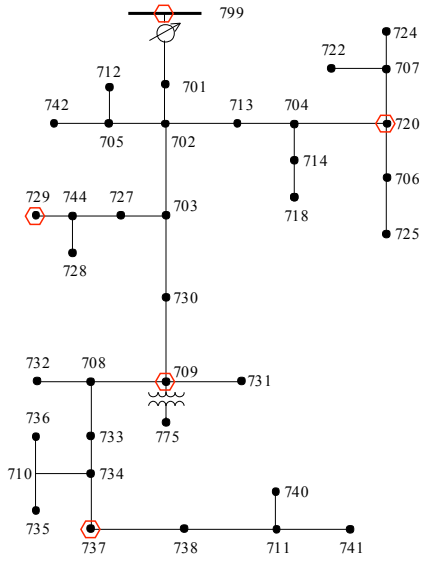


Figure 4.4: IEEE 37 node test feeder model, voltage sensors are indicated with red hexagons.

compute the Average Root Mean Square Error (ARMSE) on the voltages $\boldsymbol{V}_e$ that are not measured,

$$\text{ARMSE}(\{\hat{\boldsymbol{V}}_e[t]\}_{t=1}^{T}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{V}}_e[t] - \boldsymbol{V}_e[t]\|^2}\,. \tag{4.32}$$

Figure 4.6 shows the ARMSE metric for all buses. It is bounded by 0.2 p.u. for the forecasted values and 0.02 p.u. for the estimated values. Notice that buses with higher forecast errors benefit significantly from the estimation procedure. Buses that already have a proper accuracy on forecasted values of the order < 0.01 p.u. do not necessarily gain much from estimation. This can be attributed to the fact that these errors are in the same order as the modeling errors due to linear approximation, which are carefully studied in a separate paper [5].

4.6.2. Validation experiment on a utility testbed

Here, we apply the method to a network in the territory of *Alliander*, the largest Distribution Network Operator (DNO) of the Netherlands serving over three million customers. Alliander is experimenting with community electricity storage in Rijsenhout a suburban village close to Amsterdam, the Netherlands. The project is called "BuurtBatterij" which translates to "Neighborhood or Community Battery". Figure 4.7 depicts the Rijsenhout feeder that houses the battery project. One of the goals of the community battery experiments is to assess and improve the accuracy of the available network simulation models. As a part of the community battery

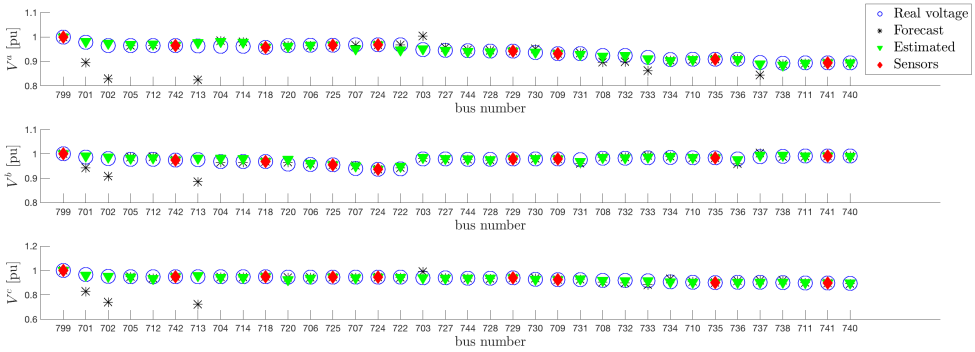


Figure 4.5: Example voltage profile with forecast and estimation update at all the buses, numbered as in Figure 4.4.

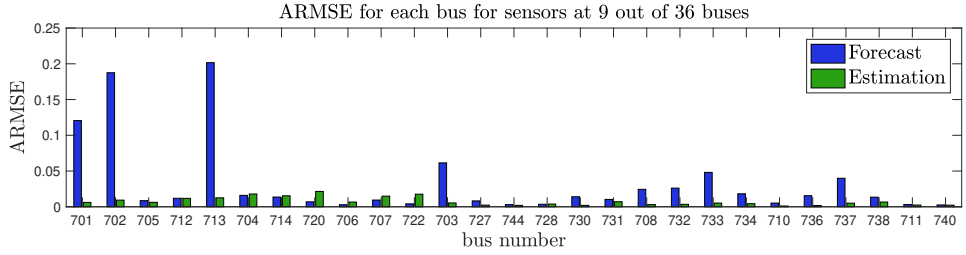


Figure 4.6: ARMSE in p.u. for each non-measured bus across all phases. Buses with higher forecast errors benefit significantly from the estimation procedure. Buses that already have good accuracy on forecasted values of the order < 0.01 p.u. do not necessarily gain much from estimation. This can be attributed to the fact that these errors are in the same order as the modeling errors due to the estimator's linear approximation.

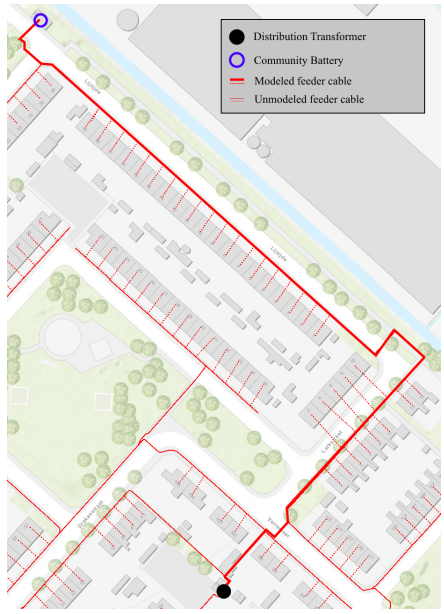


Figure 4.7: GIS view of Alliander's low voltage network of Rijsenhout. The outlined modeled network is the network that is considered for the DGSE model. The modeled part of the network consists of 34 customers. The unmodeled cables are not physically connected to the modeled network. At the distribution transformer and the community battery both power and voltage are measured. At 12 households, the power was measured. For privacy reasons, their exact location could not be displayed, but they are almost uniformly distributed along the cable.

experiments, the local low voltage power grid is modeled and measurement data is gathered.

We apply the state estimation procedure to the network, relying on a feeder model and real load and voltage measurements. The feeder contains 142 buses, of which 34 are regular household customers, one is the distribution transformer and one is the community battery. The other buses are network cable joints. The source of the network data is the Alliander GIS database, which contains the exact location and properties of the electricity cables. However, the GIS database does not contain on which phase each customer is connected, therefore the estimator is constructed using a balanced single phase model, using the formulation in Section 4.4.1. The distribution transformer is located at the top of the feeder, and the Neighborhood Battery is installed at the end of the feeder. Both the transformer and battery contain SCADA equipment for measuring power and voltage at a 1-second rate. Of the 34 households connected to this feeder, 12 customers share their power consumption data with Alliander as part of the community battery project. All data for building forecasts have been collected at a 1-minute resolution. Customers with no direct measurement were assigned the residual power load, which was defined as the total transformer load minus the sum of all measured loads. Each unmeasured customer was assigned an equal proportional share of the residual

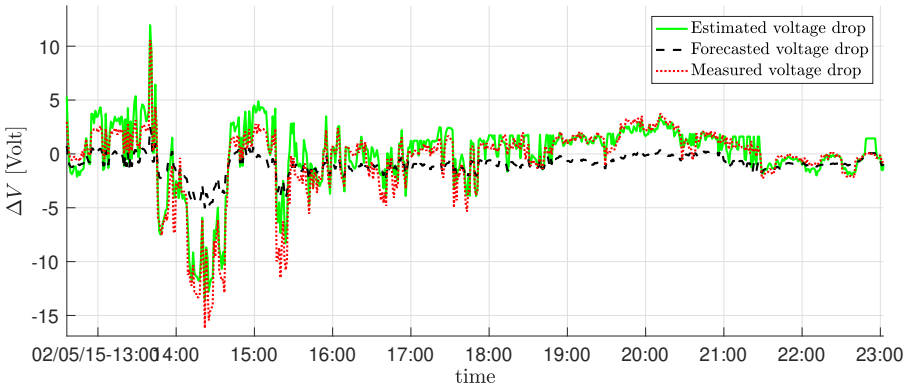


Figure 4.8: Comparison of forecasted and estimated voltage with real voltage measurement at the Neighborhood Battery bus.

load. Note that this introduces some error in the forecast procedure.

Figure 4.8 shows the results of applying SE and comparing the predicted and estimated voltage drop at a particular bus with real voltage measurements. Observe that the estimated values provide a significant improvement over the forecasted values, showing agreement with the actual values. The improvements are particularly strong for larger voltage deviations, providing critical information for safety procedures. At certain times the estimation does not improve accuracy, which has two explanations. Firstly, for smaller voltage deviations, modeling errors due to linearization of power flow are more dominant, as mentioned above. Secondly, the effect of limited real-time voltage sensors (in this case only 2 out of 140 buses) provides significant but limited improvement due to limited observability of all load flow scenarios in the network. This challenge requires revisiting the notion of network observability, which is discussed in Section 4.7. Similar to the IEEE synthetic experiment, SE significantly reduces the ARMSE across all buses in the network, on average by 60%. Given the difficulty of predicting the power consumption of individual households due to their variability, this result is useful for DSOs in improving the fidelity of their forecasting data with limited sensing capabilities, which is a likely context in most networks for the foreseeable future. As such, Alliander is implementing SE algorithms in their critical calculations, and aim to use the data for optimal sensor placement, cable health monitoring, real time overload predictions, and control of voltage and power flow.

4.7. Suggestions for observability analysis and sensor placement

Network observability characterization for state estimation using the weighted least squares (WLS) approach (as in (4.2)) was derived by Monticelli and Wu for settings that assume the DC approximation [33]. Gómez Expósito and Abur proposed an

approach for general nonlinear measurement equations, which involves taking the first order Taylor approximation and can be used to include current magnitude measurements [34].

In contrast to the conventional methods, our state estimator does not require solving a WLS problem. Instead of using the Taylor approximation for a fully nonlinear power flow model, we have expressed both our measurements $\Delta \boldsymbol{y}_m$ and our estimation variables $\Delta \boldsymbol{y}_e$ as a linear function of the load vector $\boldsymbol{s}$. Assuming we have access to a load forecast μ_s for all nodes in the network, we can argue that the voltage forecast $\boldsymbol{Z}_e \mu_s$ itself is well-defined and provides full *prior observability*; given statistical information for all loads, the mapping from load forecast to voltage forecast is well-posed.

As covered in Section 4.2.2, our estimator is a LLSE which is equivalent to *projecting* the estimation variable $\Delta \boldsymbol{y}_e$ onto the set of linear functions of the measurement $\Delta \boldsymbol{y}_m$, which can be interpreted as the best linear unbiased estimator, assuming the linear power flow model is unbiased. The projection is a result of the assumption that the estimation step considers a limited number of sensors $M < N$, which in the context of network theory means this step will never be able to capture all changes in the estimation variables. That said, it is possible to determine a sensor placement that allows the measurements $\Delta \boldsymbol{y}_m$ to capture a maximum amount of information about the estimation variables $\Delta \boldsymbol{y}_e$.

Definition 4.7.1. *A load profile $\boldsymbol{s} \in \mathbb{R}^N$ is observable if $\boldsymbol{s} \in \mathcal{R}\left(\boldsymbol{Z}_m^\top\right)$ (row space of the measurement matrix). For any $\boldsymbol{s} = \boldsymbol{s}_o + \boldsymbol{s}_u$, with $\boldsymbol{Z}_m \boldsymbol{s}_u = 0$ and $\boldsymbol{s}_o \in \mathcal{R}\left(\boldsymbol{Z}_m^\top\right)$, such that $\boldsymbol{Z}_m \boldsymbol{s} = \boldsymbol{Z}_m \boldsymbol{s}_o$, we say that $\boldsymbol{s}_o, \boldsymbol{s}_u$ are the observable and unobservable parts of the load profile.*

Our aim now is to design an estimator that minimizes

$$\begin{aligned} \Delta \boldsymbol{y}_e - L[\Delta \boldsymbol{y}_e | \Delta \boldsymbol{y}_m] &\approx \boldsymbol{Z}_e \boldsymbol{s} - \left(\boldsymbol{Z}_e \mu_s + \boldsymbol{Z}_e (\boldsymbol{s}_o^m - \mu_{s,o}^m)\right) \\ &= \boldsymbol{Z}_e (\boldsymbol{s} - \boldsymbol{s}_o^m) + \boldsymbol{Z}_e (\mu_{s,o}^m - \mu_s) \\ &= \boldsymbol{Z}_e (\boldsymbol{s}_u^m - \mu_{s,u}^m), \end{aligned} \tag{4.33}$$

with respect to some metric over all relevant load scenarios. Here, $\boldsymbol{s}_o^m$ and $\boldsymbol{s}_u^m$ are the observable and unobservable parts of the load profile $\boldsymbol{s}$, with respect to the mapping $\Sigma_s \boldsymbol{Z}_m^\top \left(\boldsymbol{Z}_m \Sigma_s \boldsymbol{Z}_m^\top\right)^{-1} \boldsymbol{Z}_m$. Notice that as Σ_s and $\left(\boldsymbol{Z}_m \Sigma_s \boldsymbol{Z}_m^\top\right)^{-1}$ are both full rank square matrices, the null space of the mapping is fully characterized by the null space $\mathcal{N}(\boldsymbol{Z}_m)$. A desired property is for the part of the load profile that is unobservable in the measurements to be insignificant, or equivalently also unobservable in the estimated variables, and thereby in the null space of the estimation matrix $\boldsymbol{Z}_e$. This means that whatever information is lost by the projection by measurement matrix $\boldsymbol{Z}_m$ does not contribute to changes in the *actual* values of the estimation variables $\Delta \boldsymbol{y}_e$. In mathematical terms, we hence may want the null spaces of $\boldsymbol{Z}_m$ and $\boldsymbol{Z}_e$ to intersect as much as possible. This can be formulated as the following optimization problem:

$$\min_{\boldsymbol{Z}_e, \boldsymbol{Z}_m} \dim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{Z}_e \\ \boldsymbol{Z}_m \end{bmatrix}\right) - \dim \mathcal{N}(\boldsymbol{Z}_m) \,. \tag{4.34}$$

Note that $\boldsymbol{Z}_e, \boldsymbol{Z}_m$ are both determined by the sensor placement. Alternatively, given a data set Ξ of historical load profiles, we can formulate a data-driven sensor placement approach which minimizes

$$\min_{\boldsymbol{Z}_e, \boldsymbol{Z}_m} \sum_{\xi \in \Xi} \|\boldsymbol{Z}_e (I - \mathcal{P}_{\boldsymbol{Z}_m}) \xi\|^2 , \tag{4.35}$$

where $\mathcal{P}_{\boldsymbol{Z}_m} = \boldsymbol{Z}_m^\top (\boldsymbol{Z}_m \boldsymbol{Z}_m^\top)^{-1} \boldsymbol{Z}_m$ is a projection matrix. Equation (4.35) should be read as trying to minimize the extent to which the parts of all historical load profiles that are unobservable with respect to $\boldsymbol{Z}_m$ affect the value of $\Delta \boldsymbol{y}_e$. This approach allows the DSO to prioritize important load flow scenarios that are more safety-critical, by weighting these differently, yielding

$$\min_{\boldsymbol{Z}_e, \boldsymbol{Z}_m} \|\boldsymbol{Z}_e (I - \mathcal{P}_{\boldsymbol{Z}_m}) \Xi W\|_F^2 , \tag{4.36}$$

where W is a diagonal weight matrix. The above sensor placement strategies are here presented as suggestions. A rigorous analysis of these is left as future work.

4.8. Conclusion

This chapter addressed the challenge of formulating a distribution grid state estimator, for scenarios where fully observed sensor arrangements are not yet feasible, and load forecasts may be subject to large uncertainties due to lack of access to data. Building on preliminary work for single-phase networks, we derived an algorithm that exploits the information in load forecasting and feeder models to construct prior statistics of relevant voltage variables. We then used a Bayesian approach, in the form of the linear least squares estimator, to update prior voltage statistics in real-time based on measured deviations at a limited set of voltage sensors. We applied the method to a benchmark IEEE network and on a real testbed in the Netherlands and showed its ability to provide accurate voltages estimates using limited historical data and real-time sensors. As such, the method is highly applicable in the typical distribution network setting in which data and sensing will remain limited for the foreseeable future.

Acknowledgments

We thank Michael Sankur for contributions to the code base for this work, Sascha von Meier for helpful suggestions in the earlier stage of this research, and colleagues at Alliander for enabling the experiments for real world integration.

References

[1] N. Seltenrich, *The new grid - plugging into california's clean-energy future,* (2013).

[2] G. Kaur and M. Vaziri, *Effects of distributed generation (DG) interconnections on protection of distribution feeders,* in *IEEE Power Engineering Society General Meeting, 2006*.

[3] A. Abur and A. Gomez Exposito, *Power System State Estimation: Theory and Implementation*, Power Engineering (Willis), Vol. 24 (CRC Press).

[4] G. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. Wollenberg, *Monitoring and optimization for power grids: A signal processing perspective,* IEEE Signal Processing Magazine **30**, 107.

[5] M. D. Sankur, R. Dobbe, A. v. Meier, E. M. Stewart, and D. B. Arnold, *Optimal voltage phasor regulation for switching actions in unbalanced distribution systems,* in *2020 IEEE Power Energy Society General Meeting (PESGM)* (IEEE, 2020) pp. 1–5.

[6] B. Xu and A. Abur, *Observability analysis and measurement placement for systems with PMUs,* in *IEEE PES power systems conference and exposition*, Vol. 2 (Citeseer) pp. 943–946.

[7] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, *Micro-synchrophasors for distribution systems,* in *IEEE 5th Innovative Smart Grid Technologies Conference, Washington, DC*.

[8] P. Rousseaux, T. Van Cutsem, and T. E. Dy Liacco, *Whither dynamic state estimation,* International Journal of Electrical Power & Energy Systems **12**, 104.

[9] E. Blood, B. Krogh, and M. Ilic, *Electric power system static state estimation through kalman filtering and load forecasting,* (Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century IEEE, 2008).

[10] Y. Xiang, *Operation of future medium voltage distribution grids: Application of statistical methods for state estimation and fault location*, Ph.D. thesis, Eindhoven University of Technology (2015).

[11] R. Singh, B. C. Pal, and R. A. Jabr, *Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement,* IET Generation, Transmission & Distribution **4**, 50.

[12] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, *Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling,* IEEE Transactions on Power Systems **27**, 1888.

[13] A. Primadianto and C. N. Lu, *A review on distribution system state estimation,* IEEE Transactions on Power Systems **32**, 3875.

[14] M. Gol and A. Abur, *A hybrid state estimator for systems with limited number of PMUs,* IEEE Transactions on Power Systems **30**, 1511.

[15] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. Von Meier, *Bayesian linear state estimation using smart meters and PMUs measurements in distribution grids,* in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 572–577.

[16] Y. Weng, R. Negi, and M. D. Ilic, *Probabilistic joint state estimation for operational planning,* IEEE Transactions on Smart Grid **Vol. PP**, 10.1109/TSG.2017.2749369.

[17] J. Walrand, *Probability in Electrical Engineering and Computer Science: An Application-Driven Course*, 1st ed., Vol. 1 (Quoi, 2014).

[18] R. E. Kalman, *A new approach to linear filtering and prediction problems,* Journal of basic Engineering **82**, 35 (1960).

[19] P. Mirowski, S. Chen, T. Kam Ho, and C.-N. Yu, *Demand forecasting in smart grids,* Bell Labs technical journal **18**, 135.

[20] S.-J. Huang and K.-R. Shih, *Short-term load forecasting via ARMA model identification including non-gaussian process considerations,* IEEE Transactions on Power Systems **18**, 673.

[21] M. Behl, A. Jain, and R. Mangharam, *Data-driven modeling, control and tools for cyber-physical energy systems,* (ACM/IEEE, 2016).

[22] C. E. Rasmussen, *Gaussian processes in machine learning,* in *Advanced lectures on machine learning* (Springer, 2004) pp. 63–71.

[23] H. Mori and M. Ohmi, *Probabilistic short-term load forecasting with gaussian processes,* (2005).

[24] R. Dobbe, D. Arnold, S. Liu, D. Callaway, and C. Tomlin, *Real-time distribution grid state estimation with limited sensors and load forecasting,* in *7th International Conference on Cyber-Physical Systems (ICCPS)* (ACM/IEEE).

[25] M. E. Baran and F. F. Wu, *Optimal sizing of capacitors placed on a radial distribution system,* Power Delivery, IEEE Transactions on **4**, 735 ().

[26] M. Baran and F. Wu, *Optimal capacitor placement on radial distribution systems,* IEEE Transactions on Power Delivery **4**, 725 ().

[27] D. B. Arnold, M. Sankur, R. Dobbe, K. Brady, D. Callaway, and A. Von Meier, *Optimal dispatch of reactive power for voltage regulation and balancing in unbalanced distribution systems,* in *16th IEEE Power & Energy Society General Meeting (PESGM)* (IEEE).

[28] M. Sankur, R. Dobbe, E. Stewart, D. Callaway, and D. Arnold, *A linearized power flow model for optimization in unbalanced distribution systems,* arXiv preprint arXiv:1606.04492 .

[29] L. Gan and S. H. Low, *Convex relaxations and linear approximation for optimal power flow in multiphase radial networks,* (2014).

[30] B. A. Robbins, H. Zhu, and A. D. Dominguez-Garcia, *Optimal tap setting of voltage regulation transformers in unbalanced distribution systems,* IEEE Transactions on Power Systems **31**, 256.

[31] *IEEE distribution test feeders,* http://ewh.ieee.org/soc/pes/dsacom/testfeeders/.

[32] *Dataport,* http://www.pecanstreet.org/ (2017).

[33] A. Monticelli and F. F. Wu, *Network observability: Theory,* IEEE Transactions on Power Apparatus and Systems **PAS-104**, 1042.

[34] A. Gomez Exposito and A. Abur, *Generalized observability analysis and measurement classification,* in *20th International Conference on Power Industry Computer Applications., 1997*, pp. 97–103.

5

Analysis of energy transition impact on the low voltage network using stochastic load and generation models

The energy transition poses a challenge for electricity distribution network design as new energy technologies cause increasing and uncertain network loads. Traditional static load models cannot cope with the stochastic nature of this new technology adoption. Furthermore, traditional non-linear power method have difficulty to evaluate very large networks with millions of cables, because they are computationally expensive. This chapter proposes a method which uses copulas for modeling the uncertainty of technology adoption and load profiles and combines it with a fast linear load flow model. The copulas are able to accurately model the stochastic behavior of solar irradiance, load measurements and mobility data, converting them into electricity load profiles. The linear load flow model has better scalability and stability compared to traditional load flow models. The models are applied to a case study which uses a real world dataset consisting of a realistic technology adoption scenario and a low voltage network with millions of cables, which considers both voltage and current problems. Results show that risk profiles can be generated for all cables in the network, resulting in a valuable map for the district network operator as to where to focus their efforts.

Parts of this chapter are published in Energies (2020), Volume 13. Co-author: Raoel Bernards

5.1. Introduction

The energy transition poses a challenge for electricity distribution network design as new energy technologies cause increasing and uncertain network loads. Distributed energy generation becomes more common as Photovoltaics (PV) get cheaper, shifting electricity production from predictable large plants to consumers [1]. Additionally, consumers are electrifying their transportation and heating needs with electric vehicles (EV) and heat pumps (HP). Each of these techniques has a significant impact on the residential electricity consumption [2].

Distribution network operators (DNO), which are responsible for maintaining the medium and low voltage (LV) network, are trying to determine the exact impact of these technologies. This is especially challenging on a low voltage level, given that the technology adoption of individual customers is hard to predict. Furthermore, because of their vast size, the impact on urban low voltage networks is hard to simulate.

For predicting the increasing and uncertain loads, this chapter proposes a method which uses copulas for modeling the uncertainty of technology adoption and load profiles. For determining the impact on the low voltage network, a fast linear load flow model is described. The copulas and load flow models are combined in a case study considering a large real world low voltage network. Results show that risk profiles can be generated for all cables in the network, resulting in a valuable map for the district network operator as to where to focus their efforts.

5.1.1. Related work

Traditionally, the impact of load on low voltage networks is determined by evaluating a single deterministic peak load. This peak load is determined via a Velander approximation [3] which accounts for peak load covariance between consumers. While this approach is straightforward and requires relatively little data, it cannot cope with the uncertain and irregular load profiles caused by the adoption of new energy technologies [2]. In more recent literature, probabilistic approaches are the norm for assessing low voltage networks. With a probabilistic peak load, it becomes possible to make risk-based decisions and consider multiple planning alternatives [4]. Considering residential load, many probabilistic load models are designed bottom-up [5–7], starting at individual appliance level and their operational states, and combining these into a specific model for each house. While accurate, bottom-up require information on a household level, which is generally not available by DNOs. Furthermore these models have a high modeling intensity, i.e. they require many individual variables and parameters to be modeled correctly. This necessitates the use of additional assumptions and renders them impractical for application [8]. Conversely, spatial load forecasting methods based on historical growth data can be employed such as [9]. However, this approach lacks the possibility to evaluate different future scenarios, due the absence of modeling the underlying drivers for changes in load.

Other work focuses on probabilistic simulation through random sampling from measured datasets, e.g. [4, 10, 11]. Either electric load profiles are sampled directly, or profiles of different variables (e.g. heat demand, or solar irradiance) are

sampled and converted to electric load. While these approaches allow for ease of use and low modeling intensity is required, they are very dependent on the actual values that occurred by chance in the used datasets.

The statistical modeling of measurement data related to electricity networks and energy technologies such as PV, EV, and HP has been discussed in for instance [12–17], using various distribution functions. However, all these models show applicability only to one specific set and no mention is made of generalizability to other data.

Another major point in the statistical modeling of load profiles is dealing with the inherent correlation between time steps. This effect can be accounted for by copulas [18], which are the basis of this chapter. Copulas can be used to model the correlation structure between different random variables, allowing for separate modeling of the correlation and the marginal distributions. Copulas have been successfully applied to modeling loads in power system studies [19–23], most often focusing on modeling wind power generation.

5.1.2. Contributions

This works adds to the existing research on load and generation modeling by extending and implementing a flexible modeling method that is able to model the statistical behavior of load profiles and is applicable to a variety of data. Existing methods to model stochastic load profiles are often specifically tailored to a single application, requiring information and data on a very low level to accurately model the behavior. In low voltage network planning such low level information is usually not available, making the models infeasible for use without relying on additional assumptions. Moreover they are difficult or impossible to adapt to different technologies, limiting their efficient use in large scale network planning which needs to include many different load components. Our method allows for an accurate and standardized way to model the electric load behavior of different connections and technologies, and can therefore be incorporated in existing network planning processes with relative ease.

Furthermore, stochastic load models are created for residential load, PV, EV and HP, yielding insight into their electrical behavior and the uncertainties herein. The profiles are generated by first modeling the statistical behavior present in different measurement datasets and subsequently modeling the electric load based on the statistical representation of the data. Additionally, we show that these profiles are accurately modeled using the proposed methodology. The models also include different correlations which are, specifically temporal correlations for residential load and PV generation, correlations between travel distance and arrival and departure times for EV charging, and the combined correlation with time and temperature for HP load. We also show that the use of a single combination of probability distribution functions (Gaussian mixture models with t-copulas) can effectively model many different datasets.

The stochastic nature of the load models puts a higher computational burden on simulations using these models. To be able to effectively apply the models in large scale network simulations, a fast network analysis methodology is therefore

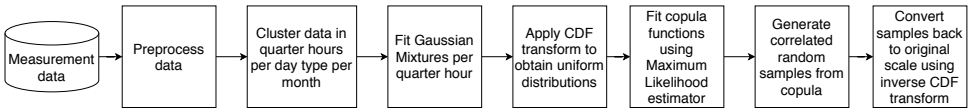


Figure 5.1: Diagram of the steps for generating the correlated load models.

discussed and applied in the final sections of this chapter. This analysis methodology allows for fast and detailed analysis of large network areas (e.g. two to three million connections). We demonstrate that by using this analysis methodology it becomes feasible to apply more detailed load models also for large scale network evaluations, which might previously have been limited by unacceptable computation times.

5.1.3. Chapter outline

In the next section first the general profile modeling methodology will be discussed. Then we show the application, fitting and output of an application of the methodology to different sets of measurement data. The methodology is applied to create individual profile models for current residential load, PV, EV and HP. In the subsequent section, the models are applied in a large scale network analysis using a fast and efficient loadflow algorithm. Here the uncertain impacts of an energy transition scenario are shown on the real-world LV network area of a distribution grid operator. Finally, the chapter concludes with several conclusions and recommendations regarding the proposed models and their application.

5.2. Stochastic profile modeling

5.2.1. Methodology

The modeling of the load and generation profiles is based on a method proposed by [24]. In this work the original method is extended with several modifications to allow the generation of load profiles based on measurements of different variables.

The core of the methodology is a two-stage approach to the modeling of stochastic correlated variables. First the marginal distributions of the data variables are modeled using Gaussian mixture models (GMMs), while the correlation structure of the variables is modeled separately using copula functions. This detachment of the marginals and the correlation allows for a flexible modeling method, which can be effectively applied to a variety of data.

The different steps in the modeling methodology are detailed in Fig. 5.1. In the first step the measurement data are structured and preprocessed to remove missing values. Then the data are clustered by quarter hour of the day, per day type (week or weekend) and per month. Consequently, GMMs are fit to model the marginal distribution of each quarter hour. The data are then transformed to uniform distributions using the CDF transform as in [25] and a copula is fit to the multivariate uniform distribution. To generate profiles first samples are generated from the copula functions, which are then converted back to the original scale of the marginal distributions using the inverse CDF transform.

Accuracy of the stochastic models is evaluated by the Continuous Rank Prob-

ability Score (CRPS), which is an often used score in ensemble forecasts [26]. It compares the difference between a forecast CDF and the empirical CDF. Additionally the Root Mean Square Percentage Error (RMSPE) [27] is used to evaluate average output profiles to the measured average to quantify the level forecast error in the output model. A relative measure is used to judge the fit to the shape of the profile, as the absolute values between (15-min) clusters may vary significantly.

More details about the modeling technique can be found in [24], in the next paragraphs the individual steps will be explained more in-depth by applying the method to several datasets. The used datasets contain electric residential load profiles, solar irradiance measurements, mobility data and heat pump electric load measurements.

5.2.2. Residential load

First we consider the modeling of the base residential load, which is the current load of households without any PV, EV and/or HP system installed. The dataset used for this is an open data set of smart meter data of residential loads in 2013 [28]. This dataset contains recorded 15-min average energy consumption values for 80 households for a whole year. The households in this dataset vary from apartments to detached houses, with a gas connection for heating and cooking, and a wide range in yearly electricity consumption varying from 1000 to over 6500 kWh. Moreover, the households also do not have any energy transition technologies already installed.

Data and preprocessing

To analyze and model the statistical behavior the measurements are first clustered into groups, based on time of day, day type (week- or weekend day) and month of the year. This partitioning is based on the seasonality of the load and time of use behavior of residential consumers [7, 29]. Each cluster now consists of all measurements for all households for all week- or weekend days in a specific month at a specific 15-minute time interval (e.g. the time interval of 12:00 - 12:15h for all weekend days in June). This results in 12 months x 2 day types x 96 quarters = 2304 clusters with each around 1700 and 700 measurements for week- and weekend days respectively. Fig. 5.2 illustrates the behavior of the residential load over time and the stochastic variation at specific time instants. The mean profiles in Fig. 5.2a show a clear daily cycle, with a peak at 18:00h of approximately 0.9 kW in the winter and 0.5 kW in summer. The wide confidence interval around the mean indicates a large spread in individual load values, which is more detailedly displayed in Fig. 5.2b and 5.2c which show the distribution of the load at two chosen 15-min intervals.

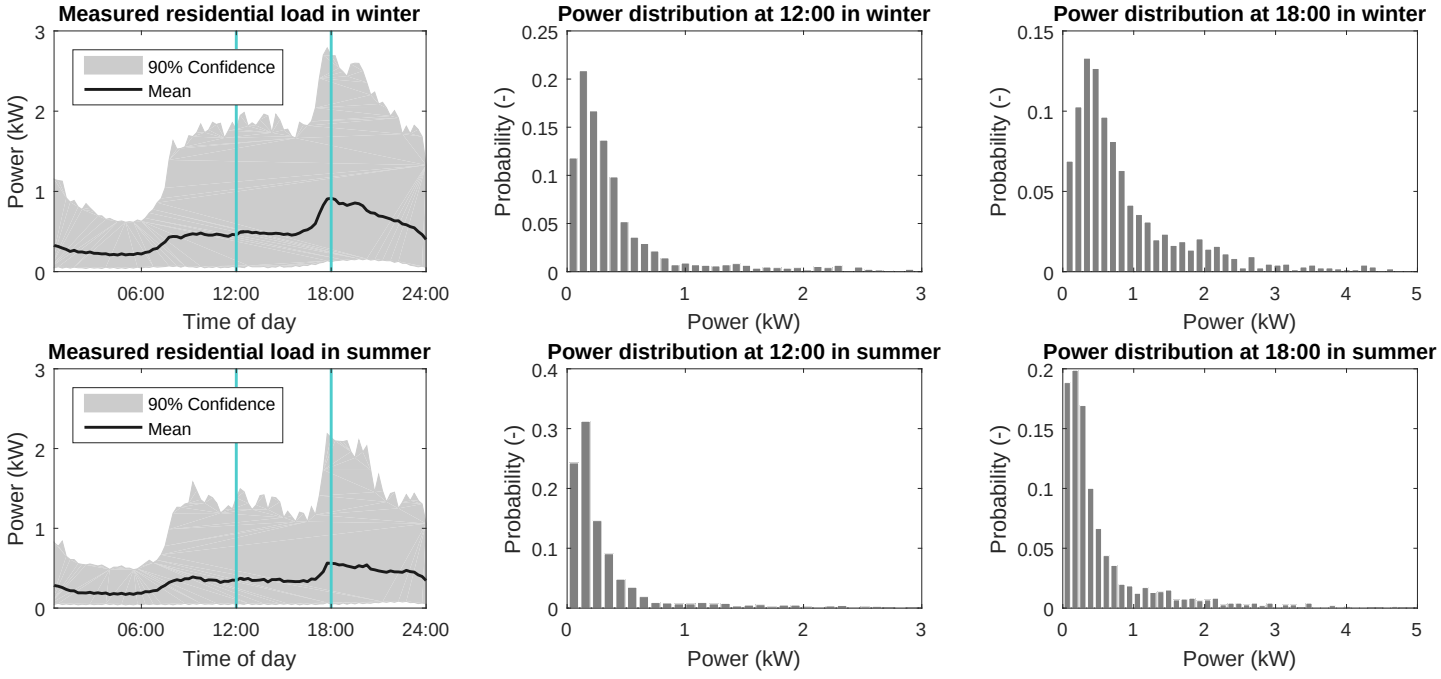
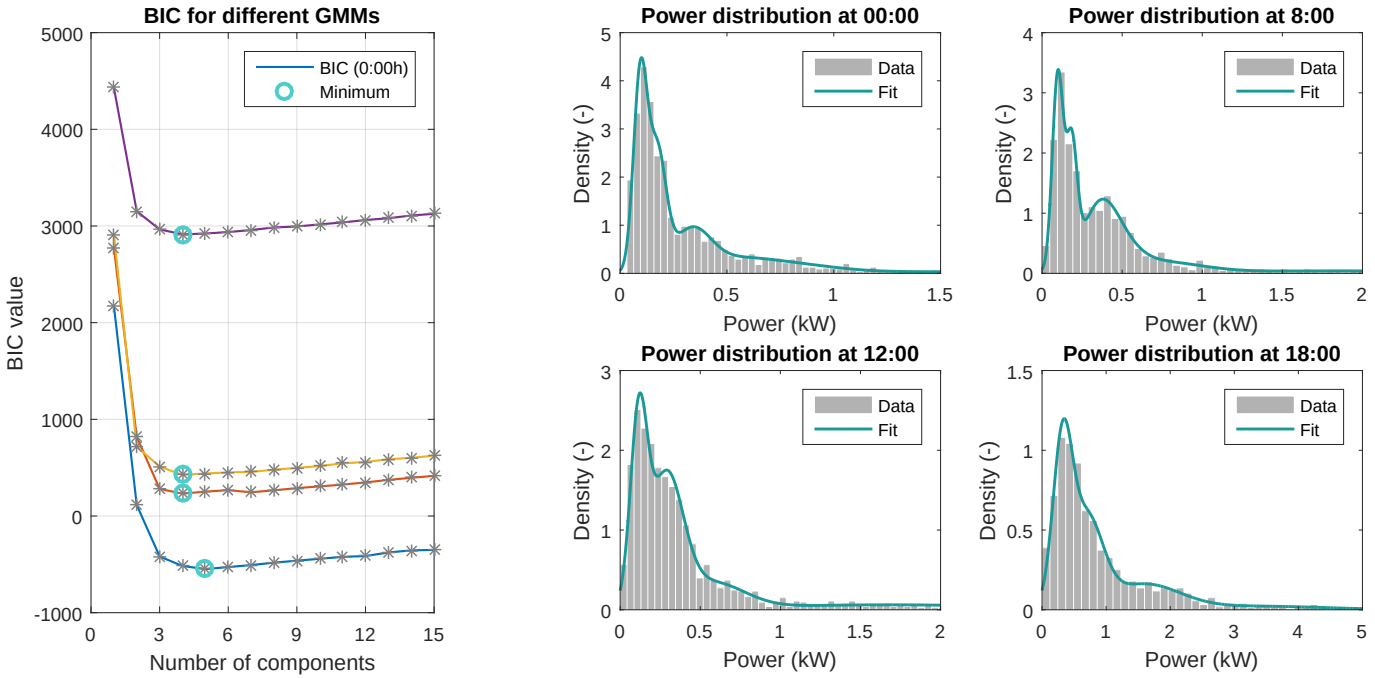


Figure 5.2: Load behavior in winter and summer. (a) mean profiles and 90% confidence interval around the mean, (b) and (c) stochastic behavior of load at specific time intervals.



(a) (b)

Figure 5.3: Fitting of mixture models to measurement data at four time intervals. (a) The BIC values of the GMMs for an increasing number of components, (b) PDFs of data and model fit using the optimal number of components.

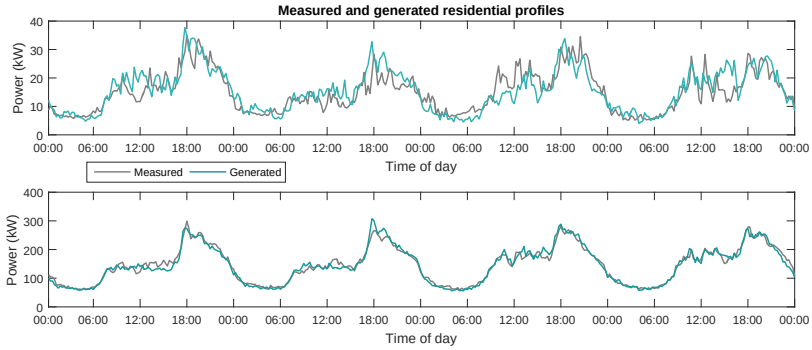


Figure 5.4: Comparison of measured and generated profiles for two week- and weekend days in January for 30 (top) and 300 (bottom) households.

Model fitting

Firstly, GMMs are fit to the measurement data to obtain the marginal distributions of each individual quarter of an hour. The optimal number of components of each GMM is selected by evaluating at what number the Bayesian Information Criterion (BIC) is minimal. Fig. 5.3a shows the evaluation of the BIC for GMMs with an increasing number of components, for several quarter of an hours in January. For all clusters the optimal number of components varies between 3 and 7 components. Fig. 5.3b show the fit of four of the mixture models to their respective data, by comparing the probability density function (PDF) of the GMM to the measured data, indicating a good fit and showing no signs of overfitting. CRPS values of the different clusters are also low, 0.01 kW on average, ranging from 0.002 to 0.02 kW.

Next, a transformation is applied to transform the data to (approximate) uniform distributions with a range between [0,1] by using the probability integral transform [18]. Copulas are fitted for modeling the dependence structure for all points in time for a each day-month date-type (e.g. a weekday in January). Different families of copulas [18, 30] were considered and their 'goodness of fit' was tested based on the log-likelihood of the transformed data points. For all studied cases, the t-copula produced the best overall results. Subsequently, these copula functions can be used to generate daily load profiles.

Profile generation and evaluation

By using the copulas, correlated samples are generated. The samples are obtained from day profile with a 15-minute interval resulting in 96 correlated time steps per day. The generated samples are normalized on a [0,1] range and to transform the samples back to their original scale (representing their actual power values present in the data) an inverse CDF transformation with the marginal distributions was applied. Given that the CDF of mixture distributions cannot be obtained analytically, we opted for a two-step numerical method. First the marginal CDFs were generated with small discrete steps. Subsequently a line search was performed resulting in the correctly scaled power values.

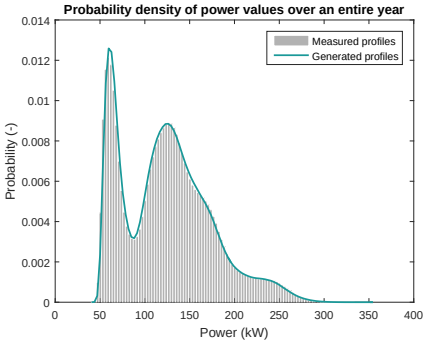


Figure 5.5: Probability density of year profiles drawn from the measured values or from the generated day profiles.

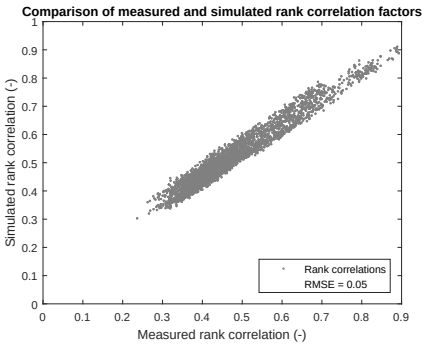


Figure 5.6: Correlation plot of the correlation matrices of the measured and simulated values.

Profiles can be now generated which represent the statistical behavior and correlations present in the original dataset. Fig. 5.4 shows a comparison of measured and generated profiles for 30 and 300 random household profiles for four days in January. By visual inspection, the model appears to represent the time behavior and peak load well. The probability density distribution of the load for a full year is shown in Fig. 5.5, indicating a proper capture of the stochastic yearly behavior by the model. Further, the correlation factors between quarters of an hour of the measured and modeled profiles are compared in Fig. 5.6, again demonstrating a good representation of these correlations by the model with a root mean squared error (RMSE) of 0.05.

Due to the flexibility of the GMMs and copulas, profiles can be generated for PV, EV and HP technologies in a similar fashion. However, for each of these cases some technology specific adaptations are required to achieve proper load- or generation profiles. The next sections will focus on these extensions of the proposed method for each technology.

5.2.3. Photovoltaics

The power output of a PV system depends on the effective solar irradiance on the surface of the panels. This is affected by several factors such as the actual irradiance levels, the changing angle of the sun over the day (and year), and the orientation and inclination angle of the panels [31]. To model the PV output, we first model the irradiance profiles and subsequently convert the irradiance values to an instantaneous power output.

Data and model fit

The modeling of the irradiance profiles is based on measurements of the solar irradiance for a period of 10 years from four weather stations in the Netherlands, obtained from the openly available meteorological data of the Dutch Royal Meteorological Institute (KNMI) [32]. The data is clustered per 15-minute interval of the day for each specific month (the distinction between week- and weekend days is

obviously not made), resulting in 12 months x 96 quarters = 1152 clusters with each around 1300 measurements. Fig. 5.7 shows the fit of six mixture models to their respective data clusters for morning, noon and evening clusters in January and June.

Model fitting

Next we again first model the marginal distributions of each individual quarter of an hour. Because there is no or negligible irradiance during night we only fit models on quarters between (nautical) dawn and dusk[1]. Evaluation of the optimal number of components showed the optimal performance for 2 to 4 components, depending on the time of day and year. Fig. 5.7 shows the fit of six mixture models to their respective data clusters for morning, noon and evening clusters in January and June. The densities of the fitted mixture models show to track the stochastic behavior of varying levels of irradiance well, with an average CRPS of 0.011 kW/m2 for summer days and 0.050 for winter days.

To model the correlation between solar irradiance at different time intervals of the same day copula functions are then fit to the transformed distributions of the data. The correlation matrices of the solar irradiance are constructed for the quarters between dawn and dusk for each month and the remaining correlation values are fixed at 0. Analyzing different model copula function fits showed the t-copula to also provide the best fit for solar irradiance data.

Profile generation and evaluation

By sampling values from the copula functions and transforming these back using the GMMs, stochastic irradiance profiles can be generated. Fig. 5.8 shows a comparison of averaged measured and generated profiles of 30 and 300 random two day profiles in summer and winter, indicating the model closely represents the measured profiles. The respective RMSPE values of the summer and winter profiles are 2.9% and 5.3%.

[1] Nautical dawn and dusk are the moments when the sun is 12 degrees below the horizon, around this time the irradiance may become large enough for PV installations to start producing.

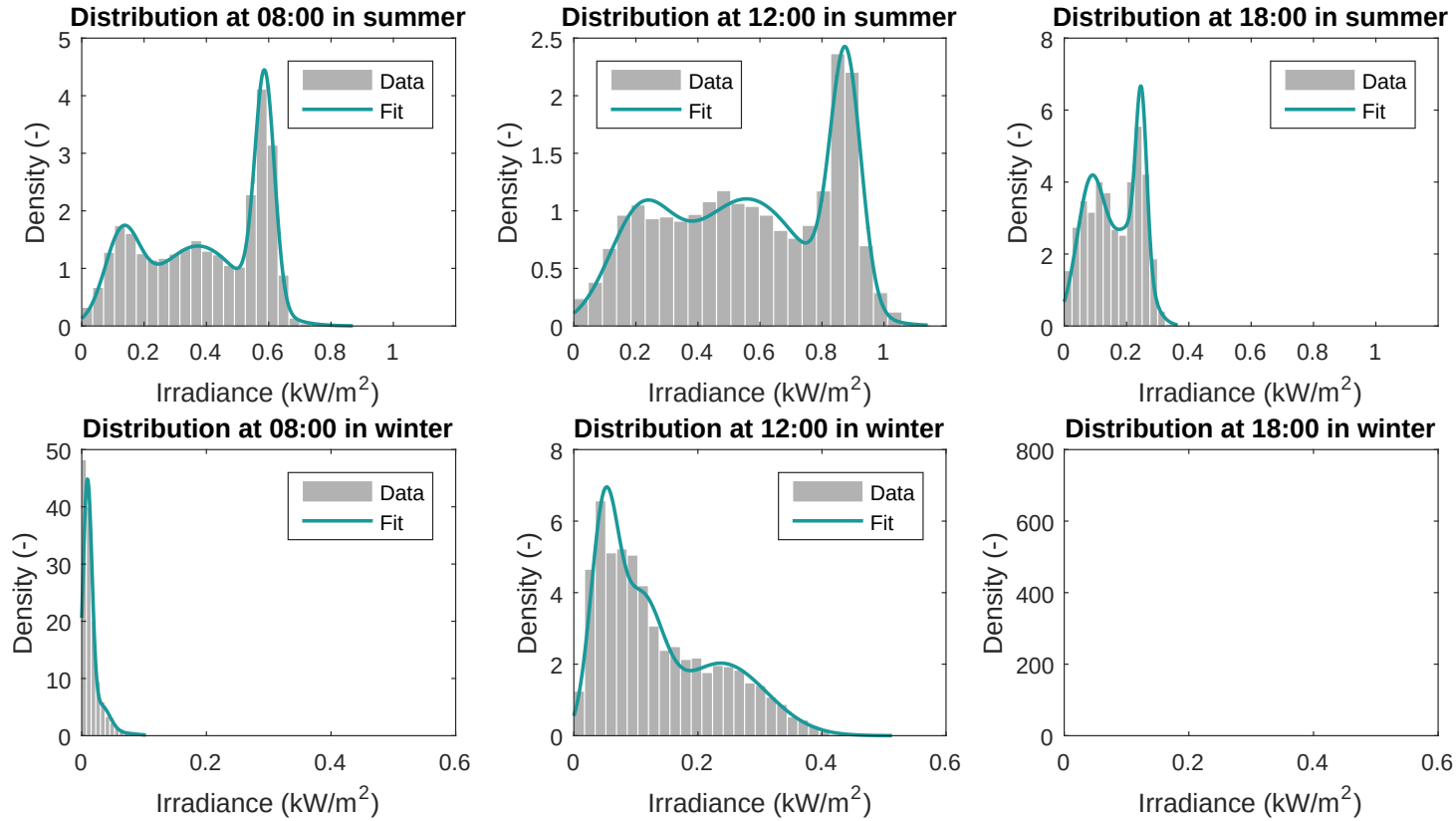


Figure 5.7: PDFs of data and fitted mixture models for irradiance measurements at three time intervals in summer and winter.

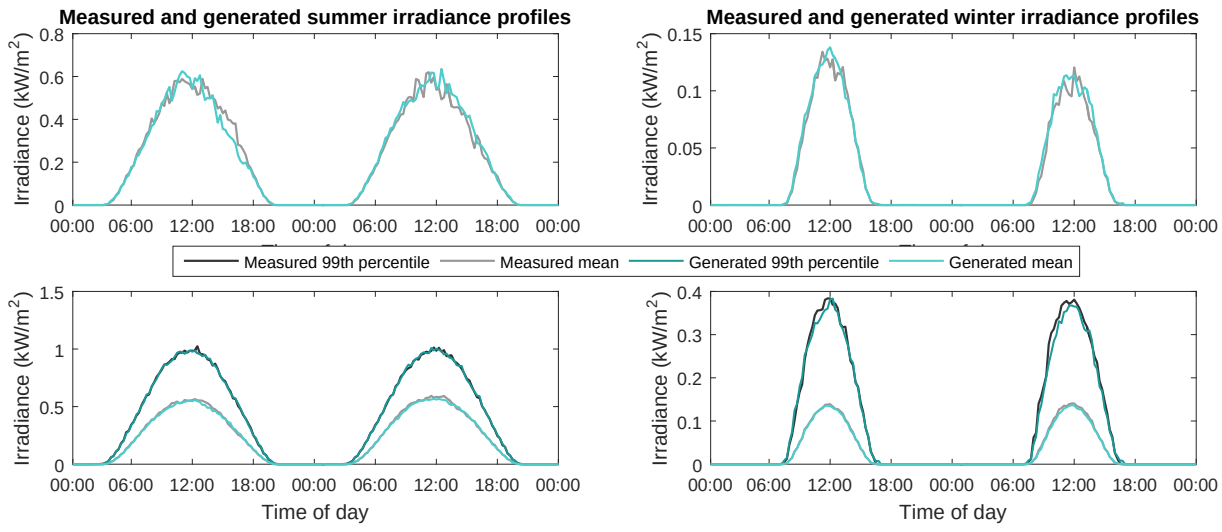


Figure 5.8: Comparison of measured and generated profiles of 30 (top) and 300 (bottom) random two day profiles in summer and winter

To obtain profiles of the power output of a PV system the global irradiance profiles are converted to the irradiance on an inclined surface using the calculation methods from [31],[33] and [34].

5.2.4. Electric vehicles

To model electric vehicle's charging profiles a modified approach is used, as rather than fitting the marginal distributions directly to measured charging profiles, we model arrival times and trip distances and convert to electric load profiles afterwards. This has several reasons, firstly because data on mobility is openly available for many countries, this way of building up the models is more generically applicable than by relying only on measured charging profiles which might not be available or not representative enough. Secondly, (in absence of some smart charging scheme) the charging profiles of electric vehicles usually take the form of block shaped curves with an event-based character, i.e. either charging at full power, or not charging. A stochastic representation of these profiles would look like a combination of Dirac (delta) functions at zero and full charging power, which are not well representable by a gaussian mixture with non-zero variances. Thirdly, the charging profiles would change with varying charging rates or when implementing smart charging to shift charging time and alter power consumption levels, this can more easily be integrated in this way.

Data and model fit

The load profiles of electric vehicles are based on surveyed mobility data from a Dutch Mobility research [35]. The dataset contains over 30.000 journeys by car, made by over 11.000 individual persons. From this dataset, the arrival and departure times, and distances traveled are extracted and modeled. To implement a (from a network perspective) worst-case situation, we order the data such that all persons are assumed to charge their car battery after the last homebound trip and the total energy needed is based on their cumulative distance traveled that day. Hereby it is assumed that the battery capacity in each car is sufficient for the total distance driven that day. The surveyed and modeled distributions of the departure and arrival times, and the distance traveled are shown in Fig. 5.9.

Model fitting

The surveyed and modeled distributions of the departure and arrival times, and the distance traveled are shown in Fig. 5.9. The plots again show a good tracking of the marginal stochastic behavior of the different variables, which is further supported by the low Continuous Ranked Probability Score (CRPS) values of 0.11 and 0.08 minutes, and 10.2 meters for respectively departure time, arrival time and distance traveled.

Profile generation and evaluation

By sampling from the fitted copula functions and applying the previously described transformation, combinations of departure time, arrival time and travel distance are generated. Using the average efficiency of currently available electric vehicles,

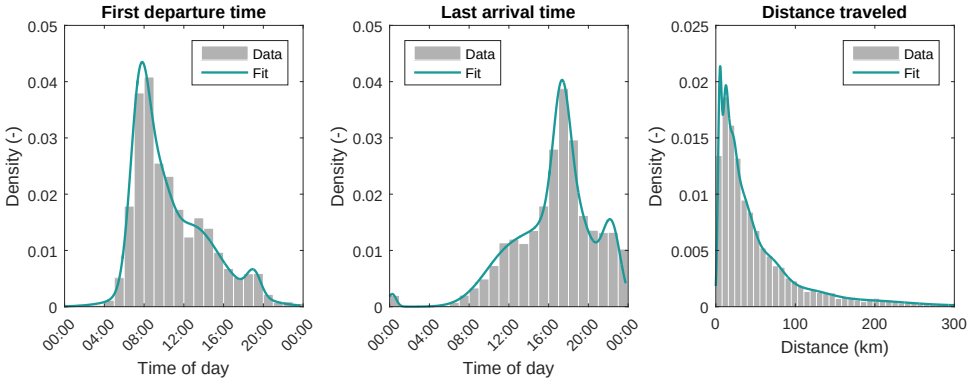


Figure 5.9: PDFs of data and fitted mixture models for first departure time, last home arrival time and cumulative distance traveled.

the power consumption while driving is assumed to be 6 km/kWh [36]. Given the available information on traveled distance, the total required energy each trip is calculated. These energies are then converted in profiles by assuming a charge rate of 3.6 kW, which corresponds to the single phase residential charge power. With the final assumption that people start charging their car at full power at the time they arrive at home, representative charging profiles have been be generated. Figure 5.10 shows individual and aggregated profiles for residential EV charging. The average charging duration is roughly 2.5 hours, which aligns with the average travel distance of 54 kilometers. The bottom graph shows a comparison of the average simulated charging profiles to the average of measured profiles from a dataset provided by ElaadNL. This dataset contains measurements EV charging at 428 (semi-)public charging stations in the Netherlands for a full year in 2017. From this dataset the charging stations which are located in residential areas and that did not show charging peak in the morning (to exclude charging when arriving at work) were selected for comparison, furthermore the set was filtered to include only day profiles with at least one charging session.

Overall the general shape of the profiles is similar, however the simulated profile has a lower peak, which also occurs a bit earlier in the day compared to the measured profile. The differences may be related to the fact that the measured profiles are from (semi-)public charging stations, and not from actual residential consumer connections, this might cause a different charging behavior at individual stations. For instance, as the charging stations are in the public space multiple cars may connect on a single day, for usually shorter charging duration, and some stations are less regularly used than would be expected for a private residential charging station. Also the driving behavior of the average EV driver in the recent past would not necessarily resemble the the average overall driving behavior of the current Dutch car fleet, e.g. the difference in peak suggests that on average current EV drivers arrive at home a bit later than the overall average driver.

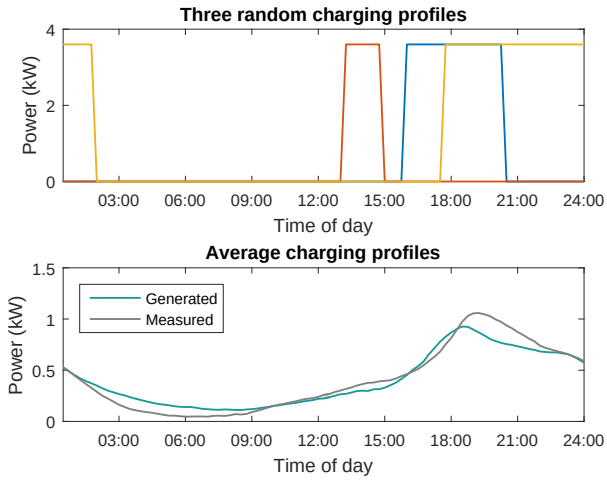


Figure 5.10: Individual and aggregated profiles for charging of electric vehicles, at a charging rate of 3.6 kW.

5.2.5. Heat pumps

For modeling the electrical loads of heat pumps, measurements from the Your Energy Moment project [37] was used. Because the heating demand is also strongly influenced by the outside temperature the interdependency between electric load, time of day and temperature will be modeled to obtain realistic profiles.

Data and preprocessing

The used measurements consist 15 minute intervals of the electric load of 37 heat pumps during one year, along with the temperature at each time interval. The heat pumps used in the pilot project had a two-step compressor with two power levels, 1 kW regular heating needs and 1.8 kW for tap water. For households with higher or lower heating demand we assume the installed power of the heat pump will be sized accordingly, thus the overall usage profile remains identical only the actual power consumption changes. The data are again clustered per month, day type and quarter of an hour of the day.

Model fitting

In Fig. 5.11, the PDF of the electric load of a heat pump is shown for evening time in winter. Three distinct peaks can be discerned, signifying the three general states of operation of the heat pump: off, room heating/cooling, and tap water heating. During the off-state, a small amount of power is still being used to keep the water flowing through the system. The other two peaks occur around the compressor steps of the heat pump, the variance being caused by the change of operation state during the measured 15-min intervals.

Once again, copula functions are used to model the correlations present in the dataset. For the heat pump case, besides the autocorrelations of the load itself, also

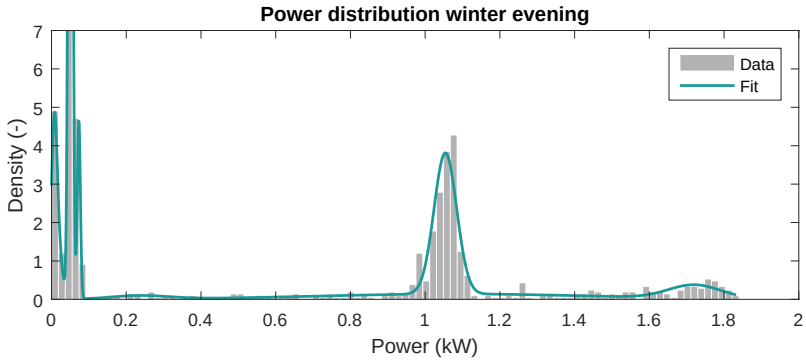


Figure 5.11: Measured and fitted PDF of heat pump electric load during winter evening, showing three peaks at the levels of power usage for the different operating states.

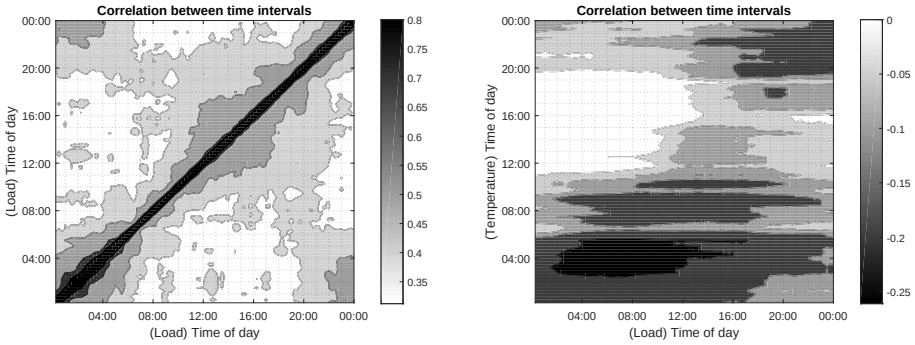
5

the correlation between load and outside temperature will be taken into account. Fig. 5.12 shows the correlation matrices of the load vs. load (Fig. 5.12a) and the temperature vs. load (Fig. 5.12b). The autocorrelation of the load is characterized by high correlation factors for subsequent time intervals, degrading for time intervals further away. The correlation between load and temperature is negative (higher temperature equals lower load), with the largest correlations in late night and early morning, and late evening time intervals, indicated by the darker shading in the figure. Notably, the correlation of the load with the temperature during the afternoon is much lower, and for time intervals further apart virtually non-existent. The correlation factors between load and temperature appear relatively low. This is mainly caused by two reasons. Firstly, the fact that only for relatively low temperatures (<5 degrees Celcius) the correlation is decidedly higher. For moderate and high temperatures the correlation quickly drops. Since the correlation plot shows the average correlations over the entire year, the moderate temperatures make up a large portion of the dataset, resulting in a low overall correlation. Secondly, The electric power use of the heat pumps is partially used for tap water heating. This type of heat demand has a much lower correlation with outdoor temperature and is much more dependent on the time of day.

Profile generation and evaluation

By sampling from the copula functions random load-temperature profiles can be generated. Fig. 5.13 shows the average load and temperature profiles generated with the models compared to the measured averages. The model shows to provide a good representation of the measurements, with an RMSPE of 2.7% for the winter load profile and 5.1% for the summer profile. The average aggregated peak load occurs in the early morning and is approximately 0.65 kW.

To apply these stochastic models for large scale network evaluations they will be incorporated in a linear network model as described in the next section.



(a) (b)

Figure 5.12: Correlation of power consumption at different time intervals, with darker shading indicating a larger magnitude of the correlation factor. (a) autocorrelation of electric load, (b) correlation of load with temperature.

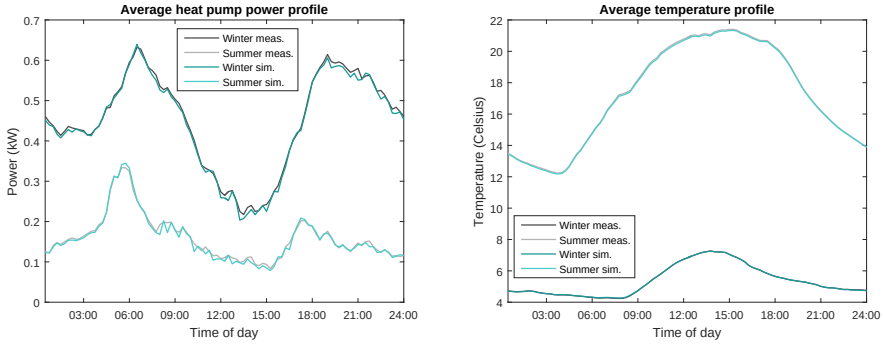


Figure 5.13: Average heat pump load and temperature profile during winter and summer.

5.3. Evaluating a large low voltage electricity network

This section introduces a linear network model which is suitable to evaluate very large low voltage distribution networks [38, 39]. Contrary to the models presented in the previous section, the proposed load flow approach is deterministic. Therefore a Monte Carlo approach is applied to simulate the impact of the stochastic peak loads. The general load flow modeling approach is displayed in Figure 5.14 and is described more thoroughly in [38, 39], where also its accuracy is assessed on a test bed. Additionally, high-efficiency solvers for this load flow formulation are considered in [40, 41].

In calculating power flows in distribution networks, it is sufficient to use a steady state electrical model. The most common way to calculate these properties is by us-

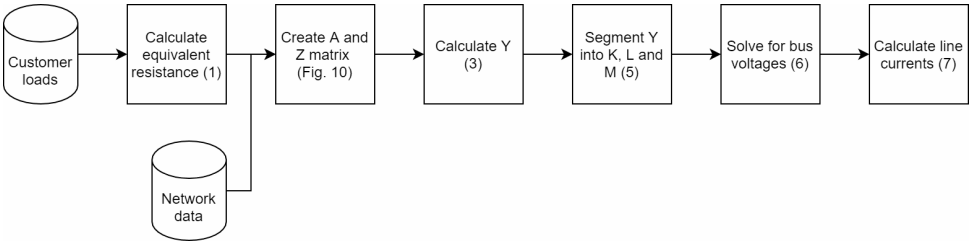


Figure 5.14: Diagram of the steps for solving the linear load flow problem.

ing a load flow model [42, 43]. The consumer load is traditionally modeled by combining constant power, constant current and constant impedance load models[43]. The constant power model (i.e. the network power load is independent from the network state) is most commonly used. However, while using a constant power load the load flow model is non-linear, which makes simulating large networks challenging due to increasing computational time and convergence problems [41].

5

To create a load flow model which is capable of modeling very large networks, we only use a constant impedance model. This results in a fast linear model, which is still sufficiently accurate [38, 40] and over 8 times faster than a non-linear approach [40].

An example constant impedance network is displayed in Figure 5.15. This network consists of two impedances which represent the consumer loads and a swing bus which represents the connection to the medium voltage grid. The customers are modeled as an impedance which is directly connected to the ground. No current enters/leaves the network at the cable joints and the influence of the transformer is neglected.

To create the constant impedance model as in Figure 5.15, all load is converted to an equivalent impedance:

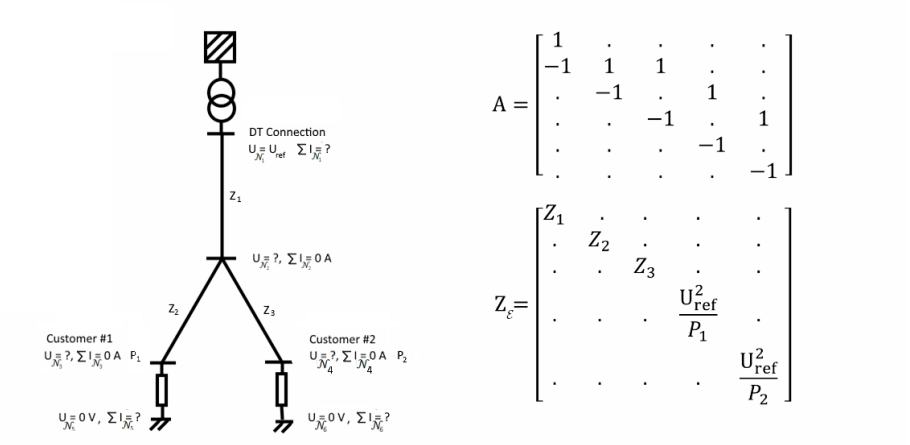
$$Z_{\text{eq}} = U_{n,\text{ref}}^2 / P_{\text{user}} \quad \forall n \in \mathcal{N} \tag{5.1}$$

In the formula above Z_{eq} is the equivalent impedance representing the load, P_{user} is the consumer load, n is an index which corresponds to the bus to which the consumer is connected and $U_{n,\text{ref}}$ is a reference voltage at the location of the consumer. Given that the actual voltage at the consumer's location is not exactly known, the reference voltage is set to the nominal voltage. This assumption introduces a small but acceptable error [38, 40]. The consumer load was assumed to have a power factor of 1.

The grid is described by a graph. The graph G is defined as $G = (\mathcal{N}, \mathcal{E})$. Here $\mathcal{N}$ represent the network buses and $\mathcal{E}$ are the network edges/lines. The modeling objective is to calculate the nodal voltages $U_{\mathcal{N}}$ and the cable currents $I_{\mathcal{E}}$.

The starting point is Ohm's law, with which the nodal voltages and bus currents can be determined:

$$I_{\mathcal{N}} = \bar{Y} U_{\mathcal{N}} \tag{5.2}$$



$$A = \begin{bmatrix} 1 & . & . & . & . \\ -1 & 1 & 1 & . & . \\ . & -1 & . & 1 & . \\ . & . & -1 & . & 1 \\ . & . & . & -1 & . \\ . & . & . & . & -1 \end{bmatrix}$$

$$Z_{\mathcal{E}} = \begin{bmatrix} Z_1 & . & . & . & . \\ . & Z_2 & . & . & . \\ . & . & Z_3 & . & . \\ . & . & . & \frac{U_{ref}^2}{P_1} & . \\ . & . & . & . & \frac{U_{ref}^2}{P_2} \end{bmatrix}$$

Figure 5.15: An example low voltage distribution network (left) and its corresponding A and $Z_{\mathcal{E}}$ matrices (right).

$I_{\mathcal{N}}$ is the bus current i.e. the current that enters/leaves the network and $\bar{Y}$ is the 'admittance matrix', which can be obtained by [44]:

$$\bar{Y} = AZ_{\mathcal{E}}^{-1}A' \tag{5.3}$$

A is the network directional connection matrix. Here a matrix row represents a network bus and a column represents a network line. Each line has no more and no less than one starting and one end point. These points are denoted by '1' and -1 respectively. Because of the quadratic nature of 5.3, the interchanging the minus sign between buses will result in the same admittance matrix $\bar{Y}$. $Z_{\mathcal{E}}$ is a diagonal matrix with a diagonal containing the inverse of the impedance of each line ($\mathcal{E}$).

Since neither $I_{\mathcal{E}}$ and $U_{\mathcal{N}}$ are completely known, solving (5.2) requires some extra steps. One way to efficiently solve this problem is to segment the matrices and solving two separate sub-problems. If one sorts $I_{\mathcal{N}}$, $\bar{Y}$ and $U_{\mathcal{N}}$ so that all swing and ground buses are $\in U_1$, then the problem can be written as:

$$I_{\mathcal{N}} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, \bar{Y} = \begin{bmatrix} K & L \\ L' & M \end{bmatrix}, U_{\mathcal{N}} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{5.4}$$

At regular network buses I_2 is equal to $\bar{0}$ as no currents enter/leave the network. U_1, which contains all the voltages at the swing/ground buses, is completely known. This results in the following load flow problem:

$$\begin{bmatrix} I_1 \\ \bar{0} \end{bmatrix} = \begin{bmatrix} K & L \\ L' & M \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{5.5}$$

A practical way to obtain U_2 is:

$$MU_2 = -L'U_1 \tag{5.6}$$

This equation can be defined as the standard matrix equation $Ax = B$, which can be solved by many of-the-shelf solvers. This formulation is efficient enough to be solve networks with millions of cables as will be demonstrated in Section 5.3. It is possible to increase the load flow speed with the numerical techniques described in [40].

Once all voltages are known, the cable currents can be obtained using:

$$I_{\mathcal{E}} = Z_E A' U_{\mathcal{N}} \tag{5.7}$$

A major advantage of the described linear approach is that it is susceptible to convergence problems which can occur by using non-linear models [43, 45]. These unfeasible solutions tend to happen more often in large networks with high loads. Given that network matrices with over 22 million cables always contain some errors and loads increase due to the energy transition, applying a linear model is a logical choice.

5.4. Case study: Congestion in a large real-world low voltage power grid

The objective of this case study is to find the number of overloaded cables and voltage problems in a large LV grid, using the stochastic peak loads generated from the copula models. An diagram giving an overview of the case study is presented in Figure 5.16. An overload problem is defined as a peak load current which is higher than the rated capacity of the corresponding cable. A voltage problem is defined as a voltage drop of more than 4.5% of the nominal voltage on the LV network, as in accordance with DNO policy.

For simulating the impact of new energy technologies on the distribution network, the network of the Dutch DNO Alliander was studied. Alliander DNO is the largest DNO of the Netherlands serving three million consumers. It is responsible for maintaining the Medium Voltage and Low Voltage distribution network. The LV network which is considered in this chapter has more than 80,000 km of cables and over 38,000 MV/LV transformers. The network spans roughly 1/3rd of the Dutch distribution network and is described with 22 million cable segments. The low voltage network operates on 400V/230V and is modeled as a single phase balanced network.

The Dutch power grid is very similar to many other European/US grids, but has a few distinguishing properties, which make this work more relevant. In contrary to most other countries, almost all electricity cables in the Netherlands are underground. This is mainly motivated by the fact that the Netherlands generally has a soft, non-rocky soil. While it greatly increases the network's resistance to stormy weather, it has as downside that it is very expensive to replace cables, making the planning methods of this chapter indispensable.

Furthermore, since the Netherlands is very densely populated, the low voltage part of the power grid is very large. An average MV/LV transformer has over 100 consumer connections, a 120-630kVA power rating and over 2.5km of cable. This is a major difference to, for example, a rural US network, where a MV/LV transformer

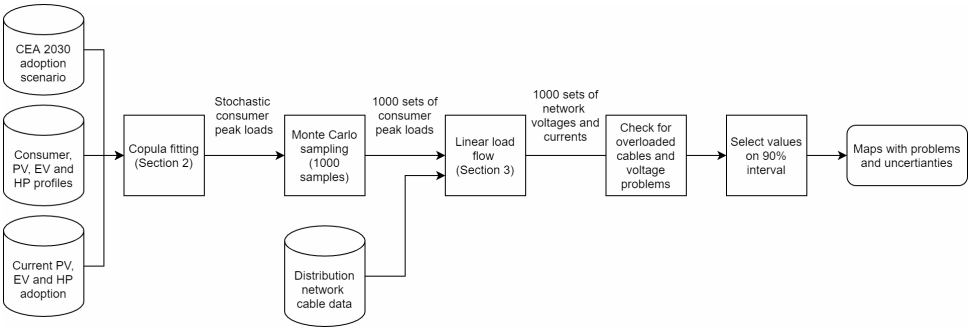


Figure 5.16: Diagram giving an overview of the steps performed in the case study.

often serves less than five consumers, has a power rating of less than 50kVA and a few hundreds of meter cable. Because of the vast size of the studied LV grid, this case study is of great value for DNOs which operate in densely populated areas.

5.4.1. General approach and scenario description

The scenario used in this study is the one provided by the Climate and Energy Agreement (CEA) [46], which is widely accepted as very plausible and is basis for the energy policy of the Dutch government.

The year of 2030 was selected for which the CEA provides adoption numbers for the PV, EV and HP technologies. According to the CEA scenario, in 2030 the Alliander area contains 4.9 GW of extra solar power generation, 1.2 million extra EVs and 2.8 million HPs compared to the current situation. This means that 40% of all low voltage electricity demand will be produced by local solar power, 39% of all cars will be an EV and 37% of the consumers will have a HP. The current adoption of PV, EV and HP is 10%, 3% and below 3 % respectively.

In accordance to the diagram in Figure 5.16, the numbers of the CEA scenario are applied to the copula models described in Section 5.2 resulting in stochastic load profiles for every consumer. These load profiles were sampled using a Monte Carlo approach, resulting in a set of customer peak loads. Each set of peak loads was evaluated using the linear load flow model from Section 5.3, following the diagram presented in Figure 5.14.

5.4.2. Simulation results

The copula model was implemented Matlab and the load flow model was implemented in the *R* programming language. By using sparse matrices and a QR decomposition method calling BLAS/LAPACK routines, the equations of the LV grid with 22 million cables were solved within 30 seconds using a single processor core of an Intel i7-6820HQ processor.

A thousand Monte Carlo iterations have been performed on the stochastic consumer peak loads resulting from the models from Section 5.2. For every iteration, all the currents and voltages in the low voltage network have been calculated. The

cable currents were compared to the rated power of each cable and the voltages were compared to the allowed voltage range. This resulted in a thousand sets with voltage and current problems and their location.

The total number of overloaded cables and voltage problems did not differ much between scenarios. The number of overloaded cables was found to be 3-4% of the total number of LV cables. This may seem like a low percentage, but since the LV of Alliander network spans over 80,000 km of underground cable, the cost implications are significant.

The number of voltage problems was calculated to be 10-12%. This is a large number and voltage problems are generally cost-intensive to solve. However, while this predicts a large future problem from a policy view (i.e. the voltages are outside the allowed design specifications), it will likely not be noticed by most consumers. Most household appliances work well within a larger voltage range than the specified 4.5% difference from the nominal voltage, with PV inverters being the exception. It remains to be seen how much cable replacement will be driven by voltage problems.

Figure 5.17 shows the geographical distribution of the load and the voltage and current problems. It can be observed that some areas in the Netherlands are expected to have much more voltage problems than others. The west of the Netherlands has the oldest network and is the most densely populated. It is not surprising that most voltage problems are expected in this area. Because Monte Carlo methods can produce outliers, the top and bottom 5% of the voltages have been filtered from the results, creating a confidence interval of 90%.

In an attempt to show the local uncertainty between Monte Carlo scenarios, Figure 5.18 shows the difference between the highest and lowest voltage within the 90% confidence interval. Darker areas have more variability in the end voltage of the consumers on average. While this is an imperfect measure of the variability between voltages because not every area has the same number of consumers, it gives an indication in which areas the effect of the new technologies is less certain. This can be used by the DNO to focus their data collection.

5.5. Conclusion

This chapter presented a general framework for modeling the effects of energy transition on the distribution network. its potential is demonstrated by applying it to a real world network.

The proposed modeling approach using Gaussian mixture models and copula functions is shown to be both accurate and flexible, and is therefore an effective method to represent the stochastic behavior and correlations of a variety of variables.

The advantages of this methodology over existing methods are that multiple different technologies or datasets can be modeled using the same approach with minor variations, and a low modeling intensity is required given an adequate dataset. This also allows the methodology to be automated, for ease of use in practical network planning situations. Furthermore, due to the parametric nature of the used models they can be scaled and altered to be adapted to local variations if required. Lastly the separate stochastic representation of individual technologies allows for

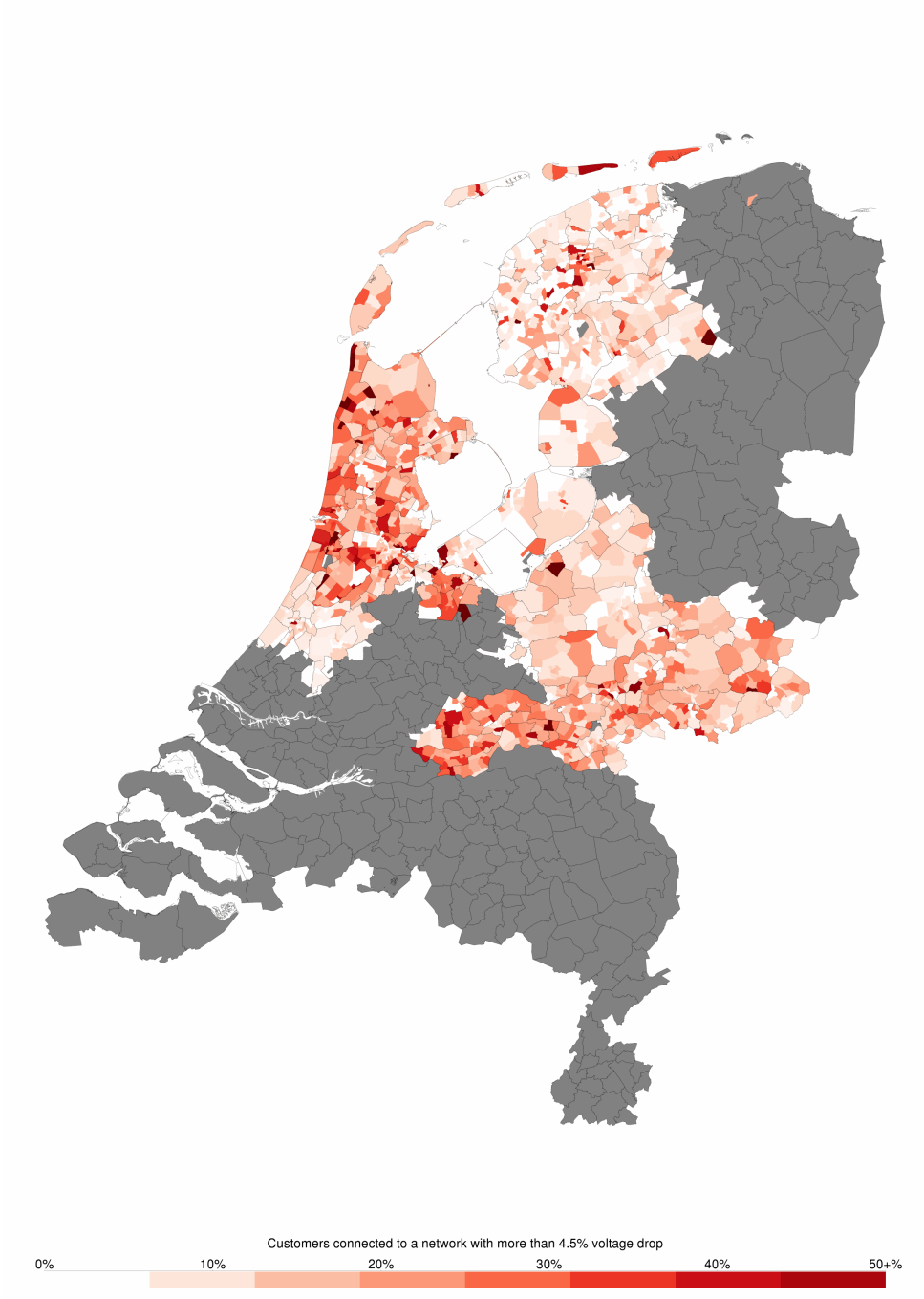


Figure 5.17: Geographic distribution of voltage problems in 2030 within 90% confidence. The grey area is The Netherlands in which Alliander operates the coloured area. The total number of customers with voltage problems in this figure is 11%.

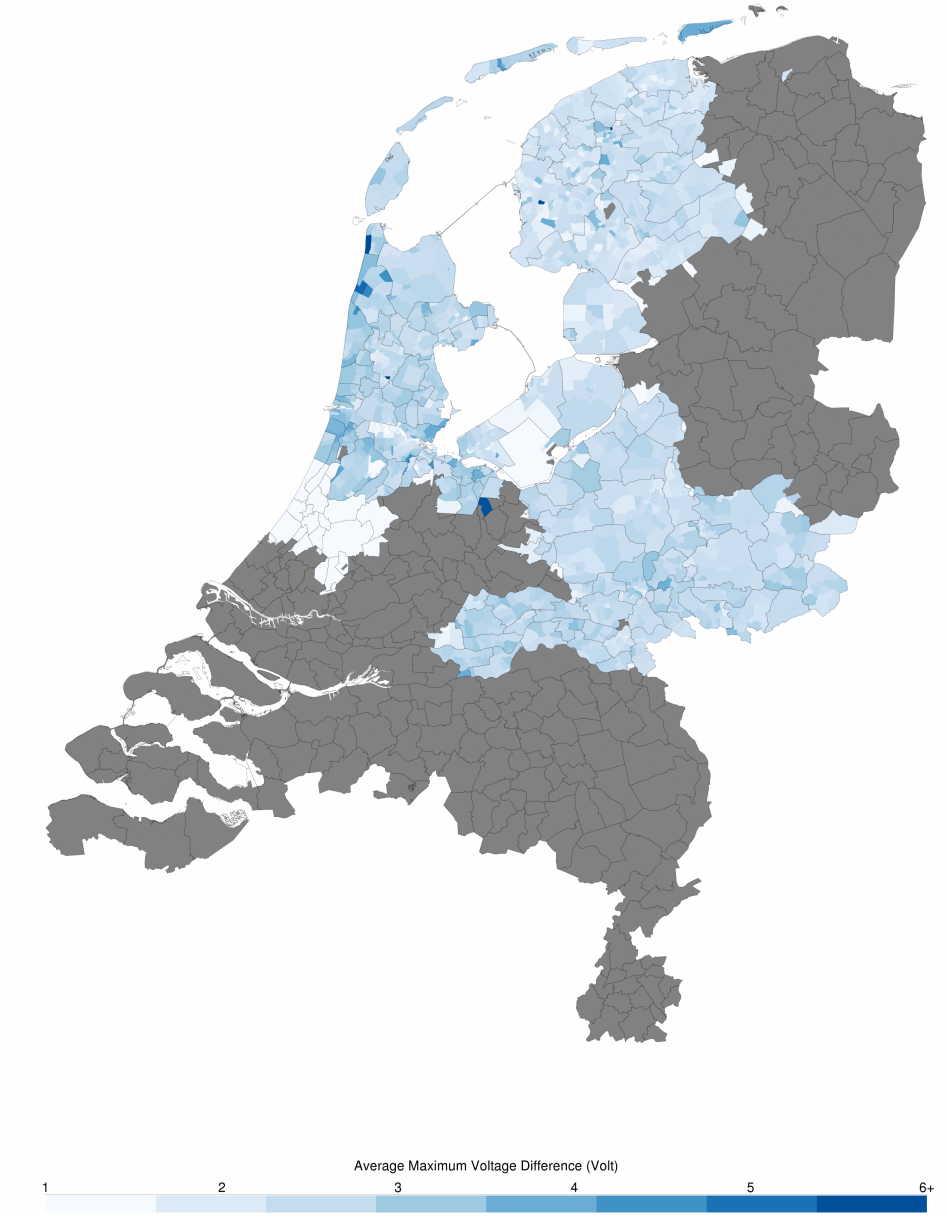


Figure 5.18: The difference between the highest and lowest voltage within the 90% confidence interval. It is a measure for the local variability between scenarios.

a risk based analysis of the change in electric load under various future scenarios. Limitation of the load modeling approach is the heavy reliance on available (measurement) data and that the relatively smooth copula functions are limited in describing very volatile correlation structures with.

A linear approach to modeling large low voltage networks has been described and applied to a very large network of Alliander DNO. The model was sufficiently fast to evaluate a the network and provided valuable insights in the impact of the new technology on the grid. Due to the higher electricity demand, 3-4% of the cable segments were overloaded and 10-12% consumers are expected to have voltage problems based on the current grid.

Given the size of the studied LV network, the number of problems is expected to pose a serious challenge for the DNO in the near future. However, using the geographical overviews from this chapter, the DNO has a valuable tool as to where to focus their efforts.

It is the author's ambition to expand the modeled network to other network voltage levels (medium and high voltage) and create an integral model for the entire Dutch electricity network. The theory presented in this chapter makes it possible to predict problems in the distribution network, which help in maintaining and securing the distribution grid for the next decades.

References

[1] U.S. Energy Information Administration, *International Energy Outlook 2016*, Vol. 0484(2016) (2016) pp. 1–2, arXiv:EIA-0484(2013) [DOE] .

[2] E. Veldman, M. Gibescu, H. J. G. Slootweg, and W. L. Kling, *Scenario-based modelling of future residential electricity demands and assessing their impact on distribution grids,* Energy Policy **56**, 233 (2013).

[3] S. Velander, *Methods of operational analysis applied to distribution of electric power,* Teknisk Tidskrif **82**, 293 (1952).

[4] A. Navarro-Espinosa and L. F. Ochoa, *Probabilistic Impact Assessment of Low Carbon Technologies in LV Distribution Systems,* IEEE Transactions on Power Systems **31**, 2192 (2016).

[5] J. Dickert and P. Schegner, *A time series probabilistic synthetic load curve model for residential customers,* in *2011 IEEE PES PowerTech* (Trondheim, 2011) pp. 1–6.

[6] A. J. Collin, G. Tsagarakis, and A. E. Kiprakis, *Development of Low-Voltage Load Models for the Residential Load Sector,* IEEE Transactions on Power Systems **29**, 2180 (2014).

[7] A. Grandjean, J. Adnot, and G. Binet, *A review and an analysis of the residential electric load curve models,* Renewable and Sustainable Energy Reviews **16**, 6539 (2012).

[8] W. Labeeuw and G. Deconinck, *Residential electrical load model based on mixture model clustering and markov models,* IEEE Transactions on Industrial Informatics **9**, 1561 (2013).

[9] V. Evangelopoulos, P. Karafotis, and P. Georgilakis, *Probabilistic spatial load forecasting based on hierarchical trending method,* Energies **13**, 4643 (2020).

[10] A. Navarro-Espinosa and P. Mancarella, *Probabilistic modeling and assessment of the impact of electric heat pumps on low voltage distribution networks,* Applied Energy **127**, 249 (2014).

[11] E. Veldman and R. A. Verzijlbergh, *Distribution grid impacts of smart electric vehicle charging from different perspectives,* IEEE Transactions on Smart Grid **6**, 333 (2015).

[12] F. Lo Franco, M. Ricco, R. Mandrioli, and G. Grandi, *Electric vehicle aggregate power flow prediction and smart charging system for distributed renewable energy self-consumption optimization,* Energies **13**, 5003 (2020).

[13] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. Van Der Sluis, *Stochastic modeling of power demand due to EVs using copula,* IEEE Transactions on Power Systems **27**, 1960 (2012).

[14] E. Pashajavid and M. Aliakbar Golkar, *Copula-based multivariate stochastic modeling of load demand due to plug-in electric vehicles in order to be integrated in distribution system planning,* in *CIRED 22nd International Conference on Electricity Distribution*, 0333 (2013) pp. 10–13.

[15] J. Widén and E. Wäckelgård, *A high-resolution stochastic model of domestic activity patterns and electricity demand,* Applied Energy **87**, 1880 (2010).

[16] Y. Mu, J. Wu, N. Jenkins, H. Jia, and C. Wang, *A Spatial-Temporal model for grid impact analysis of plug-in electric vehicles,* Applied Energy **114**, 456 (2014).

[17] S. Chen, H. Zhang, J. Guan, and Z. Rao, *Agent-based modeling and simulation of stochastic heat pump usage behavior in residential communities,* in *Building Simulation* (Springer, 2020) pp. 1–19.

[18] R. B. Nelsen, *Springer Series in Statistics*, Vol. 53 (Springer, 2013) p. 276, arXiv:arXiv:1011.1669v3 .

[19] H. Wang and B. Zou, *Probabilistic computational model for correlated wind speed, solar irradiation, and load using bayesian network,* IEEE Access **8**, 51653 (2020).

[20] W. P. J. Philippe, S. Zhang, S. Eftekharnejad, P. K. Ghosh, and P. K. Varshney, *Mixed copula-based uncertainty modeling of hourly wind farm production for power system operational planning studies,* IEEE Access **8**, 138569 (2020).

[21] M. Cui, V. Krishnan, B.-M. Hodge, and J. Zhang, *A copula-based conditional probabilistic forecast model for wind power ramps,* IEEE Transactions on Smart Grid **10**, 3870 (2018).

[22] M. T. Bina and D. Ahmadi, *Aggregate domestic demand modelling for the next day direct load control applications,* IET Generation, Transmission & Distribution **8**, 1306 (2014).

[23] Y. Zhao, Q. Liu, J. Kuang, K. Xie, and W. Du, *Modeling multivariate dependence by nonparametric pair-copula construction in composite system reliability evaluation,* International Journal of Electrical Power & Energy Systems **124**, 106373.

[24] R. Bernards, J. Morren, and J. G. Slootweg, *Statistical Modelling of Load Profiles Incorporating Correlations using Copula,* in *IEEE PES Innovative Smart Grid Technologies* (Torino, 2017).

[25] G. Papaefthymiou and D. Kurowicka, *Using copulas for modeling stochastic dependence in power system uncertainty analysis,* IEEE Transactions on Power Systems **24**, 40 (2009).

[26] M. Zamo and P. Naveau, *Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts,* Mathematical Geosciences **50**, 209 (2018).

[27] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, *Integrating metaheuristics and artificial neural networks for improved stock price prediction,* Expert Systems with Applications **44**, 320 (2016).

[28] Liander N.V., *Smart meter data,* https://www.liander.nl/over-liander/innovatie/open-data/data (2013).

[29] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Jarventausta, *Enhanced load profiling for residential network customers,* IEEE Transactions on Power Delivery **29**, 88 (2014).

[30] E. Ciapessoni, D. Cirio, A. Pitto, S. Massucco, and F. Silvestro, *A novel approach to account for uncertainty and correlations in probabilistic power flow,* IEEE PES Innovative Smart Grid Technologies Conference Europe **2015-Janua**, 1 (2015).

[31] G. Ye, M. Nijhuis, V. Cuk, and J. Cobben, *Stochastic Residential Harmonic Source Modeling for Grid Impact Studies,* Energies **10**, 372 (2017).

[32] KNMI, *Daggegevens van het weer in Nederland,* https://www.knmi.nl/nederland-nu/klimatologie/daggegevens (2017).

[33] M. Benghanem, *Optimization of tilt angle for solar panel: Case study for Madinah, Saudi Arabia,* Applied Energy **88**, 1427 (2011).

[34] W. Stine and H. Harrigan, *Power from the Sun* (John Wiley & Sons, 1986).

[35] CBS, *Onderzoek Verplaatsingen in Nederland 2014,* https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:61643 (2015).

[36] Office of Energy Efficiency and Renewable Energy, *Fuel Economy Guide 2017*, Tech. Rep. (U.S. Department of Energy, 2017).

[37] E. A. M. Klaassen, B. Asare-Bediako, C. P. De Koning, J. Frunt, and J. G. Slootweg, *Assessment of an algorithm to utilize heat pump flexibility-theory and practice,* 2015 IEEE Eindhoven PowerTech, PowerTech 2015 (2015), 10.1109/PTC.2015.7232783.

[38] W. van Westering and H. Hellendoorn, *Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation,* International Journal of Electrical Power & Energy Systems **114**, 105349 (2020).

[39] W. van Westering, B. Droste, and H. Hellendoorn, *Combined medium voltage and low voltage simulation to accurately determine the location of voltage problems in large grids,* (2019).

[40] B. Sereeter, W. van Westering, C. Vuik, and C. Witteveen, *Linear power flow method improved with numerical analysis techniques applied to a very large network,* Energies **12**, 4078 (2019).

[41] B. Sereeter, C. Vuik, and C. Witteveen, *Mathematical formulations and algorithms for fast and robust power system simulations*, Ph.D. thesis, Delft University of Technology (2020).

[42] H. Le Nguyen, *Newton-raphson method in complex form [power system load flow analysis],* IEEE Transactions on Power Systems **12**, 1355 (1997).

[43] W. H. Kersting, *Distribution system modeling and analysis* (CRC press, 2001).

[44] J. Kirtley, *6.061 Introduction to Power Systems Class Notes Chapter 5 Introduction To Load Flow* (MIT Open Courseware, 2018).

[45] J. E. Sarmiento, C. A. Alvez, B. de Nadai N., A. C. Zambroni de Souza, E. M. Carreno, and P. F. Ribeiro, *A complex-valued three-phase load flow for radial networks: High-performance and low-voltage solution capability,* IEEE Transactions on Power Systems **34**, 3241 (2019).

[46] M. Hekkenberg and R. Koelemeijer, *Analyse van het voorstel voor hoofdlijnen van het klimaatakkoord,* Planbureau van de Leefomgeving (2018).

6

Conclusion

The chapter summarizes improvements over existing methods found in literature. Furthermore, after presenting a brief overview of future research opportunities, this chapter considers multiple applications of the theory considered in this thesis.

6.1. Improvements over literature

This thesis presents several improvements over the available literature which are summarized in this section. The back bone of these improvements is a fast linear approach of real time load flow simulations, which is applied to several urgent DNO problems.

While linear modeling has been well studied, it could still be improved upon by applying numerical analysis techniques for additional speed and Bayesian interference theory for situations with limited sensing. Furthermore the applications in this thesis, namely large scale network simulations and real time control, are pushing the state of art in the field of district network operation.

Linear load flows are a powerful tool to speed up electricity network simulations. Naturally, the linearization has intrinsic limitations. One should exercise caution when applying linear methods to networks with large voltage drops (over 10%) and significant phase imbalances. However, even while the linear methods are not very accurate in these kinds of situations, they can still be applied for certain purposes. For example, inaccurate simulations will still be able to identify 'weak' networks with relatively high load.

Community battery control and deployment

Chapter 2 provides a solid foundation for integrating residential and community-level storage in the existing LV network. A fast linear LV model was developed and applied to the LV feeder of the community-battery in Rijsenhout. The model was proven to be sufficiently accurate for network stabilization purposes.

The battery control theory was formulated as a linear optimization problem. A receding horizon controller was developed to be used in a continuous way. The controller is suited very well to be integrated with other battery control goals, while still securing the voltages and currents within the network. It has been shown that a community battery is able to stabilize and control the loads in a real world low voltage network to a large extend.

A step-by-step method was proposed for quickly modeling the impact of new and existing batteries on the LV grid. Both the stabilizing and destabilizing potential regarding steady-state voltages and currents of the battery can be quickly estimated. By unlocking the potential of battery storage on a DNO level, a fast and secure energy transition is one step closer.

Numerical methods applied to very large load flow problems

Chapter 3 proposes a fast LPF method using a constant impedance load model to simulate both the entire LV and MV networks in a single simulation. Mathematical modeling of power systems and transformers is given and the algorithm of the LPF approach is explained in detail. The performance of this LPF approach is assessed by comparing it with the Newton power flow method and a commercial network design tool (Vision) on various distribution networks.

The results show that this LPF method can be as accurate as classical NPF methods using a constant power load model and additionally it is much faster than NPF

computations. In this chapter, it is shown that voltage problems can be identified more accurately when MV and LV are integrally evaluated. Moreover, Numerical Analysis techniques are applied to the LLPF problem in order to improve the computation time by studying the properties of the linear system. In the numerical analysis, reordering techniques (RCM) and numerous direct solvers (Cholesky, IC, LU, and ILU) and various Krylov subspace methods (CG, PCG, GMRES, and BiCGSTAB) are chosen and applied to the LLPF problem with both real and complex components. Finally, the original computation times of LLPF problems with real and complex components are reduced by 2.8 times and 5.7 times respectively as a result of the application of NA techniques.

Efficient Bayesian state estimation with limited sensor coverage

Chapter 4 addresses the challenge of formulating a distribution grid state estimator, for scenarios where fully observed sensor arrangements are not yet feasible, and load forecasts may be subject to large uncertainties due to lack of access to data. Building on preliminary work for single-phase networks, we derived an algorithm that exploits the information in load forecasting and feeder models to construct prior statistics of relevant voltage variables. We then used a Bayesian approach, in the form of the linear least squares estimator, to update prior voltage statistics in real-time based on measured deviations at a limited set of voltage sensors. We applied the method to a benchmark IEEE network and on a real testbed in the Netherlands and showed its ability to provide accurate voltages estimates using limited historical data and real-time sensors. As such, the method is highly applicable in the typical distribution network setting in which data and sensing will remain limited for the foreseeable future.

Strategic scenario evaluation using copulas and load flow

Chapter 5 presents a combination of a stochastic modelling approach with a fast network analysis model which makes it possible to assess in detail the impact of various future scenarios on large scale low voltage networks. The proposed modeling approach using Gaussian mixture models and copula functions is shown to be both accurate and flexible, and is therefore an effective method to represent the stochastic behavior and correlations of a variety of variables. A linear approach to modeling large low voltage networks has been described and applied to a very large network of Alliander DNO. The model was sufficiently fast to evaluate the network and provided valuable insights in the impact of the new technology on the grid. A geographical overview was shown to be a valuable tool for a DNO to focus its attention on the postal codes with relatively weak networks. Because of early problem detection, the methods provided in this chapter will help in securing the electricity transportation for the next decades.

6.2. Further applications of the presented research

Several real world applications have already been described in this thesis. The most notable are real time control of a community battery, integral MV/LV simulation for a better detection of voltage problems, real time Bayesian state estimation applied

on a network with very limited sensors and simulating the impact of the energy transition.

This section describes several other applications in which the theories described in this thesis can be applied to. Some of these applications are already known to be in active development.

Network losses prediction and network reconfiguration

The power loss in power distribution networks is a significant issue, both financial and environmental. For example, Alliander DNO, which operates the electricity distribution for over three million customers in the Netherlands, estimates the total power losses caused by energy transportation in 2014 of 1.1 TWh [1], with a net worth of 72 million euros and equivalent to 747 Gigaton of CO2 emissions. It is estimated that roughly a third of these losses are the result of the resisitive losses in the medium voltage power grid. Given these numbers, a reduction of a few percent of power loss will already result in a significant benefit.

The medium voltage distribution network of the Netherlands is to be operated in a radial tree-like topology for short-circuit detection purposes. However, to ensure redundancy the network is strongly interconnected and the actual radial topology is obtained by disengaging power switches. Since there are many possible switching configurations which yield such a radial topology and the network topology influences the energy losses, it is possible to save energy by finding the optimal switching configuration. The problem of *finding the optimal network topology* by changing the switching configuration is known as the Distribution Network Reconfiguration Problem (DNRP).

Generic optimization algorithms are used to solve the DNRP problem [2]. However, given the binary nature of the problem, the limiting factor of the optimization is often the number of load flow simulations that can be performed. The linear approach presented in this thesis is very useful speeding up the optimization algorithms, resulting in better solutions and more loss reduction. Preliminary research shows that *in two case studies the network resistive losses can be reduced by 15%-27%* by applying a genetic optimization approach to the network switching configuration [2].

Network capacity maps for connecting renewable energy generation

Currently, DNOs are struggling to incorporate renewable generation in their network because of the large number of applications. Requests for new grid connections are usually considered one-by-one and simulations are performed manually. *Determining if enough grid capacity is available can be a very time consuming job*.

Luckily, the process can be automatized using the load flow algorithms from this thesis. By exploiting the linear property as described in Chapter 2 and applying 2.15, the available capacity of the network can be determined in a single simulation iteration. Applying this algorithm for all customers results in *a valuable network capacity map*. These maps can be used for many purposes, such as determining locations with high network load, informing customers and providing data for making policy decisions.

Non-invasive low voltage real time network monitoring

Currently, *regional network operators have little knowledge about the day-to-day power flows through their network*. Traditionally, measurements in literature are performed by Power Measurement Units (PMUs) which are assumed to provide accurate and synchronized measurements of the voltage and its phase angle. In practice, however, these measurement devices are often too costly to be deployed on a large scale.

The deployment of smart meters provides the opportunity for a new perspective on measuring the loads in a MV/LV network. While smart meter measurements are not close to the level of accuracy PMUs can provide, they are abundantly available. However, *using smart meter data can have severe privacy implications*. Data about electricity consumption is considered sensitive data, since it can be directly related to the behaviour of the customer.

Instead one can use smart meter voltage data, which cannot be directly related to the behaviour of the customer. This presents a new opportunity to apply the state estimation algorithms presented in Chapter 4. *Using the (incomplete) smart meter voltage data and network topology, an estimate can be made of the load on the MV/LV transformer*. Since the MV/LV transformer has over 100 customers on average, the calculated load cannot be traced back to individual customers. This way a DNO can obtain valuable data on the network load by using smart meter data, but without harming the privacy of its customers.

6.3. Suggestions for future research

This section considers some angles for future research expanding on the theory of this thesis.

Community battery control schemes

Regarding the community battery, many more sophisticated control algorithms can be studied. For example, it would be interesting to see how a set of distributed community batteries can stabilize the grid or how to optimally distribute them among customers [3]. Since real world data connections never have 100% uptime, another interesting topic is decentralized control of multiple batteries.

Contrary to an example in literature [4], the community battery of Rijsenhout is not ready for providing backup power in case of a power outage. This is mainly because of safety reasons. Current practice is that in case of an outage, the affected network is unpowered so the engineers can safely inspect the network assets. Theory, policy and guidelines have to be developed to make it possible to use local electricity storage for backup power.

Large scale Bayesian state estimation

Combining the premises of Chapter 3 (Large scale simulation) and Chapter 4 (Bayesian State Estimation using limited sensing) yields an interesting opportunity for future research: large scale Bayesian state estimation. Applying the theory of Chapter 4 to large scale networks is not trivial, however. For example, the Bayesian approach

currently depends on several permutation matrices, which tend to scale badly. Furthermore, the load prediction models in Chapter 4 depend on customer load time series, which are hard to obtain for both technical and privacy reasons. If these practical problems can be solved, large scale applications have the potential to be very valuable to DNOs as they will greatly improve the DNO's understanding of the loads in their own electricity network. It will be easier to identify overloaded cables, allowing for more efficient network operation.

Applications to other types of critical infrastructure

While different, there are many fundamental similarities between the electricity network and other types of critical infrastructural networks, such as the water network, gas network, sewage network, traffic network and IT network. All these networks can be described by large graphs and can often be approached by linear equations with an acceptable accuracy. Furthermore, all these networks are governed by their own version of 'voltage' and 'current'. For water, sewage and gas networks, the equivalent principles are 'pressure' and 'mass flow'. Similar notions exist in traffic networks and even IT networks, although in the latter 'pressure' does not have a straightforward direct physical meaning. The implication of these similarities is that many of the applied theories in this thesis can be applied to other networks with minor modifications. The best example of such an application would be the Bayesian approach on interpreting sensor data in a network with relative low sensor coverage, which is a known problem in most other types of large scale critical infrastructure.

6.4. Final remarks

Using the theory and applications developed in this thesis, DNOs are able to cope better with the uncertainties caused by the energy transition. The challenge of accommodating renewable energy sources in the power grid is very large and will require a great effort, but can become a bit more manageable by smartly applying the simulation principles from this thesis.

The subtitle of this thesis is 'Towards a controllable electricity network'. While this thesis is a step in this direction, the 'controllable electricity network' is still a thing of the future. However, once the simulations are in place and large scale state estimation is deployed, new network control methods can be developed. Measures such as decentralized congestion management and anomaly detection provide new ways to maintain the network. Therefore, a controllable electricity network is closer to realization than it has ever been before.

References

[1] Alliander N.V., *Alliander N.V. annual report 2019,* https://2019.jaarverslag.alliander.com/verslagen/annual-report-2019 (2019), accessed: 29-06-2020.

[2] W. van Westering, M. van der Meulen, and W. Bosma, *Evaluating electricity dis-*

tribution network reconfiguration to minimize power loss on existing networks, in *CIRED Workshop 2016* (2016) pp. 1–4.

[3] R. de Groot, *Optimal placement and operation of novel grid technologies in distribution grids*, Ph.D. thesis, Department of Electrical Engineering (2017), proefschrift.

[4] R. J. de Groot, B. M. Vonk, H. J. Beckers, and J. G. Slootweg, *Development of a charge path optimization controller block for a battery energy storage system,* IFAC Proceedings Volumes **47**, 8583 (2014).

Acknowledgements

First of all, I would like to thank my promotor Prof. dr. ir. J. Hellendoorn for his support during this 6-year project. He is an exceptional adviser. While I was struggling with publishing my results, he gave the invaluable advice to focus on collaborating with other researchers, which changed the course of the project. Without this essential piece of advice, I would likely not have succeeded. Prof. dr. ir. J. Hellendoorn also gave very valuable advice on matters beyond my PhD and regarding my personal struggles. I am still very grateful for that. Also thanks to my co-promotor prof. dr. ir. J.G. Slootweg and other members of the committee for their feedback, time and effort.

I would like to thank all my co-authors: Roel Dobbe, Baljinnyam Sereeter and Raoul Bernards. Without your collaboration, this thesis would not be here. I enjoyed our cooperation and am happy that I was able to contribute to your PhD projects as well.

A special thanks to all my colleagues, whom without I could not have completed this work. Thanks to my supervisors Robin Hagemans, René Snijders and Dirk Jan Boon for having confidence in me during this long project and for creating the context in which my PhD was possible. Thanks to Robin especially for letting me initiate the PhD.

There are many other colleagues to thank. Pieter Gockel, for arranging funding for 'project state estimation' in a time where the result was still unclear. Thanks to Barbera Droste, Jacco Heres and Jaap van Wijck who spent a lot of time and effort on said project. Especially the help regarding data and validation was greatly appreciated. Also thanks to my 'Alliander friends' Guillaume Goijen, Jan Willem Eising and Rolf van der Velde for their support.

Finally, many thanks to my friends and family. The most thankful I am for my mother for always supporting me at any time.

Werner H.P. van Westering
Delft, January 2021

A

Reference networks

The numerical analysis performed in Chapter 3 was applied to the networks displayed in this appendix.

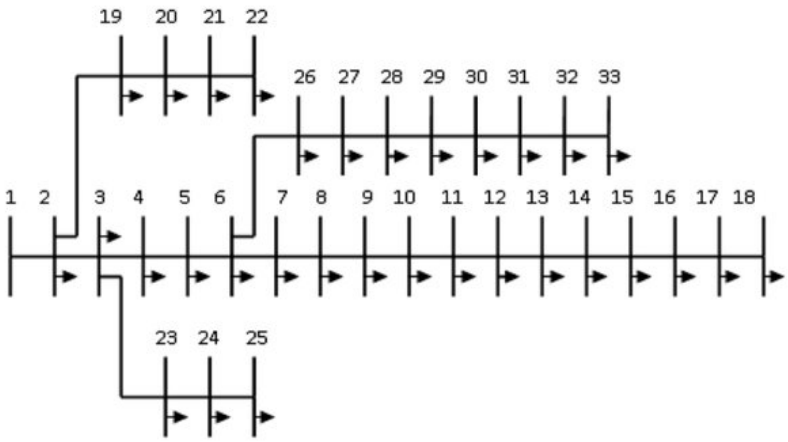


Figure A.1: The standard IEEE 33 bus network. Picture taken from [1].

References

[1] M. P S and S. Hemamalini, *Optimal siting of distributed generators in a distribution network using artificial immune system,* International Journal of Electrical and Computer Engineering **7**, 641 (2017).

[2] I. Pisica, C. Bulac, and M. Eremia, *Optimal distributed generation location and sizing using genetic algorithms,* (2009) pp. 1 – 6.

A

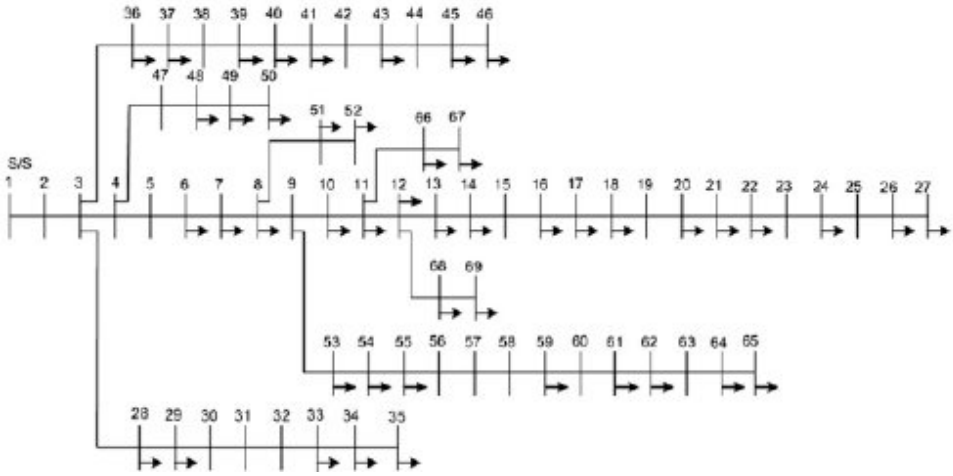


Figure A.2: The standard IEEE 69 bus network. Picture taken from [2].

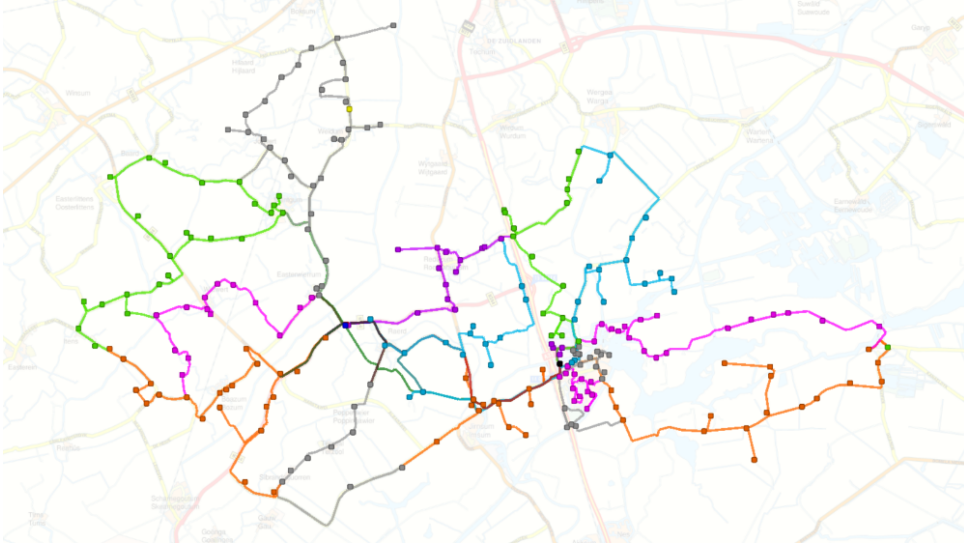


Figure A.3: Case 991, also known as substation Rauwerd. The lines of the MV network have been plotted on a map of the Rauwerd area, located in Friesland, The Netherlands.

Curriculum Vitæ

Werner H.P. van Westering

04-04-1989 Born in Haarlem, The Netherlands.

Work Experience

Education

2007–2010 BSc. Mechanical Engineering
Delft University of Technology

2011–2013 MSc. Systems & Control
Delft University of Technology
Thesis: Gas network modeling
Supervisor: Prof. dr. J. Hellendoorn

Work Experience

2013–2020 Junior/senior/principal data scientist
Data & Insights, Alliander NV

Awards

2014 Thesis prize (3rd)
Koninklijke Vereniging van Gasfabrikanten in Nederland

2014 Best Student Paper Award (1st)
11th IEEE International Conference on Networking, Sensing and Control

2011 Debating tournament (1st)
Student association C.S.R. Delft

2008 1st-year Students Design Competition (1st)
Faculty of Mechanical, Maritime and Material Engineering

List of Publications

Journal articles

5. **W. van Westering and J. Hellendoorn**, *Low voltage power grid congestion reduction using a community battery: Design principles, control and experimental validation*, International Journal of Electrical Power & Energy Systems **114**, 105349 (2019).

4. **B. Sereeter, W. van Westering, C. Vuik and C. Witteveen**, *Linear Power Flow Method Improved With Numerical Analysis Techniques Applied to a Very Large Network*, Energies **12**, 4078 (2019).

3. **R. Dobbe, W. van Westering, S. Liu, D. Arnold, D. Callaway and C. Tomlin**, *Linear Single- and Three-Phase Voltage Forecasting and Bayesian State Estimation with Limited Sensing*, IEEE Transactions on Power Systems (2020).

2. **R. Bernards, W. van Westering, J. Morrena, H. Slootweg**, *Analysis of Energy Transition Impact on the Low Voltage Network using Stochastic Load and Generation Models*, Energies (2020), Volume 13.

1. **W. van Westering and J. Hellendoorn**, *Operational Real Time State Estimation on a Very Large MV/LV Distribution Grid*, In Preparation.

Conference proceedings and other publications

8. **W. van Westering, B. Droste and J. Hellendoorn**, *Combined Medium Voltage and Low Voltage simulation to accurately determine the location of Voltage Problems in large Grids*, CIRED conference - Madrid June 2019 (2019).

7. **J. Sijm, P. Gockel, M. van Hout, Ö. Özdemir, J. van Stralen, K. Smekens, A. van der Welle, W. van Westering and M. Musterd**, *The supply of flexibility for the power system in the Netherlands, 2015-2050. Report of phase 2 of the FLEXNET project*, Technical study published by ECN (2017).

6. **J. Heres, W. van Westering, G. van der Lubbe and D. Janssen**, *Stochastic Effects of Customer Behaviour on Bottom Up Load Estimations*, CIRED conference - Glasgow 12-15 June 2017 (2017).

5. **W. van Westering, J. Heres, T. Dekker, M. Danes and J. Hellendoorn**, *Assessing the Voltages and Currents in a large Dutch Regional Power Distribution Grid*, Energy-open Conference (2017).

4. **W. van Westering, M. van der Meulen, W. Bosma**, *Evaluating Electricity Distribution Network Reconfiguration to Minimize Power Loss on Existing Networks*, CIRED Workshop 2016 (2016).

3. **W. van Westering, A. Zondervan, A. Bakkeren, F. Mijnhardt and J. van der Els**, *Assessing and Mitigating the Impact of the Energy Demand in 2030 on the Dutch Regional Power Distribution Grid*, 2016 IEEE 13th International Conference on Networking, Sensing, and Control (2016).

2. **W. van Westering, J. Hellendoorn**, *Optimal sensor placement using gas distribution network models: A case study*, ICNSC 2015 - 2015 IEEE 12th International Conference on Networking, Sensing and Control (2015).

1. **W. van Westering, J. Hellendoorn, B. Brasjen and R. van der Linden**, *Natural gas distribution network modelling and leak minimization*, 2014 IEEE 11th International Conference on Networking, Sensing and Control (ICNSC) (2014).