Master Thesis of Mathematics

Scoring voltage measuring locations in power distribution grids

Jeroen Slot
s4476905

Supervised by
Werner van Westering, Alliander N.V.
Wieb Bosma, Radboud University

November 27, 2020

Radboud Universiteit

Preface

When I started my Master Mathematics in September 2018 I did not immediately know what my specialization would be. I changed it a few times and finally chose Mathematical Foundations of Computer Science (MFoCS). At that point I had already quite a few years of studying mathematics behind me, and I wanted something new. I wanted to experience the working life and as such chose a specialization geared more towards learning coding languages and wished to apply this knowledge with an external research internship.

I am an avid supporter of fighting climate change and as such searched for companies with close ties to energy. Alliander fits that frame perfectly. It is a public utility company (*nutsbedrijf*) that is currently in a transition to be less dependent on non-renewable sources of energy. I went to Alliander with this in mind and spoke to my current internship supervisor Werner van Westering. He was willing to take me under his guard and placed me into the system optimization team. In the first month of my internship I had great fun getting to know my new colleagues, finding out how Alliander operates, and what it has to offer.

This first month was February 2020, less then two months before the first lockdown due to Covid-19. My goals of experiencing the working life were cut short and I once again found myself working five days a week from home, the exact pattern I was trying to break. Working from home and the global pandemic effected my mental state I became more anxious. Often I would get stuck on the non-mathematical parts of my thesis, not finding the courage to ask my new colleagues. Instead I often put it off, decreasing my productivity, thinking I could not get any further at that moment. This caused me to delay the end date of my internship and to not be completely satisfied with the results.

In retrospect I feel that I could have worked harder and achieved more, though this does not take away from the research I did achieve. There are still unanswered questions that I would have liked to incorporate into the thesis. Looking back it feels like quite the journey, making it hard to see the things that I did achieve, both academically and on a personal level. I only wish I had done more.

I am very grateful for my supervisor Werner van Westering for guiding me in this process. He was always willing to help and knew exactly where I needed to improve when it came to the non-mathematical parts of my thesis.

Thanks to my second supervisor dr. Wieb Bosma who was always kind and helpful. His feedback always got me thinking and he more then once helped me get unstuck when I did not know how to further proceed with my project. Moreover I want to thank Jelte Zwetsloot, a friend and colleague who always helped me evaluate my position and convince me life was not as bleak as I made it out to be. Especially in the beginning of the Covid-19 pandemic he was of great moral support, having recorded a vlog every morning at 8.30 with me and two more of our friends, Vincent Roovers and

Bas Swinkels.

Lastly I wish to thank my partner, Merel van der Heijden. She was always there to listen to my problems and provided much needed support. Having talked to her as much as I did, I think that at this point she might be the second most knowledgeable person regarding the proposed methods and theories in this thesis.

So great thanks to all the people who supported me and helped me with my thesis. I can not thank you enough.

Contents

Chapter 1

Introduction

Can we improve the predictions of a forecaster in a power distribution grid by introducing sensors and what is the optimal configuration of these sensors given certain (budget) constraints? That, in a nutshell, is the focus of this thesis.

The placement of smart components and sensors in a power distribution net has become more common over the years at Alliander. Components like smart cable guards and iMSR's are deployed to detect disturbances in the cable and consequently reduce outage time. Algorithms have been developed that can produce an optimal configuration of these sensors such that outage time is as low as possible.

But what if outage reduction is not the only goal of these sensors? They are capable of more. In particular they may be able to measure electrical quantities such as current and/or voltage. Knowing the true state of a distribution system is a valuable asset and affects the decisions making process regarding anything in said system. Are cables overloaded? Is fraud being committed? Where is the most power being used? Therefore it might be valuable to take measuring accuracy into considering when placing the aforementioned components.

Currently a machine learning algorithm is used to predict the state of power distribution systems. This thesis explores how the accuracy of this algorithm can be improved if a set of measuring components are added within the network. Some locations might be more valuable to measure than others, and thus an optimal configuration can be searched for. We introduce a scoring function that determines a score for any set $\mathcal{O} \subseteq \mathcal{V}$ of nodes. Here a given network is interpreted as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$.

The question arises: given a budget B, what is the subset $\mathcal{O}$ of measurable locations that optimizes the scoring function while $\text{cost}(\mathcal{O}) \leq B$. Similarly we can ask the question: given a certain score threshold, is there a subset $\mathcal{O}$ that exceeds this thresholds and if so, what is the subset with the lowest placing cost?

The main focus of this thesis is to develop such a scoring function. For that we need an understanding of state estimation and explore some previous research on this topic. A few assumptions will be made throughout this thesis and when such an assumption is made it will be clearly stated and explained why the assumption is being made. Lastly, we explore some algorithms that optimize with respect to our scoring function.

Chapter 2

Sensor Covering

In this chapter we provide an overview of the problem at hand. We discuss what we consider a state of a power distribution grid, what we consider a forecaster for these states and how we can score such forecasters. We also briefly mention how we can optimize over measuring different locations within the network.

2.1 Power distribution grid

The distribution network of electrical power in the Netherlands is categorized in three sections. There is the transmission grid consisting of high voltage cables, which are long cables running either underground or above ground via transmission towers. They transport large amounts of electricity over long distances and are managed by TenneT.[1] Connected to the transmission grid are (power) distribution grids. A substation is an electrical installation that connects high voltage transmission grids to medium voltage distribution grids. Such a distribution grid distributes energy to industrial consumers and rural networks and can be connected to small power plants, solar and wind farms.

One level lower there is the low voltage network, where each rural network distributes the power to individual houses. A direct current (DC) runs through a low voltage network, but this is not the case for the distribution grid. We focus on distribution grids where a three-phase alternating current (AC) is applied. Voltages can be measured by sensors and are currently being predicted by a forecaster, as in a typical distribution grid only the substation is measured. The main difference between predictions and measurements is their accuracy. Measurements are typically done via real world devices that can determine the actual Voltage up to a relatively small error. A prediction is less accurate as it determines the Voltage via some theoretical approximation.

[1] There also exists extreme high voltage cables, of which in the Netherlands only two exist. These cables are typically connected to facilities such as large nuclear power plants and hydro-electric plants.

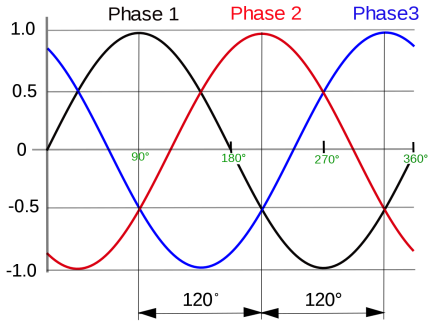

Figure 2.1: One voltage cycle of a three-phase system, labeled 0 to 360° along the time axis.

Mathematically speaking a power distribution grid can be interpreted as a directed graph by associating a node to each connection that goes out of the network. The bundle of cables that connect two of these nodes is considered an arrow, i.e. an directed edge. It is policy that in a distribution grid no loops occur. This is mainly to ensure that faulty cables can be efficiently detected and to ensure that mechanics are not electrocuted when operating on said faulty cables. Therefore the graph is a directed tree where we consider the substation as its root. The different branches sprouting from the substations are called routes.

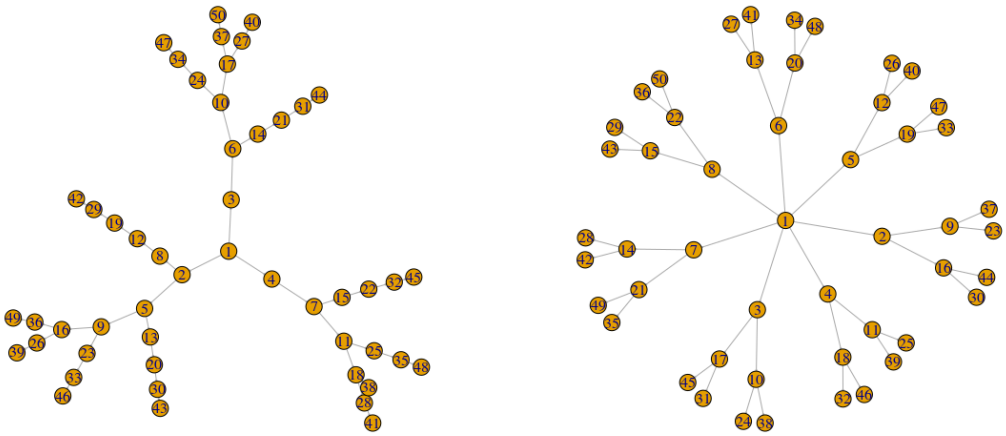

Figure 2.2: Two graphs with 50 nodes, the left indicates a typical small network with three routes, the right indicates a larger network where all paths in between intersections have been shortened to length 1.

Traditionally, all power originated from the transmission grid and was distributed downwards. Large power generators connected to the transmission grid generated power which was the distributed. With the rise of alternative energy many households have access to solar panels on their roofs and other forms of energy that does not originate from the transmission grid. This changes the possible power flow through a distribution grid and might stress the cables as some cables are old and not prepared for this change. Knowledge about the flow of power in a given power distribution grid can thus be a valuable resource to have.

2.2 Sensing a distribution grid

What is there to measure in a distribution grid? When we talk about measuring quantities in a power distribution grid more often then not we talk about measuring a given state. A state of a power distribution grid is the state of that grid at a given moment in time. What variables we associate with that state depends on the scope and focus of the project. A categorization of the different state variables types is given in [1]. In [1] They also note that collecting and determining this data often is a multi-stage process comprising of:

1. Sensing, measurement and data acquisition – the basic processes of obtaining raw grid data, with conversion from analog to digital form.

2. Filtering, linearization, scaling and unit conversion – conversion and processing of raw digital data from uncompensated integer counts to compensated, linearized values, scaled to engineering or physical quantity units as opposed to dimensionless integers.

3. Representation – conversion of physical variables into forms suitable for analysis and use in control in any of several domains: time, frequency, geospatial, or electrical distance from a reference points such as a substation.

4. State formation – construction of actual grid state elements; may involve several computational processes such as extraction of parameters from data sets, estimation where necessary, and then assembly and aggregation of grid state elements.

5. Distribution and persistence – grid state elements must be made available to various decision and control processes, and may have to be persisted in any of several tiers of data storage, depending on the various uses for the data.

We focus on sensing the nodal voltages $\mathbf{V_n}$ and assume all line resistances $\mathbf{Z_{mn}}$ to be known a priori. Through what is called *state estimation* other electrical entities can be calculated, namely the line currents $\mathbf{I_{mn}}$, line powers $\mathbf{S_{mn}}$ and nodal powers $\mathbf{s_m}$. In this thesis we will only be considering nodal voltages $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and nodal powers $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$. As long as the resistances are known the nodal powers $\mathbf{s}$ can be calculated from the nodal voltages $\mathbf{V}$. Considering the resistances are constant with respect to time[2], with $(\mathbf{V}, \mathbf{s})$ we will denote *the state* of a power distribution system at a given time.

Measuring in a power distribution grid is expensive, so we wish to limit the amount of sensors placed. We consider a prediction algorithm that predicts the nodal voltages and observe how we can improve this algorithm by smartly placing sensors in the system. This means that in the multi-stage process of collecting and determining data, as mentioned above, we focus on step 4. We assume we have access to sensors which can achieve step 1,2 and 3 can be used and that they provide us with accurate measurements that we can compare to the predicted data. Moreover, we assume that this sensing is more accurate then predicting and thus measuring errors are relatively insignificant. Step 5 is not within the scope of this thesis.

The natural question arises, what is a valid observability requirement? We need a scoring mechanism that can compare achieved observability of different sets of observed nodes. We can use any algorithm that fits the purpose of predicting nodal voltages and in particular we can base it on the forecaster produced by Alliander ([3]), as it is used within Alliander for state estimation. For a node n the voltage $\mathbf{V_n}$ is a vector of three elements, V_n^a, V_n^b and V_n^c where a, b, c indicate the three different phases in the power distribution grid. The nodal powers $\mathbf{s_n}$ are similarly decomposed. The vector $(s_1^\alpha, \ldots, s_N^\alpha)$ for $\alpha \in \{a, b, c\}$ of the true state can be compared to the predicted state $(\tilde{s}_1^\alpha, \ldots, \tilde{s}_N^\alpha)$ which is generated by a forecaster. There is a baseline prediction where no nodes are measured, and all knowledge of the state comes from this prediction (given a set of line resistances). Then, when some voltages are measured instead of predicted, accuracy should rise. Different sets of nodes give rise to different predicted states $(\tilde{s}_1^{\alpha,\mathcal{O}}, \ldots, \tilde{s}_N^{\alpha,\mathcal{O}})$ where $\mathcal{O}$ indicates the set of observed nodes. We want a score for each $\mathcal{O}$ that compares a large range of true states $(s_1^\alpha, \ldots, s_N^\alpha)$ to a large range of predicted states $(\tilde{s}_1^{\alpha,\mathcal{O}}, \ldots, \tilde{s}_N^{\alpha,\mathcal{O}})$ for each $\alpha \in \{a, b, c\}$. For that we first need a notion of forecasting.

[2] not taking into account any slight alterations due to temperature

2.3 Forecasting

Forecasting is the process of making predictions based on past and present data and most commonly by analysis of trends. The power distribution state predictions can be interpreted as such, but they are not the only examples of forecasters. A typical example of a forecaster is a weather forecasting. Partly because it is intuitive, and partly because meteorologists were among the first to realize the importance of this problem. Weather forecasting is a good example to illustrate forecasting concepts, in fact, whenever you are trying to empirically compare different prediction algorithms or scientific theories you are in a situation similar to this ([14]). We start with an example of forecasting found in betting. It leads to a notion indicating coherent bookkeepers ought to represent their uncertainty by probability calculus, i.e. with a probability distribution.

Imagine a horse race in mid 1800s England. Such a race attracted betters and bookmakers alike and any bookmaker is an example of a forecaster and has traits similar to a modern forecaster. Any reasonable bookmaker would only post odds that favor himself in the long run. For example, say there is a race between two horses A and B and our bookmaker announces: "bet a single pound on A and triple your money when he wins. But! If you bet on B you can win back four times your wager!" Any smart better can take advantage of this by betting one pound on horse A and one pound on horse B. Independent of the outcome, he will have made a profit. When horse A wins he profits one pound and when horse B wins he goes home with a two pound profit. More formally we can say:

Definition 2.3.1. *A bookmaker's betting odds are coherent if a client cannot place a bet or a combination of bets such that no matter what outcome occurs, the bookmaker will lose money.*

The next theorem, due to de Finetti (1937), formalizes the claim that coherence requires prices that satisfy the axioms of probabilities. For a more formal development see Shimony (1955). The conditions of the theorem are as follows.

DBT1: The odds are fair to the bookmaker. That is the bookmaker is willing to both sell and buy bets on any of the events posted.

DBT2: There is no restriction about the number of bets that clients can buy or sell, as long as these bets are finite.

The first condition is required to guarantee that the odds reflect the bookmaker's knowledge about the relevant uncertainties, rather than desire to make a profit. The second condition is used, in de Finetti's words, to purify the notion of probability from the factors that affect utility. It is strong: it implies, for example, that the bookmaker values the next dollar just as much as if it were his or her last dollar.

Theorem 2.3.2 (**Dutch Book theorem**). *If DBT1 and DBT2 hold, a necessary condition for a set of prices to be coherent is to satisfy Kolmogorov's*[3] *axioms, that is:*

A1) *For every event* θ *we have* $0 \leq \mathbb{P}(\theta) \leq 1$.

A2) *For the universe* Θ *it holds that* $\mathbb{P}(\Theta) = 1$.

A3) *If* θ_1 *and* θ_2 *are independent, then* $\mathbb{P}(\theta_1) + \mathbb{P}(\theta_2) = \mathbb{P}(\theta_1 + \theta_2)$.

[3] which are the axioms of probability calculus

Proof. See [14]. □

The De Finetti's Dutch Book theorem guaranteed that, if one wants to avoid a sure loss (that is, be coherent), then probability calculus ought to be used to represent uncertainty. The conclusion is that, regardless of one's beliefs, it is incoherent not to express such beliefs in the form of some probability distribution. Whenever you are trying to empirically compare different prediction algorithms or scientific theories, you are in a situation similar to this.

Forecasters can be divided into two categories. Single-valued forecasters, and probabilistic forecasters. A forecaster of the latter variant has a probability distribution indicating where a value lies, and is more detailed then a single-value forecast. A single-valued predictive algorithm may output that the voltage on node k is $\mathbf{V}$ at a given time t, but this does not indicate the certainty of that prediction. A prediction that the forecaster is certain about is treated equally to any prediction that is not certain. That is to say, the uncertainty of the prediction is not taken into account. As indicated by the Dutch Book theorem, probabilistic forecasters are more reliable.

Initially this inspired me, the author, to convert single-valued forecaster to probabilistic forecasters, but this turned out to be ineffective. This was chosen to be left out, as it did not contribute to the main focus of this thesis. We focus on single-valued forecasters, such as the one presented in [3], which is currently used by Alliander.

For this thesis the choice was made to simulate any results that comes from a forecaster, instead of using an actual forecaster. This is done by some random process where the predicted stated is dependent on the true state and differs by some random variable which is most likely Gaussian by nature. As to why this choice was made, see Section 2.6. This allows us to consider any algorithm as long as we can determine its behavior via a statistical process.

2.4 Scoring functions

Scoring function are originally a notion for probabilistic forecasters. We briefly discuss how single valued forecasters can be interpreted as such. For more information see [14]. Let q be a possible output of a forecaster.

Definition 2.4.1 (**Scoring rule**). *A scoring rule (or function) s for the probability distribution q is a function assigning a real number $s(\theta, q)$ to each combination (θ, q), where θ indicates a possible event.*

So when the output of a forecaster $\mathcal{F}$ is probabilistic, its output can be assigned a scoring function. Recall that the state of a given power distribution system is denoted by $(\mathbf{V}, \mathbf{s})$ and influences the outputted distribution. This means that a scoring function is dependent on the state of the power distribution grid and the used forecaster. We want it to be dependent on neither one. Dependency on the state can be alleviated by a Monte Carlo process, where we sample many different states. Sampling different states that together indicate an average day and calculating a score for each sample produces a vector of scores for each event θ, which we can average with a certain confidence interval. More mathematically, each true state $(\mathbf{V}, \mathbf{s})$ inherits a scoring function $s_{(\mathbf{V},\mathbf{s})}$ that can score any event θ and q, where q represent predictions made by $\mathcal{F}$. Then we can define the average score

$$s(\theta, q) \coloneqq \frac{1}{|\Phi|} \sum_{\mathbf{V} \in \Phi} s_{(\mathbf{V},\mathbf{s})}(\theta, q)$$

where Φ represents the set of all possible voltages (given a set of resistances). This set can be represented as an n dimensional interval, where n is the size of the network. Note that the interval Φ contains unaccountably many elements. We can merely look at the set of most common states to ensure convergence, or replace the infinite sum with an integral,

$$s(\theta, q) \coloneqq \frac{1}{l(\Phi)} \int_{\mathbf{V} \in \Phi} s_{(\mathbf{V},\mathbf{s})}(\theta, q),$$

where $l(\Phi)$ indicates the length of the interval Φ.

2.5 Optimizing with respect to budget restraints

Given a scoring rule that assigns a score to any set of observed nodes, optimization is natural and search space reduction techniques are needed. This is due to the amount of subsets $\mathcal{O} \subseteq \mathcal{V}$ of which a score can be calculated. Recall that we define a power distribution network as a directed tree $G = (\mathcal{V}, \mathcal{E})$ with the substation as its root. In practice, a budget constraints the amount of sensors we can place, otherwise the best solution would always be to observe every node in the system. Assuming the scoring function has output range $[0, 1]$, with 0 being the best score and 1 being the worst, we can optimize with respect to a budget: Given a budget B, what is the subset $\mathcal{O} \subseteq \mathcal{V}$ of measurable locations such that $\text{cost}(\mathcal{O}) \leq B$ and

$$\text{score}(\mathcal{O}) = \min_{\substack{\text{cost}(\mathcal{U}) \leq B; \\ \mathcal{U} \subseteq \mathcal{V}}} \text{score}(\mathcal{U}).$$

The set of $\{\mathcal{U} \subseteq \mathcal{V} \mid \text{cost}(\mathcal{U}) \leq B\}$ is never empty as it always contains the empty set as an element, i.e. not measuring is always an option. Note that this budget constraint is equivalent to a constraint $m \in \{1, \ldots, |\mathcal{V}|\}$ that limits the amount of nodes sensed, If costs are uniform. That is to say, if placing a sensor costs X euro, independent of location, then finding a subset $\mathcal{U}$ such that $\text{cost}(\mathcal{U}) \leq B$ is equivalent to finding a subset $\mathcal{U}$ such that $|\mathcal{U}| \leq \frac{B}{X}$.

We can also optimize with respect to a certain threshold in observability: given $T \in [0, 1]$, is there a subset $\mathcal{O} \subseteq \mathcal{V}$ with $\text{score}(\mathcal{O}) \leq T$, if so what is $\mathcal{O} \subseteq \mathcal{V}$ with $\text{score}(\mathcal{O}) \leq T$ and

$$\text{cost}(\mathcal{O}) = \min_{\substack{\text{score}(\mathcal{U}) \leq T; \\ \mathcal{U} \subseteq \mathcal{V}}} \text{cost}(\mathcal{U}).$$

Finding $\mathcal{O}$ is quite a task. Let $N = |\mathcal{V}|$, then for $m \in \{0, \ldots, N\}$ there are $\binom{N}{m} = \frac{N!}{m!(N-m)!}$ subsets of size m. To get a sense of how large this is, observe the following table for networks with 50 and 100 nodes. For $m = 5$ this is manageable but already quite large, beyond that it is simply too large to calculate each score. We will need search space reduction techniques to limit the search space. For example, if nodes $v_1, \ldots v_m$ are the m nodes with the highest individual scores, the set $S = \{v_1, \ldots, v_m\}$ is a good candidate for highest score on m nodes.

2.6 Summary

Intially we posed the question: can we improve the predictions of a forecaster on a power distribution grid by introducing sensors, and what is the optimal configuration given certain (budget)

m \ N	50	100
5	$2 \cdot 10^6$	$7 \cdot 10^7$
10	$10 \cdot 10^9$	$2 \cdot 10^{13}$
15	$2 \cdot 10^{12}$	$3 \cdot 10^{17}$
20	$4 \cdot 10^{13}$	$5 \cdot 10^{20}$

Table 2.1: The number of subsets on N nodes of size m

constraints? It has been chosen to give any forecaster the measurements post priori instead of a priori. If applied a priori, the forecaster can train itself more accurately, resulting in an improved prediction compared to the original. Although this would give preciser scores, it was chosen to append knowledge post priori for the following reasons.

First and foremost, inserting measurements a priori requires an in-depth analysis of the forecaster, which was not in the scope of this thesis. Simply put, it is easier to predict the model post priori based on the knowledge of experts within Alliander than to try and work with the algorithm itself. Moreover, each score would take considerably longer to calculate as each $\mathcal{F}_{\mathcal{O}}$ needs time to train itself. Post priori insertion takes that out of the equation, at the cost of some accuracy.

Whether the measurements are applied a priori or post priori makes a difference. Both in the model and in the resulting optimal scoring functions.

In this thesis we discuss a notion of state estimation in Chapter 3 and in Chapter 4 introduce a scoring function named Incident Voltage score, that predicts the $2-$norm score and is independent of any forecaster. In Chapter 4 we also explore some of the Incident Voltage score's properties and determine its validity with a set of case studies in Chapter 5. This is a relatively new approach and in Chapter 8 we discuss some future work that can increase the realism of the model is.

Chapter 3

State estimation

3.1 Power flow equations

State estimation is the art of estimating the currents, voltages, impedances and powers in a (distribution) network. In this section we show how nodal powers can be calculated from nodal voltages, given a set of resistances. Then we present and prove an equivalent form using the full and partial incidence matrices. As briefly described in Chapter 2, a power distribution grid can be described as a directed graph, in particular a directed tree.

Definition 3.1.1. *A graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ is called a tree if it is connected and contains no cycles. It is called a network if it is an oriented (directed) tree. The root of the directed tree is called the substation.*

The vertices $v \in \mathcal{V}$ are denoted as nodes. Directed edges $e \in \mathcal{A}$ are named arrows.

Definition 3.1.2. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with substation identified as node 1, let $n \in \mathcal{V}$ such that $(1, n) \in \mathcal{A}$. Then the subgraph $\mathcal{G}_n$ originating from n is called the route originating from n. That is to say $m \in \mathcal{G}_n$ iff node n lies on the path from the substation to m.*

In the case of Figure 3.1 node 1 would be our substation, and we would have one route. It is convention to label the first node as the substation.

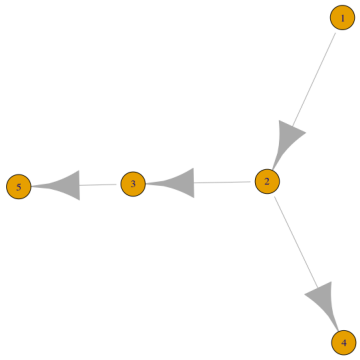

Figure 3.1: A small directed graph with five nodes with node 1 as the substation, and node 2 the origin of the only route.

To understand the relations between the voltages and the power variables in the distribution systems we need two concepts, Ohm's law and the power equations. We translate these concepts to a system of equations which can be solved using incidence matrices. In contrast to the power equations in a transmission grid, these equations now live within a complex space, as a power distribution system uses alternating currents. A complex resistance, also called impedance, on a line can be represented as

$$\mathbf{Z}_{mn} = \begin{pmatrix} Z_{mn}^{aa} & Z_{mn}^{ab} & Z_{mn}^{ac} \\ Z_{mn}^{ba} & Z_{mn}^{bb} & Z_{mn}^{bc} \\ Z_{mn}^{ca} & Z_{mn}^{cb} & Z_{mn}^{cc} \end{pmatrix}.$$

where $Z_{mn}^{\alpha\beta} = r_{mn}^{\alpha\beta} + jx_{mn}^{\alpha\beta}$, where $j = \sqrt{-1}$ denotes the complex impedance of line (m, n) across phases α and β for $\alpha, \beta \in \{a, b, c\}$. In practice these impedances are often represented by diagonal matrices

$$\mathbf{Z}_{mn} = \begin{pmatrix} Z_{mn}^{aa} & 0 & 0 \\ 0 & Z_{mn}^{bb} & 0 \\ 0 & 0 & Z_{mn}^{cc} \end{pmatrix},$$

and hence invertible. Representing $\mathbf{Z}_{mn}$ as a diagonal matrix ensures no interference between phases. To understand what this means we need Ohm's law. We can write the current $\mathbf{I}_{mn}$ through line (m, n) as

$$\mathbf{I}_{mn} = \begin{pmatrix} I_{mn}^{a} \\ I_{mn}^{b} \\ I_{mn}^{c} \end{pmatrix},$$

and $\mathbf{V}_m$, the voltage on a given node m as a complex vector of three phases,

$$\mathbf{V}_m = \begin{pmatrix} V_m^a \\ V_m^b \\ V_m^c \end{pmatrix}.$$

Ohm's law, also known as Kirchhoff′s Voltage laws, provide the relationship

$$\mathbf{V}_m - \mathbf{V}_n = \mathbf{Z}_{mn}\mathbf{I}_{mn}. \tag{3.1.1}$$

Because all impedances are considered diagonal matrices we can talk about a three-phase system as if it was a single-phase system. More formally, for diagonal $\mathbf{Z}_{mn}$ and phase $\alpha \in \{a, b, c\}$ we have

$$\mathbf{V}_m^\alpha - \mathbf{V}_n^\alpha = \mathbf{Z}_{mn}^{\alpha\alpha}\mathbf{I}_{mn}^\alpha, \tag{3.1.2}$$

which would not be the case for non-diagonal matrices. All the lemmas, propositions and theorems we prove in this chapter and Chapter 4 will be shown for a single phase, unless stated otherwise, and can be extended to three phases. The code that was built alongside the thesis is build for a three-phase system. That code shall be used to support our mathematical findings using case studies in Chapter 5.

Equations 3.1.1 are linear with respect to comparing the difference vector $\mathbf{V}_{mn} \coloneqq \mathbf{V}_m - \mathbf{V}_n$ to the current if all matrix $\mathbf{Z}_{mn}$ are known a priori. Thus linear solvers can compute

$$\mathbf{Z}_{mn}\mathbf{I}_{mn} = \mathbf{V}_{mn},$$

for given $\mathbf{Z}_{mn}$ and $\mathbf{V}_{mn}$.

From calculating currents we can continue to calculating the line powers $\mathbf{S}_{mn}$ and consequently the nodal powers $\mathbf{s}_m$. The line powers $\mathbf{S}_{mn}$ can be obtained through

$$\mathbf{S}_{mn} = \mathbf{V}_n \circ \mathbf{I}_{mn}^*. \tag{3.1.3}$$

Here $\mathbf{X} \circ \mathbf{Y}$ is the element-wise product of two vectors $\mathbf{X}, \mathbf{Y}$ and $\mathbf{X}^*$ indicates the complex conjugate of $\mathbf{X}$. The element-wise product is also called the Hadamard product.

With the common use of solar panels and other modern power generators, power can flow in each direction in a modern power distribution system. We still represent these systems as (directed) networks. The arrow attributes $\mathbf{S}_{mn}$ will be positive when flowing along the direction of the network, and negative when flowing in the opposite direction.

Definition 3.1.3. *The Hadamard product of two matrices $A = (a_{i,j})_{i,j}$ and $B = (b_{i,j})_{i,j}$ is*

$$A \circ B \coloneqq (a_{i,j} \cdot b_{i,j})_{i,j} \tag{3.1.4}$$

The Hadamard product has the following properties.

Proposition 3.1.4. *The Hadamard product is commutative, associative and distributive over addition. That is, if A, B and C are matrices of the same size, and k is a scalar then*

i) $A \circ B = B \circ A$,

ii) $A \circ (B \circ C) = (A \circ B) \circ C$,

iii) $A \circ (B + C) = A \circ B + A \circ C$,

iv) $(kA) \circ B = A \circ (kB) = k(A \circ B)$,

v) $A \circ 0 = 0$.

Moreover, if A and B are vectors we have

$$(A \circ B)_e = A_e \cdot B_e.$$

Proof. Follows directly from the fact that multiplication of scalars is commutative, associative and distributive over addition. □

Lastly we find the nodal powers $\mathbf{s}_m$ by

$$\mathbf{s}_m = \sum_{l\colon (l,m)\in\mathcal{A}} \mathbf{S}_{lm} - \sum_{n\colon (m,n)\in\mathcal{A}} \mathbf{S}_{mn} + \mathbf{L}_{mn}. \tag{3.1.5}$$

That is to say the nodal powers are the difference between all incoming powers, and all outgoing powers with some error $\mathbf{L}_{mn}$. The term $\mathbf{L}_{mn} \in \mathbb{C}^{3\times 1}$ is a nonlinear and non-convex loss term. We assume that these losses are negligible compared to the line flow, that is to say $|\mathbf{L}_{mn}| \ll |\mathbf{S}_{mn}|$ for all arrows, thus linearizing equation (3.1.5). We now have a method for finding all nodal powers from nodal voltages, given a set of line impedances. There exist matrices A and B such that the entire system can be reformulated to solving

$$\mathbf{ZI} = A^t\mathbf{V}, \tag{3.1.6}$$

$$\mathbf{S} = (B^t\mathbf{V}) \circ \mathbf{I}^*, \tag{3.1.7}$$

$$\mathbf{s} = -A\mathbf{S}, \tag{3.1.8}$$

for all three phases. These equations can be solved if the system is overdetermined, meaning there are more variables $\mathbf{V}_n$ known then there are equations to solve in the linear system (3.1.1).

When we define A and B we can show that solving (3.1.1), (3.1.3) and (3.1.16) for all m, n is equivalent to solving (3.1.6), (3.1.7) and (3.1.8) for all three phases individually.

We want the topology of the network to be represented beforehand. Representing the relationship between arrows and nodes in a matrix is needed to rewrite the power flow equations. First we need a notion of when an arrow and node are incident.

Definition 3.1.5. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $e = (m, n) \in \mathcal{A}$ an arrow, we call node m the source of e and n the sink of e. Additionally, we call m is the parent of n and n is the child of m.*

Definition 3.1.6. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $e \in \mathcal{A}$ and $m \in \mathcal{V}$ we call m is incident with e if m is either the source or the sink of e.*

Note that any arrow always has exactly two nodes incident to it, the source and the sink. Naturally there is also a notion of incidence when a graph is not directed, but since we only consider directed graph, we use this definition.

Definition 3.1.7. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ we call the binary matrix $A \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{A}|}$ defined by*

$$A_{me} = \begin{cases} 1 & \text{if } m \text{ is the source of } e \\ -1 & \text{if } m \text{ is the sink of } e \\ 0 & \text{else.} \end{cases} \tag{3.1.9}$$

the full incidence matrix of $\mathcal{G}$.

The graph in Figure 3.1 has the following full incidence matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{3.1.10}$$

when we index the arrows by $e_1 = (1,2), e_2 = (2,3), e_3 = (2,4)$ and $e_4 = (3,5)$. Labeling of the rows and columns can be changed, resulting in a different incident matrices. We see that the rows of A indicate the nodes, and show what lines are incident to that node. We can extract the degree of any node m by counting all non-zero elements of row m of A.

Definition 3.1.8. *For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ the degree of $n \in \mathcal{V}$ equals*

$$\#\{e \in \mathcal{A} \mid m \text{ is incident with } e\},$$

the number of edges that are incident to e.

Likewise we can observe all leaves of the graph. A leaf is a node with no outgoing arrows. Such a node only has one non-zero entry in its row, and that entry is -1. The transpose of A would be

$$A^t = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix},$$

switching the roles of the rows and columns. The full incidence matrix relates nodes to arrows, similar to what is happening in Ohm's law (Equation (3.1.1)).

Definition 3.1.9. *For any matrix $A \in \mathbb{C}^{n\times m}$ the $n-th$ row of A is indicated by $A_{(n,\cdot)} \in \mathbb{C}^m$ and the $m-th$ column is indicated by $A_{(\cdot,m)} \in \mathbb{C}^n$.*

Lemma 3.1.10. *For any matrix $A \in \mathbb{C}^{n\times m}$ with transpose A^t, and vectors $x \in \mathbb{C}^m, y \in \mathbb{C}^n$, we have*

$$(Ax)_k = \langle A_{(k,\cdot)}, x\rangle \tag{3.1.11}$$

and

$$(A^t y)_l = \langle A_{(\cdot,l)}, x\rangle. \tag{3.1.12}$$

Proof. By definition of matrix operations we have

$$(Ax)_k = \sum_i A_{k,i}x_i,$$

which is the inner product of $A_{(k,\cdot)} \coloneqq (A_{k,i})_i$ and $x \coloneqq (x_i)_i$. Using an analogue argument and noting that $A^t_{l,j} = A_{j,l}$ yields (3.1.12). □

The full incidence matrix can be used to calculate (3.1.1) for all nodes simultaneously, which saves computation time. Each time (3.1.1) is calculated, A can be known a priori, saving computation time if Ohm's law needs to be calculated many times.

The set of $|\mathcal{A}|$ equations in (3.1.3) can also be changed to an equivalent matrix form by using a variation of the incidence matrix, called the partial incidence matrix.

Definition 3.1.11. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ we call the binary matrix $A \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{A}|}$ defined by*

$$B_{em} = \begin{cases} 1 & \text{if } m \text{ is the sink of } e \\ 0 & \text{else.} \end{cases} \tag{3.1.13}$$

the partial incidence matrix of $\mathcal{G}$.

The partial incidence matrix is very similar to the full incidence matrix of Definition 3.1.7. The partial version only accounts for the sinks of each node, and does not consider where each arrow originates. It can be inferred from the full incidence matrix A by removing all entries equal to 1 and replacing all entries that equal -1 in A to 1. The partial incidence matrix can be used to solve all equations in (3.1.3), as these equations only need the information of where each sink node is.

We can now prove the equivalence that was presented earlier.

Proposition 3.1.12. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ let each node k inherit a voltage property $\mathbf{V}_k$ and each arrow $e = (m, n)$ an impedance $\mathbf{Z}_{mn}$ and a current $\mathbf{I}_{mn}$. Then solving the system of equations*

$$\mathbf{Z}_{mn}\mathbf{I}_{mn} = \mathbf{V}_m - \mathbf{V}_n \quad \forall (m, n) \in \mathcal{A}, \tag{3.1.14}$$

for a single phase is the same as solving the matrix equation

$$\mathbf{ZI} = A^t\mathbf{V}, \tag{3.1.15}$$

where A is the full incidence matrix, $\mathbf{Z} = diag(\mathbf{Z}^{\alpha\alpha}_{mn})$ is the diagonal matrix of all impedance, $\mathbf{I} = (\mathbf{I}^{\alpha}_{mn})_{(m,n)\in\mathcal{A}}$ is the vector of all currents and $\mathbf{V} = (\mathbf{V}^{\alpha}_{k})_{k\in\mathcal{V}}$ is the vector of all voltages of phase $\alpha \in \{a, b, c\}$.

Proof. Let $e = (m, n) \in \mathcal{A}$. Then $(\mathbf{ZI})_e = \langle \mathbf{Z}_{(e,\cdot)}, \mathbf{I}\rangle$ by Lemma 3.1.10. The vector $\mathbf{Z}_{(e,\cdot)}$ is zero everywhere except on the e-th element, where it equals $\mathbf{Z}_{mn}$. Thus $\langle \mathbf{Z}_{(e,\cdot)}, \mathbf{I}\rangle = \mathbf{Z}_{mn}\mathbf{I}_{mn}$. Likewise we find $(A^t\mathbf{V})_e = \langle A^t_{(\cdot,e)}, \mathbf{V}\rangle$ with $A_{(\cdot,e)}$ the $e-$th column of A. In Definition 3.1.7 we see that this equals the vector that is zero everywhere except at the source m and sink n, where it equals 1 and -1 respectively. As such, we find $\langle A^t_{(\cdot,e)}, \mathbf{V}\rangle = \mathbf{V}_m - \mathbf{V}_n$. This shows that each row of the matrix equation (3.1.15) represents exactly one of the equations in (3.1.14). □

Proposition 3.1.13. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ let each node k inherit a voltage property $\mathbf{V}_k$ and each arrow $e = (m, n)$ inherit an impedance $\mathbf{Z}_{mn}$, a current $\mathbf{I}_{mn}$ and a line power $\mathbf{S}_{mn}$. Then solving the system of equations*

$$\mathbf{S}_{mn} = \mathbf{V}_n \circ \mathbf{I}^*_{mn} \quad \forall (m, n) \in \mathcal{A}, \tag{3.1.16}$$

for a single phase is the same as solving the matrix equation

$$\mathbf{S} = (B^t\mathbf{V}) \circ \mathbf{I}^*, \tag{3.1.17}$$

where B is the partial incidence matrix, $\mathbf{S} = (\mathbf{S}_{mn})_{(m,n)\in\mathcal{A}}$, $\mathbf{I} = (\mathbf{I}_{mn})_{(m,n)\in\mathcal{A}}$ and $\mathbf{V} = (\mathbf{V}_k)_{k\in\mathcal{V}}$ are the vectors of all line powers, currents and voltages respectively.

Proof. Let $e = (m, n) \in \mathcal{A}$. The left-hand side of Equation (3.1.17) directly equals the left hand side of (3.1.16). By Lemma 3.1.10 and Lemma 3.1.4 we find

$$\begin{aligned} \left((B^t\mathbf{V}) \circ \mathbf{I}^*\right)_e &= (B^t\mathbf{V})_e \cdot \mathbf{I}^*_e \\ &= \langle B_{(\cdot,e)}, \mathbf{V}\rangle \cdot \mathbf{I}^*_{mn} \\ &= \mathbf{V}_m \cdot \mathbf{I}^*_{mn}. \end{aligned}$$

By definition of the partial incidence matrix, we have $\langle B_{(\cdot,e)}, \mathbf{V}\rangle = \mathbf{V}_m$.

A single phase of

$$\mathbf{S}_{mn} = \mathbf{V}_n \circ \mathbf{I}^*_{mn} \quad \forall (m,n) \in \mathcal{A}$$

is also of the form $\mathbf{V}_m \cdot \mathbf{I}^*_{mn}$ for given $e = (m,n) \in \mathcal{A}$.

□

Proposition 3.1.14. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ let each node k inherit a line power $\mathbf{s_k}$ and let each arrow $e = (m,n)$ inherit a line power $\mathbf{S}_{mn}$. Then solving the system of equations*

$$\mathbf{s}_m = \sum_{l\colon (l,m)\in\mathcal{A}} \mathbf{S}_{lm} - \sum_{n\colon (m,n)\in\mathcal{A}} \mathbf{S}_{mn} \quad \forall m \in \mathcal{V}, \tag{3.1.18}$$

for a single phase is the same as solving the matrix equation

$$\mathbf{s} = -A\mathbf{S}. \tag{3.1.19}$$

where A is the full incidence matrix and $\mathbf{S} = (\mathbf{S}_{mn})_{(m,n)\in\mathcal{A}}$, and $\mathbf{s} = (\mathbf{s_k})_{k\in\mathcal{V}}$ are the vectors of all line powers and nodal powers respectively.

Proof. Let $m \in \mathcal{V}$. Then $(-A\mathbf{S})_m = -\langle A_{(m,\cdot)}, \mathbf{S}\rangle$ by Lemma 3.1.10. We know that for $e \in \mathcal{A}$ a column $A_{(\cdot,e)}$ of A always has exactly one entry equal to 1 (the source) and one to -1 (the sink). That means, that for a given node m, the e-th entry of row $A_{(m,\cdot)}$ equals 1 if m is the source of that arrow, -1 if m is the sink, or 0 if neither. So we have

$$-\langle A_{(m,\cdot)}, \mathbf{S}\rangle = -\left(\sum_{n\colon (m,n)\in\mathcal{A}} \mathbf{S}_{mn} - \sum_{l\colon (l,m)\in\mathcal{A}} \mathbf{S}_{lm}\right).$$

Unfolding the minus sign gives us the right hand side of (3.1.18) for any given m. □

3.2 Previous work

The optimal sensor placement in a power distribution grid has been discussed in [2], [3], [4], [5]. They all focus on turning an undetermined system into an overdetermined system, though many make use of a expensive measuring unit, a Phasor Measurement Unit (PMU), which can measure both the voltage magnitude and the voltage phase angle. In [2] the focus lies on an optimal sensor placement solution that enables outage detection through a statistical test based on sensor measurements. A power flow-based approach to infer the unknown power injections at non-metered grid nodes from metered grid nodes is proposed in [4]. They provide a mathematically sound condition that guarantees solvability of a relaxation of the power flow problem. It states that if there exists a set of node-disjoint paths connecting every unmeasured node to a fully measured node, i.e. a node that uses a PMU, then the relaxed problem is locally solvable everywhere. In [3] a concrete method is given for estimating the unknown voltages for given for a three-phase distribution network. By estimating the voltages and using those estimations as input the system becomes overdetermined and thus solvable.

Comparing measuring errors to estimation errors that arise from this method differ an order of magnitude in favor of measuring, meaning measuring nodes has more advantages than estimating.

The advantage of estimating is that it does not require measuring equipment and is hence less expensive. We can use the estimator considered in [3] as a baseline on which we can improve by adding sensors. This is the estimator currently used by Alliander.

The advantages of the methods used in the other papers are the following. We assume that all voltages have the same phase angle. Considering we can set the substation as the standard, we may assume all voltage angles are zero. Consequently, equipment that only measures the voltage magnitude is in our case equivalent to a PMU, while being considerably cheaper (a PMU can cost up to 15.000 euro). As such, we can use any study that considers PMUs in their research.

According to [1], when we compare investment costs to achieved data accuracy the following should be considered: the complexity of modern grids is such that real concerns are arising about the limits of observability. Unstructured additions to power distribution grids can cause a degree of architectural chaos that makes the determination of grid state a challenge. This lack of structure, combined with severe complexity may place limits on achievable observability.

Power grids use a wide array of sensing devices. A key tradeoff for sensor network design has been the use of many low costs sensors versus the use of a smaller number of high-end sensors. As modern distribution grid complexity rises, a move toward high-end sensors might be required. Despite this possible trend, it is still valid to consider the use of low-end sensors and in the more sophisticated approaches, to employ a mix of sensor types. In our case, we only focus on low-end sensors, as we assume voltage angles to be zero which implies high-end sensors are not necessary. Research has been done to optimize the sensor placement through a greedy algorithm when PMUs are the only sensor type used (see [12] and [13]). Researchers have also proposed hybrid state estimation schemes (see [7]), integrating both PMU and SCADA data. Some of these methods incorporate the PMU measurements into the iterative state estimation updates (see [8]-[9]), while others use the PMU data to refine the estimates obtained from SCADA data (see [10] and [11]). It has also been proposed to use a different greedy algorithm to account for the use of different sensors ([5]), though this increases the calculation time to $o(p^5)$, where p is number of measurements. A metric has been introduced in [6] depending on convergence, observability and performance to evaluate the estimation performance and numeric stability of a state estimation, which can be used to optimized PMU placement. We shall go into paper [4], which shows the amount and locations of sensors needed to solve a relaxed power flow problem.

In [2] they consider a more general and flexible optimization of a weighted sum of the errors of all possible electrical quantities in the network. They say that such flexibility is important, in particular in distribution grids where the budget is limited and only a few measurement points may be chosen with respect to the size of the network. In [2] they propose a greedy solution for the case of measuring equipment with non-uniform cost, based on some general results on submodular functions.

In this thesis we can ignore the need for non-uniform measuring equipment by assuming there is no change in voltage angles, making a high-end sensor equivalent to its cheaper counterparts. No change in voltage angles means that we can assume all $\mathbf{V}_m$ to be real numbers.

3.3 Scoring state estimation predictions

In this section we introduce a notion of prediction for a true state $\mathbf{V}$ and calculate the difference between the corresponding true nodal powers and predicted powers. At the end of this section we introduce an upper bound for the $2-$norm of this difference, which will be the main contribution of this thesis.

As in Section 3.1 we can calculate the nodal powers from the voltages if all resistances are known a priori. In this section we do this for a single-phase system. A conversion to a three-phase system follows directly as we assume all impedances to be diagonal matrices, which assures each phase acts independently (see Equation (3.1.2)). We represent the topological aspects of the power distribution network by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $\mathcal{V}$ its vertices and $\mathcal{A}$ its arrows. This graph is a tree with source node 1, the substation. We start of by formalizing the notion of a prediction and a measurements. This is done for a single phase, and can be generalized by repeating the process for all three phases.

Definition 3.3.1. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = N$ and voltages $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_N)$ we call $\mathbf{V}' = (\mathbf{V'}_1, \dots, \mathbf{V'}_N)$ a prediction of $\mathbf{V}$ if for each $n \in \mathcal{V}$ we have $\mathbf{V'}_n = \mathbf{V}_n + \mathbf{X}_n$ for some $\mathbf{X}_n \in \mathbb{C}$. We often refer to $\mathbf{V}$ as the true state and $\mathbf{V}'$ as the the predicted state.*

In most cases we will only consider real valued predictions, but it is possible to consider complex predictions as well. When we compute actual case studies to support our claims the values $\mathbf{X}_n$ are taken from some distribution. This distribution represents the forecaster and how it would differ from reality. In this thesis we consider some common distributions that mostly behave as a normal distribution. Not all forecasters will behave by such a proper distribution. For example, a machine learning algorithm might not behave in a manner that can be simulated by a random variable. Which forecasters can be represented by a random variable and which cannot is not within the scope of this thesis.

Definition 3.3.2. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = N$, we call node $n \in \mathcal{V}$ is measured if for any voltage vector $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_N)$ and any prediction $\mathbf{V}' = (\mathbf{V'}_1, \dots, \mathbf{V'}_N)$ we have $\mathbf{V}_n = \mathbf{V'}_n$. Equivalently this occurs when $\mathbf{X}_n = 0$.*

An actual measurement of voltages is not this exact. We should assume that in the real world there are measuring errors. However, these are generally of a different magnitude then prediction errors. Prediction errors are far greater than measurement errors and hence can be considered zero in this context.

Remark. For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = N$, true state $\mathbf{V}$ and predicted state $\mathbf{V}'$ we have

$$\mathbf{V}' = \mathbf{V} + \sum_{n \in \mathcal{V}} \delta_n \mathbf{X}_n,$$

where $\delta_{\mathbf{n}} \in \mathbb{R}^N$ is the Kronecker delta

$$\delta_n(k) = \begin{cases} 1 & \text{if } n = k \\ 0 & \text{if } n \neq k. \end{cases}$$

For a pair $(\mathbf{V}, \mathbf{V}')$ we can calculate the corresponding true and predicted nodal powers $(\mathbf{s}, \mathbf{s}')$ by the equations discussed in Section 3.1. Consider the following function.

Definition 3.3.3. *the $2-$norm score of true state $(\mathbf{V}, \mathbf{s})$ and predicted state $(\mathbf{V}', \mathbf{s}')$ is defined as*

$$\|\mathbf{s}' - \mathbf{s}\|_2 \coloneqq \sqrt{\sum_k |\mathbf{s}_k - \mathbf{s}'_k|^2}$$

We can see this as a scoring function for the true and predicted states $(\mathbf{V}, \mathbf{s})$ and $(\mathbf{V}', \mathbf{s}')$ if we see the predicted state as the (single valued) probability distribution and the true state as the event. Indeed, the norm outputs a (non-negative) real number and hence we can consider it as a scoring function. In this section we give an upper bound for this score, which shows us some patterns as to which nodes score best if we consider the predicted state when a given node is measured.

Let us recall that for m, n nodes in $\mathcal{G}$ we have

$$\mathbf{V}_m - \mathbf{V}_n = \mathbf{Z}_{mn}\mathbf{I}_{mn}, \tag{3.3.1}$$

and

$$\mathbf{S}_{mn} = \mathbf{V}_n \circ \mathbf{I}^*_{mn}, \tag{3.3.2}$$

and lastly

$$\mathbf{s}_m = \sum_{l\colon (l,m)\in\mathcal{A}} \mathbf{S}_{lm} - \sum_{n\colon (m,n)\in\mathcal{A}} \mathbf{S}_{mn}. \tag{3.3.3}$$

In Section 3.1 we showed that solving these equations from $\mathbf{V}$ to $\mathbf{s}$ is equivalent to solving

$$\begin{aligned} \mathbf{ZI} &= A^t\mathbf{V}, \\ \mathbf{S} &= (B^t\mathbf{V}) \circ \mathbf{I}^*, \\ \mathbf{s} &= -A\mathbf{S}, \end{aligned}$$

for all three phases, where $\mathbf{Z}$ is the diagonal matrix with all $\mathbf{Z}_{mn}$ on it's diagonal, A is the full incidence matrix, and B is the partial incidence matrix as in Definition 3.1.7 and 3.1.11 respectively.

Now we try to write $\mathbf{s}$ as a function of $\mathbf{s}'$. First we need some Propositions.

Proposition 3.3.4. *The Hadamard product of two vectors $x, y \in \mathbb{R}^N$ is the same as matrix multiplication of one vector by the corresponding diagonal matrix of the other vector:*

$$x \circ y = \mathbf{D}_x y,$$

where by corresponding diagonal matrix $\mathbf{D}_x$ of x we mean a matrix with vector x on its diagonal.

Proof. Calculating $\mathbf{D}_x y$ directly yields the result. □

Corollary 3.3.5. *For vectors x, y and diagonal matrix $\mathbf{Z}$ we have*

$$x \circ \mathbf{Z}y = \mathbf{Z}(x \circ y).$$

Proof. by Proposition **??** we find

$$x \circ \mathbf{Z}(y) = \mathbf{D}_x\mathbf{Z}(y),$$

where $\mathbf{D}_x$ and $\mathbf{Z}$ commute as they are both diagonal matrices. Using Proposition 3.3.4 once more we find

$$\mathbf{Z}\mathbf{D}_x(y) = \mathbf{Z}(x \circ y).$$

□

Proposition 3.3.6. *Let $*$ denote the complex conjugate and t denote the transpose of a matrix. For any real matrix A and complex vector $\mathbf{z}$ we have*

$$(A\mathbf{z})^* = A^t\mathbf{z}^*.$$

Proof. Let $\mathbf{z}_1 = \mathbf{x}_1 + j\mathbf{y}_1$ and $\mathbf{z}_2 = \mathbf{x}_2 + j\mathbf{y}_2$ with x_i, y_i real vectors. Then

$$\begin{aligned}(\mathbf{z}_1 + \mathbf{z}_2)^* =& (\mathbf{x}_1 + \mathbf{x}_2 + j(\mathbf{y}_1 + \mathbf{y}_2))^* \\ =& \mathbf{x}_1 + \mathbf{x}_2 - j(\mathbf{y}_1 - \mathbf{y}_2) \\ =& \mathbf{z}_1^* + \mathbf{z}_2^*.\end{aligned}$$

Furthermore, any matrix is linear, thus for $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ with $\mathbf{x}, \mathbf{y}$ real vectors we find

$$\begin{aligned}(A\mathbf{z})^* &= (A\mathbf{x} + jA\mathbf{y})^* \\ &= (A\mathbf{x})^* + (jA\mathbf{y})^* \\ &= (\overline{A})^t\mathbf{x} - j(\overline{A})^t\mathbf{y} \\ &= A^t(\mathbf{x} - j\mathbf{y}) \\ &= A^t\mathbf{z}^*,\end{aligned}$$

as $\overline{A} = A$ for any real matrix A. □

From these propositions we can calculate $\mathbf{s}'$ from $\mathbf{V}'$, where $\mathbf{V}'$ is given by $\mathbf{V}' = \mathbf{V} + \sum_{k=1}^{|\mathcal{V}|} \delta_{\mathbf{k}}\mathbf{X}_k$. Firstly, by Equation (3.1.6) the predicted current $\mathbf{I}'$ is

$$\begin{aligned}\mathbf{I}' &= \mathbf{Z}^{-1}A^t\mathbf{V}', \\ &= \mathbf{Z}^{-1}A^t(\mathbf{V} + \sum_{k=1}^{n} \delta_{\mathbf{k}}\mathbf{X}_k), \\ &= \mathbf{Z}^{-1}A^t\mathbf{V} + \sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_k, \\ &= \mathbf{I} + \sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_k,\end{aligned}$$

by invertibility of $\mathbf{Z}$ and linearity of A and $\mathbf{Z}$. Consequently we find from Equation (3.1.7) that

$$\begin{aligned}\mathbf{S}' &= (B^t\mathbf{V}') \circ \mathbf{I}'^*, \\ &= \left(B^t(\mathbf{V} + \sum_{k=1}^{n} \delta_{\mathbf{k}}\mathbf{X}_k)\right) \circ (\mathbf{I} + \sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_k)^*, \\ &= (B^t\mathbf{V}) \circ \mathbf{I}^* + \left(\sum_{k=1}^{n} B^t\delta_{\mathbf{k}}\mathbf{X}_k\right) \circ \mathbf{I}^* + \left(B^t\mathbf{V}\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_k\right)^*, \\ &\quad + \left(\sum_{k=1}^{n} B^t\delta_{\mathbf{k}}\mathbf{X}_k\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_{\mathbf{k}}\right)^*, \\ &= \mathbf{S} + \left(\sum_{k=1}^{n} B^t\delta_{\mathbf{k}}\mathbf{X}_k\right) \circ \mathbf{I}^* + \left(B^t\mathbf{V}\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_k\right)^*, \\ &\quad + \left(\sum_{k=1}^{n} B^t\delta_{\mathbf{k}}\mathbf{X}_k\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1}A^t\delta_{\mathbf{k}}\mathbf{X}_{\mathbf{k}}\right)^*,\end{aligned}$$

by the distributive property of the Hadamard product (Proposition 3.1.4) and linearity of the complex conjugate. This can be further simplify using Proposition 3.1.4, Corollary 3.3.5 and Proposition 3.3.6 by noting that $\mathbf{X}_k$ are scalars, $\mathbf{Z}^{-1}$ is a diagonal matrix (as $\mathbf{Z}$ is one) and that matrix B, vectors $\delta_{\mathbf{k}}$ and vector $\mathbf{V}$ are real elements. It is only $\mathbf{Z}$ and the scalars X_k that carry any complex components. With that we find that the remaining parts of $\mathbf{S}'$ are firstly

$$\left(\sum_{k=1}^{n} B^t \delta_{\mathbf{k}} \mathbf{X}_k\right) \circ \mathbf{I}^* = \left(\sum_{k=1}^{n} B^t \delta_k \mathbf{X}_k\right) \circ (\mathbf{Z}^{-1} A^t \mathbf{V})^* \tag{3.3.4}$$

$$= (\mathbf{Z}^{-1})^* \sum_{k=1}^{n} \mathbf{X}_k \left((B^t \delta_k) \circ (A^t \mathbf{V})\right), \tag{3.3.5}$$

secondly

$$\left(B^t \mathbf{V}\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1} A^t \delta_{\mathbf{k}} \mathbf{X}_k\right)^* = (\mathbf{Z}^{-1})^* \sum_{k=1}^{n} \mathbf{X}_k^* \left((B^t \mathbf{V}) \circ (A^t \delta_{\mathbf{k}})\right), \tag{3.3.6}$$

and thirdly

$$\left(\sum_{k=1}^{n} B^t \delta_{\mathbf{k}} \mathbf{X}_k\right) \circ \left(\sum_{k=1}^{n} \mathbf{Z}^{-1} A^t \delta_{\mathbf{k}} \mathbf{X}_{\mathbf{k}}\right)^* = (\mathbf{Z}^{-1})^* \sum_{k=1}^{n} \sum_{m=1}^{n} \mathbf{X}_k \mathbf{X}_m^* \left((B^t \delta_{\mathbf{k}}) \circ (A^t \delta_{\mathbf{m}})\right). \tag{3.3.7}$$

For now we call these three components R_1, R_2 and R_3, where R stands for remainder. Before we expand on R_1, R_2 and R_3 we finish the calculation from nodal voltages to nodal powers. From Equation (3.1.8) we calculate $\mathbf{s}'$ by

$$\begin{aligned} \mathbf{s}' &= -A\mathbf{S}' \\ &= -A(\mathbf{S} + R_1 + R_2 + R_3) \\ &= \mathbf{s} - A(R_1 + R_2 + R_3). \end{aligned}$$

And so the difference between the true and predictive nodal powers is bounded by

$$\|\mathbf{s}' - \mathbf{s}\| = \|A(R_1 + R_2 + R_3)\| \tag{3.3.8}$$

$$\leq \|A\| \cdot \|R_1 + R_2 + R_3\| \tag{3.3.9}$$

$$\leq \|A\| \cdot (\|R_1\| + \|R_2\| + \|R_3\|), \tag{3.3.10}$$

where $\|A\|$ is the operator norm for A. We consider the Euclidean norm, also called the 2-norm.

Definition 3.3.7. *the 2-norm of a matrix is*

$$\|A\|_2 = \sup\{\|Ax\|_2 \colon \|x\| \leq 1\}, \tag{3.3.11}$$

where $\|y\|_2 = \left(\sum_{k=1}^{n} |y_k|^2\right)^{\frac{1}{2}}$ *is the 2-norm of the vector* $y \in \mathbb{C}^n$ *and* $|y_k|$ *is the euclidean complex norm.*

A direct upper bound for $\|A\|_2$ with A the full incident matrix can be found in Proposition 3.3.12. Inequality (3.3.9) arises from the following lemma and is called the matrix norm inequality.

Lemma 3.3.8. *For any matrix A, vector x in the domain of A and norm $\|\cdot\|$ we have*

$$\|Ax\| \leq \|A\| \cdot \|x\|. \tag{3.3.12}$$

Proof. When $\|x\| = 1$ then $\|Ax\| \leq \|A\|$ by definition of the matrix norm. For any x in the domain of A we have

$$\left\| \frac{x}{\|x\|} \right\| = 1,$$

thus by definition of $\|A\|_2$

$$\left\| A\frac{x}{\|x\|} \right\| \leq \|A\|. \tag{3.3.13}$$

The element $\frac{1}{\|x\|}$ is a scalar and A is a linear operator, so we find

$$\left\| A\frac{x}{\|x\|} \right\| = \frac{1}{\|x\|} \|Ax\| .$$

Combine this with (3.3.13) and multiply both sides with $\|x\|$ to find

$$\|Ax\| \leq \|A\| \cdot \|x\|.$$

□

Some quick simulations show that $\|A\|_2$ generally does not exceed 3 for small graphs. We calculated the mean to be 2.3 over a random set of 100 networks having around 10 to 100 nodes and 3 routes. We can also provide a mathematical upper bound, which is dependent on the amount of routes.

Definition 3.3.9. *The 1-norm of a vector x is*

$$\|x\|_1 = \sum_i |x_i|,$$

and its $\infty-$norm is

$$\|x\|_\infty = \max_i |x_i|$$

Lemma 3.3.10. *The matrix norm as in Definition 3.3.11 for the $1-$norm can be rewritten as*

$$\|W\|_1 = \max_j \sum_i |w_{ij}|,$$

and the $\infty-$matrix norm can be rewritten as

$$\|W\|_\infty = \max_i \sum_j |w_{ij}|.$$

That is to say, the 1-norm is the maximum absolute column sum and the $\infty-$norm is the maximum absolute row sum.

Lemma 3.3.11. *For any matrix W we have*

$$\|W\|_2^2 \leq \|W\|_1 \|W\|_\infty.$$

Proof. Note that if this inequality holds for any vector x, it must also hold for the operator norms. Let x be a vector. Then

$$\|x\|_2^2 = \sum_k |x_k|^2 \tag{3.3.14}$$

$$\leq \sum_k |x_k| \cdot \max_j |x_j| \tag{3.3.15}$$

$$= \|x\|_1 \cdot \|x\|_\infty. \tag{3.3.16}$$

□

Proposition 3.3.12. *For the full incidence matrix A of a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $D = \max_{k\in\mathcal{V}} \deg(k)$ we have*

$$\|A\|_2 \leq \sqrt{2D}.$$

Proof. We show that $\|A\|_1 = 2$ and $\|A\|_\infty = D$. Then by Lemma 3.3.11 we are done. The maximum absolute column sum of A is always two, as each column attains 1 exactly once, -1 exactly once and is zero everywhere else. The maximum absolute row sum equals the largest degree D. By Lemma 3.3.10 we thus have $\|A\|_1 = 2$ and $\|A\|_\infty = D$. □

The number D is often the number of routes in the network, as the maximum degree is almost always attained at the substation. The amount of routes is dependent on $|\mathcal{V}|$ and larger networks often have more routes, but this amount is often only a small fraction of the amount of nodes.

The 2−norm can be considered a scoring function, as proposed by Definition 3.3.3.

With Equations (3.3.8), (3.3.9) and (3.3.10) a direct upper bound for the 2−norm score is found, and it gives rise to a new scoring function, one that is independent of the predicted states and impedances. We introduce this upper bound as a theorem and prove this theorem in the next section.

Theorem 3.3.13. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with full incidence matrix A and partial incidence matrix B be a network with impedances $\mathbf{Z} = diag(\mathbf{Z}_{mn})$, true state $(\mathbf{V}, \mathbf{s})$ and predicted state $(\mathbf{V}', \mathbf{s}')$ such that $\mathbf{V}' = \mathbf{V} + \sum_{k=1}^{|\mathcal{V}|} \delta_k \mathbf{X}_k$ and let $\lambda = \min |\mathbf{Z}_{mn}|$. Then*

$$\|\mathbf{s}-\mathbf{s}'\|_2 \leq \frac{\|A\|_2}{\lambda} \left(\sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \left(|\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2} \right) + \sum_{k=1}^{|\mathcal{V}|} \left(|\mathbf{X}_k|^2 + |\mathbf{X}_k| \cdot |\mathbf{X}_{\mathbf{n_k}}| \right) \right), \tag{3.3.17}$$

where e_k is the unique arrow with node $k \in \mathcal{V}$ as its sink and n_k the parent of k.

The right-hand side of (3.3.17) is dependent on the term

$$|\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2}, \tag{3.3.18}$$

which in turn is dependent on A, B and $\mathbf{V}$. The right-hand side of Inequality (3.3.17) is therefore called the Incident Voltage bound and the term given in (3.3.18) is called the Incident Voltage score.

Recall that we consider a node $k \in \mathcal{V}$ measured if $\mathbf{X}_k = 0$. The Incident Voltage bound is lowered when nodes are measured, because the Incident Voltage score of that node is then multiplied with

zero. Choosing $\mathbf{X}_k = 0$ where (3.3.18) is most often large (given a set of common $\mathbf{V}$) might thus be a valid strategy for choosing which nodes to measure in a power distribution grid.

This bound is the main focus and contribution of this thesis. It shall be proven in Chapter 4, where we also show some of its properties and explore the accuracy of the bound. Then, in Chapter 5, we discuss some case studies where we pick different assumptions such as the size and shape of the network, and different distributions that simulate the predictor.

Chapter 4

Incident Voltage score and Incident Voltage bound

In Section 4.1 we proof that the Incident Voltage bound is a bound for the 2−norm. Once this is proven we formally introduce the Incident Voltage score, which can be used to score nodes in a network to show which of them should be measured. In Section 4.2 we discuss its properties. For example, the Incident Voltage score prefers nodes with a high degree. We also explore the variance for different $\mathbf{V}$ to determine if we can set up a profile of possible $\mathbf{V}$. Lastly, in Section 4.3 we discuss the loss of accuracy which is created by the estimation.

4.1 Proof of Incident Voltage bound

Recall R_1, R_2 and R_3 as used in equations (3.3.8), (3.3.9) and (3.3.10). To further calculate the norms of R_1, R_2 and R_3 we mainly need to know what

$$\|(B^t\delta_k) \circ (A^t\mathbf{V})\|_2 \tag{4.1.1}$$
$$\|(B^t\mathbf{V}) \circ (A^t\delta_k)\|_2 \tag{4.1.2}$$
$$\|(B^t\delta_k) \circ (A^t\delta_m)\|_2, \tag{4.1.3}$$

are for nodes k and m.

Lemma 4.1.1. *For any Kronecker delta δ_k we have $A^t\delta_k = A_{(k,\cdot)}$, the $k-$the row of A. This correspondence is bijective.*

Proof. Noting that $(A^t)^t = A$ and $(AB)^t = B^tA^t$ directly yields the results. Indeed, for any k we have $(A^t\delta_k)^t = \delta_k^t A$, which indicates the $k-th$ row of A through a bijective correspondence. □

Lemma 4.1.2. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with node $k \in \mathcal{V}$ and associated full incidence matrix A we have*

$$\|(A^t\delta_k) \circ y\|_2 = \sqrt{\sum_{j \in \mathcal{A}:\ j \ incident\ to\ k} |y_j|^2},$$

for any vector $y \in \mathbb{C}^{|\mathcal{A}|}$.

Proof. Let $k \in \mathcal{V}$, then $A^t\delta_k = A_{(k,\cdot)}$, the $k-$the row of A. By definition of the full incidence matrix this means that $A^t\delta_k$ is a vector which entries equal to

$$(A^t\delta_k)(j) = \begin{cases} 1 & \text{if } \exists\ n \in \mathcal{V}\colon j = (k,n) \\ -1 & \text{if } \exists\ n \in \mathcal{V}\colon j = (n,k) \\ 0 & \text{else.} \end{cases} \tag{4.1.4}$$

Now let $y \in \mathbb{C}^{|\mathcal{A}|}$, then

$$(A^t\delta_k \circ y)(j) = \begin{cases} y_j & \text{if } \exists\ n \in \mathcal{V}\colon j = (k,n) \\ -y_j & \text{if } \exists\ n \in \mathcal{V}\colon j = (n,k) \\ 0 & \text{else.} \end{cases} \tag{4.1.5}$$

The sought after equation then follows directly from the definition of the $2-$norm. □

A likewise result can be found for B^t.

Lemma 4.1.3. *For a network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with node $k \in \mathcal{V}$ and associated partial incidence matrix B we have*

$$\|(B^t\delta_k) \circ y\|_2 = |y_{e_k}|,$$

for any vector $y \in \mathbb{C}^{|\mathcal{A}|}$. Here e_k is the arrow with k as its sink.

Proof. Same as Lemma 4.1.2, except there is always exactly one source, so no summation is needed. □

Note that this is well-defined because in a tree there is always exactly one arrow that has a given node k as its sink, unless that node is the substation.

Lemma 4.1.4. *The norm of a diagonal matrix D equals its largest entry.*

Proof. Let $D = \text{diag}(\lambda_k)$. As $|\lambda_k| \geq 0$ for each k, we have

$$\begin{aligned} \|D\|_2 &= \max_{\|x\|_2=1} \|Dx\|_2 \\ &= \max_{\|x\|_2=1} \left(\sum_k \lambda_k^2 x_i^2 \right)^{\frac{1}{2}} \\ &\leq \max_{\|x\|_2=1} \max_k |\lambda_k| \left(\sum_k x_i^2 \right)^{\frac{1}{2}} \\ &= \max_k |\lambda_k| \max_{\|x\|_2=1} \|x\|_2 \\ &= \lambda. \end{aligned}$$

On the other hand, let x be an eigenvector of the largest $\lambda := \max_k \lambda_k$ with $\|x\|_2 = 1$. Then $Dx = \lambda x$ and so

$$\begin{aligned} \|D\|_2 &\geq \|Dx\|_2 \\ &= \|\lambda x\|_2 \\ &= |\lambda| \|x\|_2 \\ &= \max_k \lambda_k. \end{aligned}$$

□

Theorem 4.1.5. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with full incidence matrix A and partial incidence matrix B be a network with impedances $\mathbf{Z} = diag(\mathbf{Z}_{mn})$ true state $(\mathbf{V}, \mathbf{s})$ and predicted state $(\mathbf{V}', \mathbf{s}')$ such that $\mathbf{V}' = \mathbf{V} + \sum_{k=1}^{|\mathcal{V}|} \delta_k X_k$ and let R_1, R_2, R_3 be as in (3.3.5), (3.3.6) and (3.3.7),*

$$R_1 \coloneqq (\mathbf{Z}^{-1})^* \sum_{k=1}^{|\mathcal{V}|} \mathbf{X}_k \left((B^t \delta_k) \circ (A^t \mathbf{V}) \right), \tag{4.1.6}$$

$$R_2 \coloneqq (\mathbf{Z}^{-1})^* \sum_{k=1}^{|\mathcal{V}|} \mathbf{X}_k^* \left((B^t \mathbf{V}) \circ (A^t \delta_{\mathbf{k}}) \right) \tag{4.1.7}$$

and

$$R_3 \coloneqq (\mathbf{Z}^{-1})^* \sum_{k=1}^{|\mathcal{V}|} \sum_{m=1}^{|\mathcal{V}|} \mathbf{X}_k \mathbf{X}_m^* \left((B^t \delta_{\mathbf{k}}) \circ (A^t \delta_{\mathbf{m}}) \right). \tag{4.1.8}$$

Then with $\lambda \coloneqq \min_{(m,n \in \mathcal{A})} |\mathbf{Z}_{mn}|$ we have

$$\|R_1\|_2 \leq \frac{1}{\lambda} \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \cdot |A_{(\cdot, e_k)} \mathbf{V}|, \tag{4.1.9}$$

$$\|R_2\|_2 \leq \frac{1}{\lambda} \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \cdot \sqrt{\sum_{e \in \mathcal{A} \colon e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V} \rangle|^2}, \tag{4.1.10}$$

and

$$\|R_3\|_2 \leq \frac{1}{\lambda} \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k|^2 + |\mathbf{X}_k| \cdot |\mathbf{X}_{n_k}|, \tag{4.1.11}$$

where $e_k = (n_k, k) \in \mathcal{A}$ is the arrow with node $k \in \mathcal{V}$ as its sink and n_k is the parent of k, i.e. $(n_k, k) \in \mathcal{A}$.

Proof. The matrix $\mathbf{Z}$ is a diagonal matrix, and so $(\mathbf{Z}^{-1})^*$ is a diagonal matrix as well. Lemma 4.1.4 then says

$$\|(\mathbf{Z}^{-1})^*\| = \max_{(m,n) \in \mathcal{A}} |(\mathbf{Z}_{mn}^{-1})^*|.$$

Any norm is invariant under the transpose and in general we have

$$\max_k x_k^{-1} = (\min_k x_k)^{-1},$$

so

$$\begin{aligned} \|(\mathbf{Z}^{-1})^*\| &= \max_{(m,n) \in \mathcal{A}} |(\mathbf{Z}_{mn}^{-1})^*| \\ &= \left(\min_{(m,n) \in \mathcal{A}} |\mathbf{Z}_{mn}| \right)^{-1} \\ &= \frac{1}{\lambda}, \end{aligned}$$

where $\lambda \coloneqq \min_{(m,n\in\mathcal{A})} |\mathbf{Z}_{mn}|$.

With the triangle inequality and Lemma 3.3.8 we now have

$$\begin{aligned}\|R_1\|_2 &\leq \|(\mathbf{Z}^{-1})^*\|_2 \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \cdot \| \left((B^t\delta_k) \circ (A^t\mathbf{V})\right) \|_2 \\ &\leq \frac{1}{\lambda} \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \cdot \| \left((B^t\delta_k) \circ (A^t\mathbf{V})\right) \|_2.\end{aligned}$$

Next we show that for each $k \in \mathbf{V}$ we have

$$\| \left((B^t\delta_k) \circ (A^t\mathbf{V})\right) \|_2 = |\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle|,$$

where e_k is the unique arrow with node $k \in \mathbf{V}$ as its sink. Luckily for us, we already proved this in Lemma 4.1.3. We only need to recall Lemma 3.1.10, which states that the e_k-th entry of $A^t\mathbf{V}$ is the inner product of the e_k-th column of A and $\mathbf{V}$. As such

$$\|R_2\|_2 \leq \frac{1}{\lambda} \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \cdot \left((B^t\mathbf{V}) \circ (A^t\delta_\mathbf{k})\right),$$

and

$$\left((B^t\mathbf{V}) \circ (A^t\delta_\mathbf{k})\right) = \sqrt{\sum_{e\in\mathcal{A}\colon e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2}$$

by Lemma 4.1.2. It remains to show that

$$\sum_{k=1}^{|\mathcal{V}|} \sum_{m=1}^{|\mathcal{V}|} |\mathbf{X}_k\mathbf{X}_m^*| \cdot \| \left((B^t\delta_\mathbf{k}) \circ (A^t\delta_\mathbf{m})\right) \|_2 = \sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k|^2 + |\mathbf{X}_k| \cdot |\mathbf{X}_{n_k}|,$$

where n_k is the parent of k. By Lemma 4.1.3 we have

$$\| \left((B^t\delta_\mathbf{k}) \circ (A^t\delta_\mathbf{m})\right) \|_2 = |A_{(\cdot,e_k)}\delta_m|,$$

for any pair of nodes $k, m \in \mathcal{V}$. We can directly calculate this inner product. The vector $A_{(\cdot,e_k)}$ equals zero everywhere except the at node k and its parent n_k, where it equals -1 and 1 respectively. The vector δ_m is zero everywhere, except its $m-$th entry. And thus

$$\begin{aligned}|A_{(\cdot,e_k)}\delta_m| &= |\langle (0,\ldots,0,1,0,\ldots,0,-1,0\ldots,0), (0,\ldots,0,1,0,\ldots,0)\ \rangle| \\ &= \begin{cases} 1 & \text{if } m = n \text{ or } m = k \\ 0 & \text{else.}\end{cases}\end{aligned}$$

Thus, for a given $k \in \mathcal{V}$ we have

$$\sum_{m=1}^{|\mathcal{V}|} |\mathbf{X}_k\mathbf{X}_m^*| \cdot \| \left((B^t\delta_\mathbf{k}) \circ (A^t\delta_\mathbf{m})\right) \|_2 = |\mathbf{X}_k\mathbf{X}_k^*| + |\mathbf{X}_k\mathbf{X}_{n_k}^*|.$$

Noting that $|\mathbf{X}_k\mathbf{X}_k^*| = |\mathbf{X}_k|^2$ and $|\mathbf{X}_k\mathbf{X}_{n_k}^*| = |\mathbf{X}_k| \cdot |\mathbf{X}_{n_k}|$ finalizes the proof. □

With Equations (3.3.8), (3.3.9) and (3.3.10) we find an upper bound for $\|\mathbf{s}' - \mathbf{s}\|_2$. This indicates what nodes should be measured when trying to minimize $\|\mathbf{s} - \mathbf{s}'\|_2$. When these bounds lower, the bound on $\|\mathbf{s} - \mathbf{s}'\|_2$ lowers. Having $\mathbf{X}_m = 0$ is equivalent to having no prediction noise on node m, that is to say, measuring node m. This is very important, because it means that when trying to minimize (4.1.9), we want $\mathbf{X}_k = 0$ preferably there where $|\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle|$ is large. Likewise, from equations (4.1.10) we refer that it is wise to have $\mathbf{X}_k = 0$ when $\sqrt{\sum_{j\in\mathcal{A}:\ e \text{ incidence to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2}$ is large. Where the degree of k is higher than average, the sum in the squared root becomes bigger, meaning those X_k should take priority. Later we will give empirical evidence that having a high degree indeed constitutes a higher score. In general we have $|\langle B_{(\cdot,e)}, \mathbf{V}\rangle| \gg |A_{\cdot,e}\mathbf{V}|$, meaning the second property will be considered with a higher weight than the first.

Definition 4.1.6. *For* $\mathbf{s}$ *calculated by* $\mathbf{V}$ *through the state estimation equations as described in Section 3.1 we say the* $\mathcal{IV}$ *(Incident Voltage) score is the vector defined for each* $k = 1, \dots, |\mathcal{V}|$ *as*

$$\mathcal{IV}(\mathbf{s}, \mathbf{s}')_k \coloneqq |\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2},$$

where A *is the full incidence matrix,* B *is the partial incidence matrix and* e_k *is the arrow with node* k *as its sink.*

Note that this score does not score a single node, instead it produces a vector of length equal to the amount of nodes. A larger values means that it would be a better candidate for placing measuring equipment. Moreover, it is not dependent on $\mathbf{s}'$, which a typical scoring function is. As such, we will often write $\mathcal{IV}(\mathbf{s})$ instead of $\mathcal{IV}(\mathbf{s}, \mathbf{s}')$. Moreover, considering that the values of $\mathbf{V}$ and $\mathbf{s}$ are directly related, modulo a set of impedances, we can write either $\mathcal{IV}(\mathbf{s})$ or $\mathcal{IV}(\mathbf{V})$.

Corollary 4.1.7. *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ *with full incidence matrix* A *and partial incidence matrix* B *be a network with impedances* $\mathbf{Z} = diag(\mathbf{Z}_{mn})$*, true state* $(\mathbf{V}, \mathbf{s})$*, and predicted state* $(\mathbf{V}', \mathbf{s}')$ *such that* $\mathbf{V}' = \mathbf{V} + \sum_{k=1}^{|\mathcal{V}|} \delta_{\mathbf{k}} X_k$ *and let* $\lambda = min|\mathbf{Z}_{mn}|$*. Then*

$$\|\mathbf{s} - \mathbf{s}'\|_2 \leq \frac{\|A\|_2}{\lambda} \left(\sum_{k=1}^{|\mathcal{V}|} \mathcal{IV}(\mathbf{s})_k \cdot |\mathbf{X}_k| + \sum_{k=1}^{|\mathcal{V}|} \left(|\mathbf{X}_k|^2 + |\mathbf{X}_k| \cdot |\mathbf{X}_{n_k}| \right) \right), \tag{4.1.12}$$

where n_k *is the parent of* k*.*

Proof. Follows directly from Theorem 4.1.5 and Equations (3.3.8), (3.3.9) and (3.3.10). □

4.2 Properties of Incident Voltage score and Incident Voltage bound

Let us discuss some of the properties that the Incident Voltage scores has. The first property that we discuss here is that of independence. The $2-$norm $\|\mathbf{s} - \mathbf{s}'\|_2$ is dependent on

i) the true state $\mathbf{V}$,

ii) the predicted state $\mathbf{V}'$,

iii) the impedances $\mathbf{Z}$,

iv) and the incidence matrices A and B.

We are interested in knowing which locations to measure, meaning we would prefer a score to only be dependent on the incidence matrices, which fully encapsulates the topology of a network. Not having that luxury forces us to use statistical analysis in order to infer the influences of A and B out of a large set of samples with different $\mathbf{V}, \mathbf{V}'$ and $\mathbf{Z}$. The larger the variance, the larger the needed sample size of that variable. This increases calculation time exponentially.

In the next section we show that the influence of the impedances on the $2-$norm is very minimal. We do this by showing that isolating $\|\mathbf{Z}\|_2$ through the norm inequality is a relatively strict bound. Our newly found Incident Voltage score is even better. Indeed

$$\mathcal{IV}(\mathbf{V})_k \coloneqq |\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2},$$

is only dependent on the incidence matrices A and B, and the true states $\mathbf{V}$. We estimate a single $\mathcal{IV}(\mathbf{V})$ to be calculated in 0.65 seconds on a standard desktop. This was done using the build-in norm function of Rstudio to calculate the second component $\sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2}$, which takes up 60% of the calculation time.[1] We still need to calculate $\mathcal{IV}(\mathbf{V})$ for different $\mathbf{V}$. Let us discuss what can be considered a reasonable sample size for $\mathbf{V}$. Given that the Incident Voltage score is still dependent on $\mathbf{V}$, it is relevant to know the distribution of $\mathcal{IV}(\mathbf{V})$, given a set of true states. There are a few ways we can generate these states. Either trough a random process, or from historical data given to us by Alliander. For both methods of generating true states we calculated $\mathcal{IV}(\mathbf{V})$ for a hundred different $\mathbf{V}$ and collect all that data in one large matrix (or data frame) M where entry x_{ij} is the Incidence Voltage score of node i given the $j-th$ generated $\mathbf{V}$. We do this for multiple graphs, all with a 1000 nodes. We are interested in the variance of each row of M: for a given node i what is the variance in the different $\mathcal{IV}(\mathbf{V})_i$ for a given set of $\mathbf{V}$. First up, generating $\mathbf{V}$ by taking each entry of $\mathbf{V}$ out of the same normal distribution $N(\mu, \sigma)$ with $\mu = 1$ and $\sigma = 0.1$. Having generated one hunderd different $\mathcal{IV}(\mathbf{V})$ let MEAN($\mathcal{IV}$) be the vector of length $|\mathcal{V}|$ defined as

$$\text{MEAN}(\mathcal{IV})_i = \frac{1}{100}\sum_{j=1}^{100} x_{ij}.$$

We can do the same for the variance, maximum and minimum value of each row. In Figure 4.1 we show MEAN($\mathcal{IV}$) plotted and ordered from its highest value to its lowest value, bounded by the maximum and minimum of each row i. The substation is not considered, as it assumed to already be measured. We can directly see that the average distance between MEAN($\mathcal{IV}$) and the minimum or maximum is around 0.125, which is equivalent to 7% for most values. In Figure 4.2 we show the variance of each index, and in Figure 4.3 we show the variance of the best 100 nodes in more detail. These figures indicate that the variance very rarely exceeds 0.001 which is only a fraction of the values attained by MEAN($\mathcal{IV}$). Repeating this for different graphs of the same size and distribution of degrees we find the same results.

[1] It might be favorable to use different methods to improve calculation time

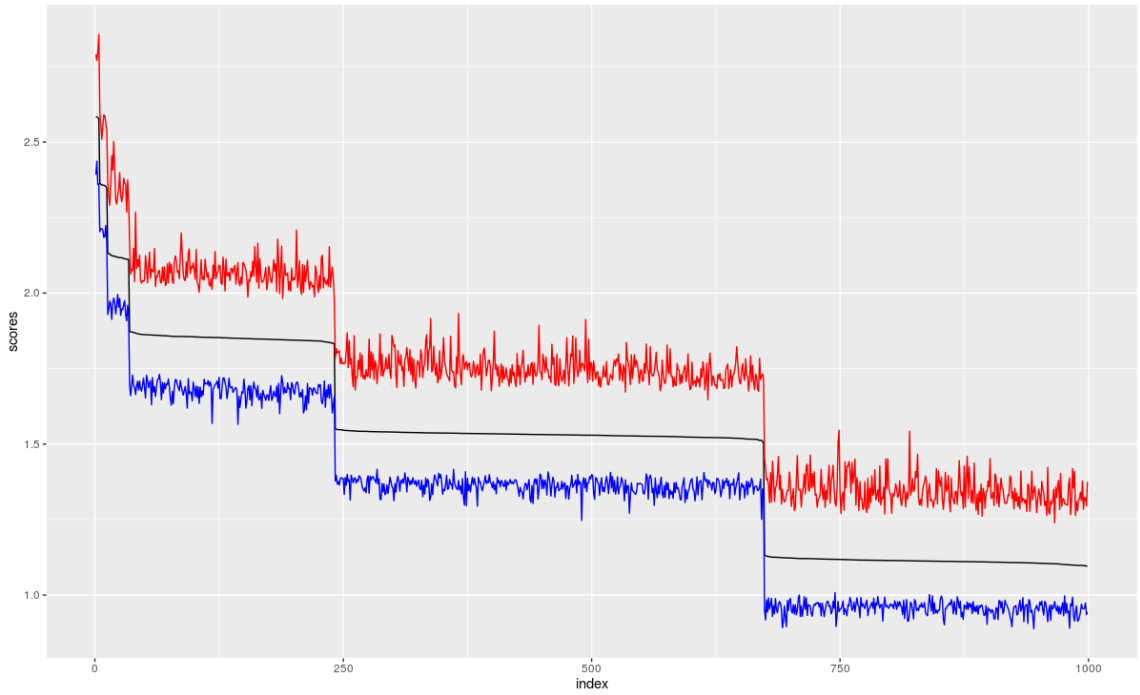

Figure 4.1: MEAN($\mathcal{IV}$) in black, ordered from highest to lowest and bounded by the minimum and maximum of each row i, in blue and red respectively.

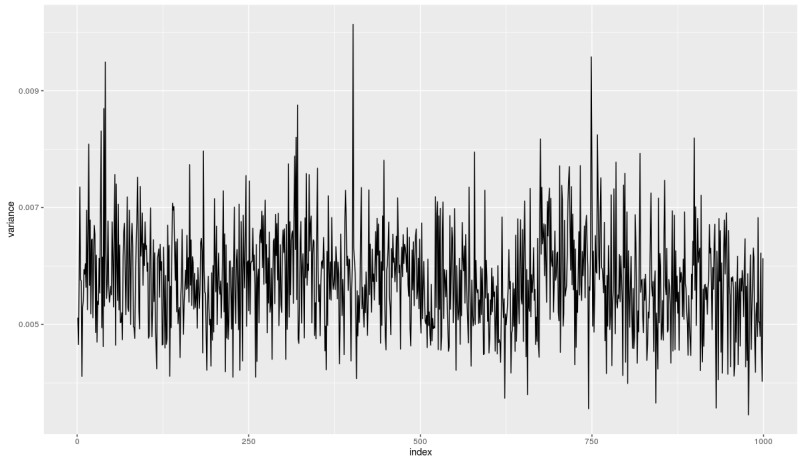

Figure 4.2: The variance of the each row i of M sorted by MEAN($\mathcal{IV}$) .

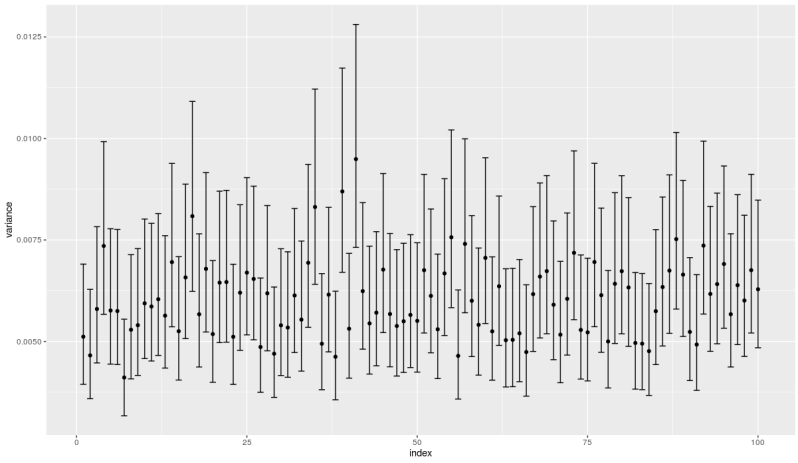

Figure 4.3: The variance of the first 100 nodes sorted by MEAN($\mathcal{IV}$) indicated by the black dots, with a 95% confidence interval given a hundred different $\mathbf{V}$ generated by a normal distribution.

Next, we show that a different distribution for $\mathbf{V}$ generates different results. A distribution that has more extreme values then the normal distribution will produce more variance. As to why, let us start by observing figure 4.1. In Figure 4.1 we see five occurrences of the mean score dropping significantly, and then remaining constant for a while. The system of equations

$$\begin{aligned} \mathbf{ZI} &= A^t\mathbf{V}, \\ \mathbf{S} &= (B^t\mathbf{V})\mathbf{I}^*, \\ \mathbf{s} &= -A\mathbf{S}. \end{aligned}$$

holds the most amount of information in nodes with a high degree. The Incident Voltage score also prefers nodes with a high degree. Recall

$$\mathcal{IV}(\mathbf{V})_k \coloneqq |\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2},$$

where A is the full incidence matrix, B is the partial incidence matrix and e_k denotes the arrow with node k as its sink. The vector $A_{(\cdot,e_k)}$ equal 1 at the parent of k, it equals -1 at k and 0 everywhere else. This means that a high variation between the parent node and its child k constitutes a high contribution of $|\langle A_{(\cdot,e_k)}, \mathbf{V}\rangle|$ to $\mathcal{IV}(\mathbf{V})_k$. The second component

$$\sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |\langle B_{(\cdot,e)}, \mathbf{V}\rangle|^2},$$

does not have the same property, as it only considers sinks. This component is influenced by the amount of arrows that are incident to k. More incident arrows imply a larger sum of non-negative real numbers. It is also influenced by the distribution of $\mathbf{V}$. For example, if a junction with a high degree, say 7, constitutes a voltage that is near zero, it is beaten by a node with a low degree if it has at least 7 times the voltage.

One final note before we move on, the score $\mathcal{IV}(\mathbf{V})_{substation}$ would exceed most other scores. In most cases the substation is the node which has a far larger degree then any other node. It will therefore always have the highest score, by the second component. At Alliander, each substation is measured, and therefore it is not relevant to consider the substation as a possible measuring location, even though its score will often exceed that of other nodes. The substation is not considered in any of the plots produced in this thesis.

We would expect a direct correlation between high degree and large $\mathcal{IV}(\mathbf{V})_k$ if the distribution of $\mathbf{V}$ is uniform among all nodes. That is, we can bound the difference between any two voltages by a relatively low number. More formally there is a suitable small $\epsilon > 0$ such that

$$\forall m, n \in \mathcal{V}\colon |\mathbf{V}_n - \mathbf{V}_m| \leq \epsilon, \tag{4.2.1}$$

A reasonable candidate for ϵ might be $\epsilon := \max_{n\in\mathcal{V}} \mathbf{V}_n$. Any $\mathbf{V}_n$ that would exceed this ϵ might influence its Incident Voltage score in a manner unlike the other $\mathbf{V}_m$.

In Figure 4.4 we can see this correlation between high degree and high Incident Voltage score when property (4.2.1) holds. Taking each $\mathbf{V}_n$ from the same normal distribution $N(\mu, \sigma)$ with $\mu = 1$ and $\sigma = 0.1$ yields the results of Figure 4.4. It directly shows how the nodes are put into different groups, depending on the degree of that node. Each jump in score correlates to a jump in degree. If Property (4.2.1) does not hold, we need to know the congestion areas of the graph: where are the voltages generally high?

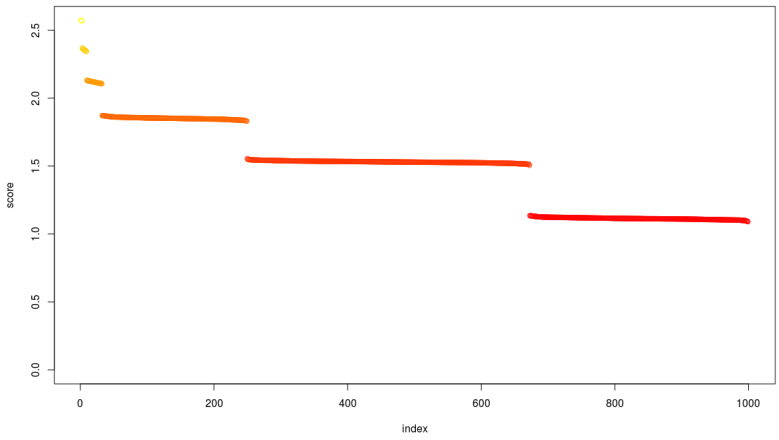

Figure 4.4: The mean score for highest to lowest score colored by the degree of said index. From yellow to red: lowest degree (1) to highest degree (6).

Let us take a different approach to simulate $\mathbf{V}$, where we consider the generalized extreme value distribution (GEV distribution). The GEV distribution is widely used in the treatment of "tail risks" in fields ranging from insurance to finance. In Figure 4.5 we see a normal distribution $N(0, 1)$ plotted versus two generalized extreme value distributions.

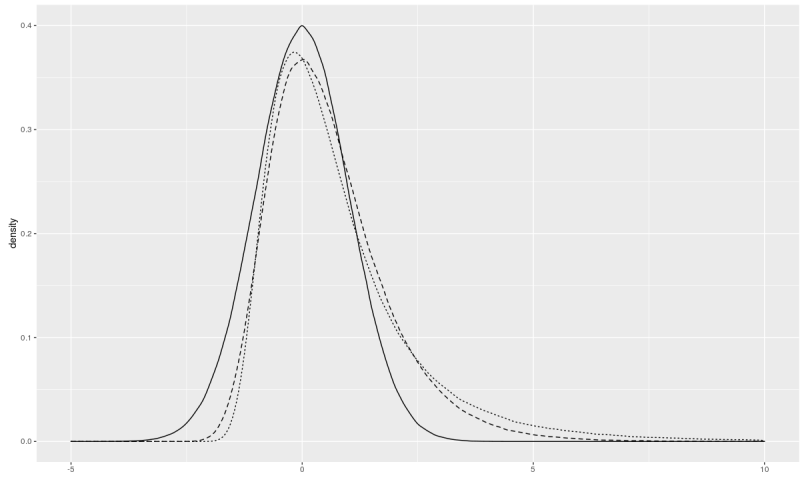

Figure 4.5: A normal distribution (continuous line) overlayed with two generalized extreme value distributions (dotted lines).

For this new method of simulating $\mathbf{V}$ there still seems to be a correlation between a high degree and a high Incident Voltage score, though be it this time with more variance. We calculate matrix M, of which entry x_{ij} is the Incidence Voltage score of node i given the $j-th$ generated $\mathbf{V}$, for 100 different $\mathbf{V}$. Figure 4.6 shows the relation between the degree of a node, and its score. There are many extreme points, which is affirmed by the variance, as plotted in Figure 4.7. The variance can now reach over 25% of the original values. Thus a larger sample of $\mathcal{IV}(\mathbf{V})$ must be calculated[2] in order to guarantee all possible voltage vectors are represented.

[2]compared to the normal distribution

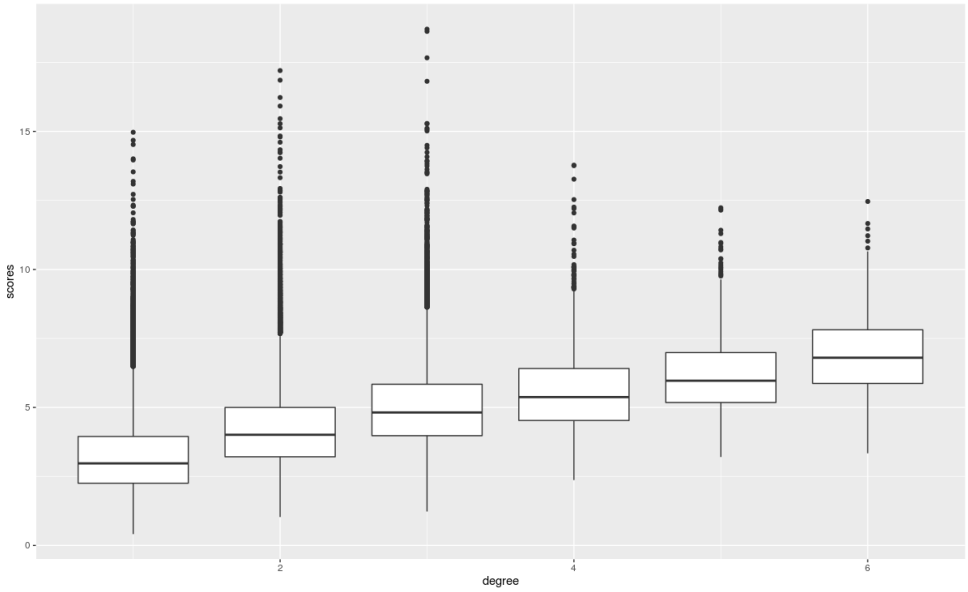

Figure 4.6: A boxplot for each possible degree a node can attain in $\mathcal{G}$, plotted against the Incident Voltage score of those nodes.

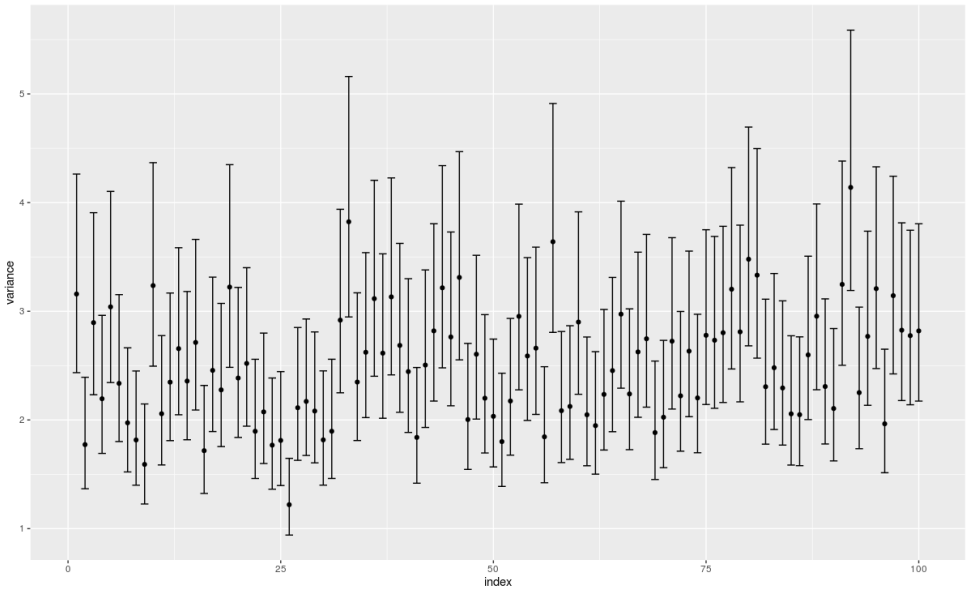

Figure 4.7: The variance of the first 100 nodes sorted by MEAN($\mathcal{IV}$) indicated by the black dots, with a 95% confidence interval given a hundred different $\mathbf{V}$ generated by GEV.

4.3 Accuracy of Incident Voltage bound and the Incident Voltage score

We now have a scoring function that is independent of the impedances and the forecaster, but the Incident Voltage bound is only an upper bound of the 2−norm score and might not be as accurate. In this section we discuss the accuracy of the Incident Voltage bound and the Incident Voltage score. Fire we discuss the growth rate of the 2−norm with respect to the size of the graph. If we calculate $\mathcal{IV}(\mathbf{s}, \mathbf{s}')$ for enough $\mathbf{s}$, we can give an upper bound that bounds $\mathcal{IV}(\mathbf{s}, \mathbf{s}')$ in 95% of all cases. The same can be done for all $\mathbf{X}_k$. Let $\mathcal{IV}$ and $\mathbf{X}$ be these 95% accurate bounds. This gives us a more direct bound.

Corollary 4.3.1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with full incidence matrix A and partial incidence matrix B[3] be a network with impedances $\mathbf{Z}$ =diag($\mathbf{Z}_{mn}$) true state $(\mathbf{V}, \mathbf{s})$ and predicted state $(\mathbf{V}', \mathbf{s}')$ such that $\mathbf{V}' = \mathbf{V} + \sum_{k=1}^{|\mathcal{V}|} \delta_{\mathbf{k}} X_k$. Moreover, let $\mathcal{IV}$ and $\mathbf{X}$ be as mentioned above. Then we can say with 95% accuracy that*

$$\|\mathbf{s} - \mathbf{s}'\|_2 \leq \frac{\|A\|_2}{\lambda} \cdot |\mathcal{V}| \cdot \left(\mathbf{X} \cdot \mathcal{IV} + 2\mathbf{X}^2\right)$$

.

With Corollary 4.3.1 we see that $\|\mathbf{s} - \mathbf{s}'\|_2$ scales (at most) nearly linearly with the amount of nodes. That is, if the forecaster behaves somewhat predictably so we can estimate $\mathbf{X}_k$. We estimated $\mathcal{IV}$ for voltages that lie between 100 and 130 Volt, which decreases as we move further away from the substation. We find that $\mathcal{IV} \approx 220$.

Empirical data shows that in this case the error $\|\mathbf{s} - \mathbf{s}'\|_2$ is actually logarithmic to the amount of nodes $|\mathcal{V}|$. For a graph of n nodes we took ten true states, and for each true state simulated 10 predicted states. The 2−norm of the error was calculated for each pair and the means of those 100 errors where plotted versus the amount of nodes n (see Figure 4.8).

[3] which is needed to calculate $\mathcal{IV}$

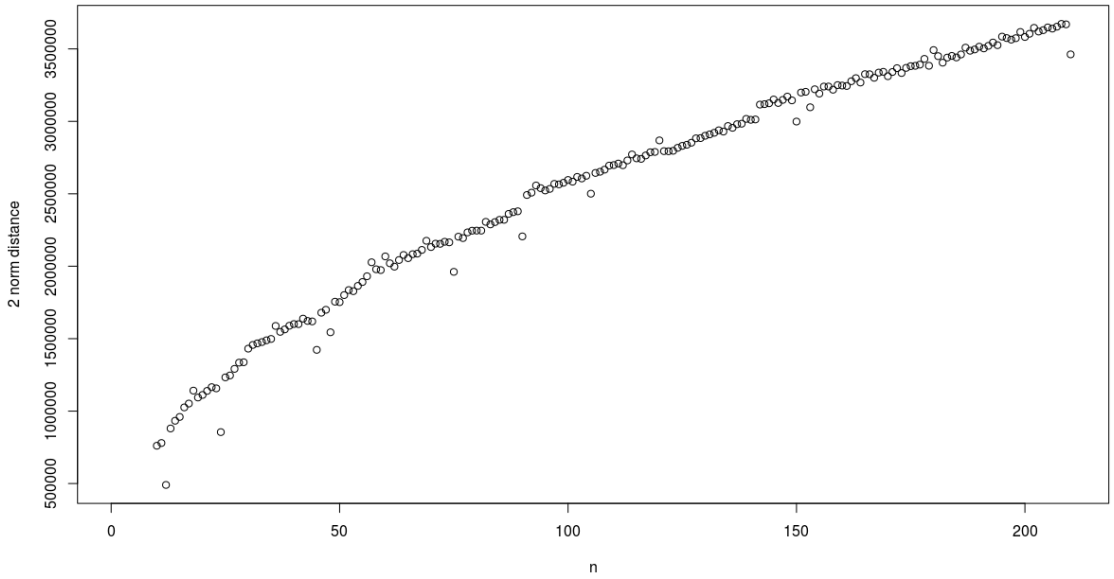

Figure 4.8: The amount of nodes n plotted versus the mean 2-norm of 100 cases.

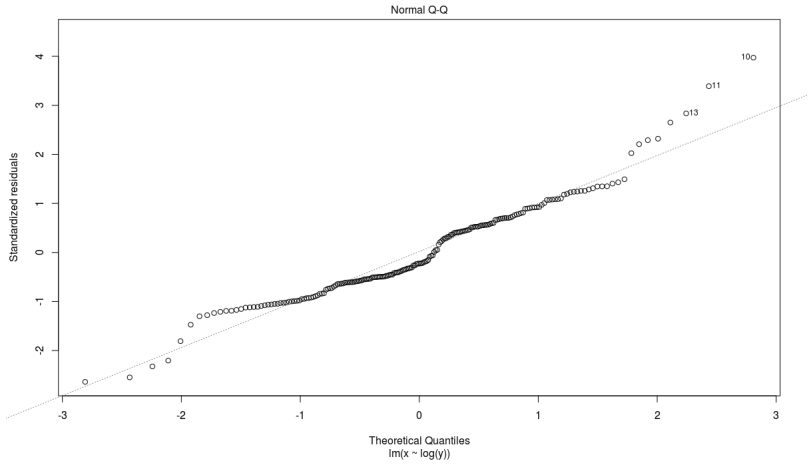

Figure 4.9: A normal Q-Q plot of the normed error x and the amount of nodes y.

If we do some statistical analyses as provided by R-studio we can plot a normal Q-Q plot and the Residuals vs Fitted, as seen in Figure 4.9 and 4.10 respectively.

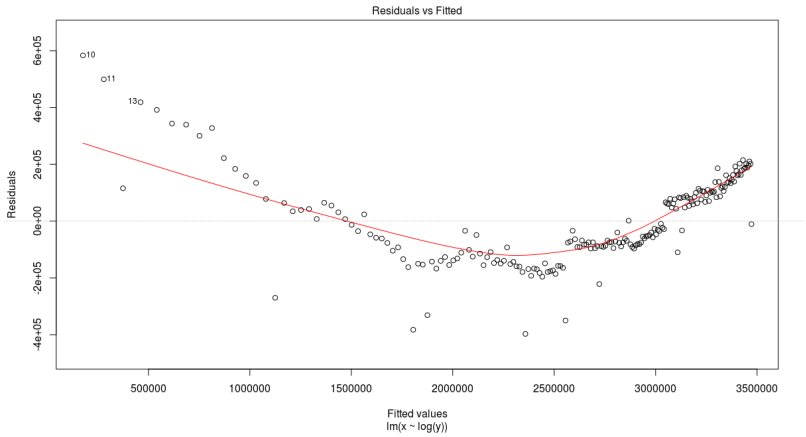

Figure 4.10: A Residual versus Fitted plot of the normed error x and the number of nodes y.

The disadvantage of the Incident Voltage score compared to the original $2-$norm score is that the bound is derived by using the triangle inequality

$$\|x - y\|_2 \leq \|x\|_2 + \|y\|_2$$

for norms and the matrix norm bound

$$\|Ax\|_2 \leq \|A\|_2 \|x\|_2.$$

Each time such a inequality is used the bound becomes less precise. The triangle inequality is most precise when the vectors are perpendicular, and the matrix norm inequality is attained if x the vector for which the matrix norm is $\|A\|_2 = \|Ax\|_2$. Considering we want to take the most common $\mathbf{V}$, who are quite random by nature, these inequalities will most likely not be attained. This means that directly calculating the $2-$norm score might result in different candidates. Direct calculations do however not have the advantage of being independent of $\mathbf{Z}$ and the predicted state $\mathbf{V}'$, meaning we are facing with (more) variance which may temper with the results. Therefore we must weigh the loss of accuracy caused by the inequalities to the loss of accuracy by our ability to asses how the initial forecaster behaves. Let us look at some case studies to further determine the accuracy of the Incident Voltage bound and the Incident Voltage score.

Recall the Incident Voltage bound

$$\frac{\|A\|_2}{\lambda} \cdot \left(\sum_{k=1}^{|\mathcal{V}|} |\mathbf{X}_k| \mathcal{IV}(\mathbf{s})_k + \sum_{k=1}^{|\mathcal{V}|} \left(|\mathbf{X}_k|^2 + |\mathbf{X}_k| \cdot |\mathbf{X}_{n_k}| \right) \right).$$

This bound is attained via some triangle inequalities and some matrix norm bounds (see Lemma 3.3.8). To be more precise, we can separate our process of obtaining the Incident Voltage bound into five steps:

1) the actual $2-$norm,

2) isolating the operator norm of the incidence matrix via Lemma 3.3.8,

3) isolating the operator norm of all impedances via Lemma 3.3.8,

4) a triangle inequality over the three components R_1, R_2 and R_3 as described in Theorem 4.1.5,[4]

5) the final triangle inequality through the sum of all nodes that results in the Incident Voltage bound.

First we show that the largest loss of accuracy comes from the last step, step 5. This is intuitively logical, as it is a triangle inequality bound over a sum of $|\mathcal{V}|$ elements, whereas the rest of the bounds are not as dependent on the size of $\mathcal{V}$.[5] In Figure 4.11 we see step 1) to 5) simulated for a graph on 318 nodes and 35 routes, a given true state $\mathbf{V}$ and size predictive states $V_1 \dots V_6$. The predicted states are generated by $\mathbf{V}' = \mathbf{V} + \sum_k \delta_k \mathbf{X}_k$ with $\mathbf{X}_k$ out of the GEV distribution with bias.

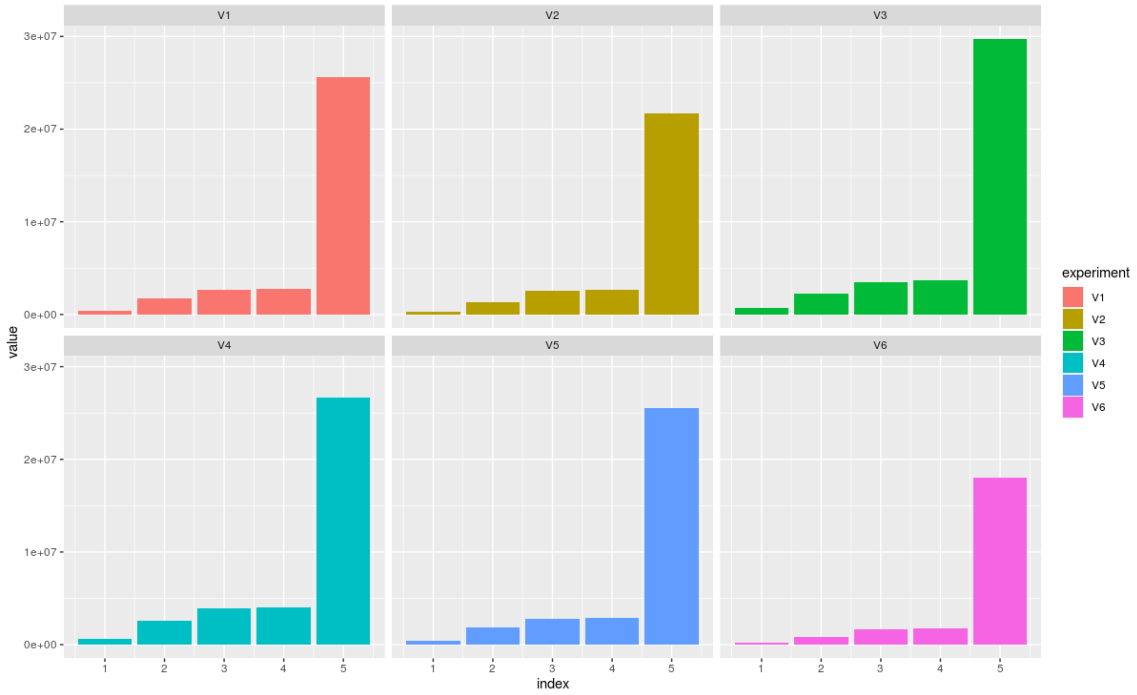

Figure 4.11: Given $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = 318$ and $\deg(1) = 35$, bound one to five are calculated for six different predictive states $V_1 \dots V_6$ with bias.

We see that the last step has significantly worsen the accuracy, compared to the other steps. Preferably, we want the Incident Voltage score to influence step 4 as much as it influences step 5. Whether or not this is true, in this case we do know that the resistances do not influence the

[4]Note that these components are slightly different then the ones described in Theorem 4.1.5, as we already isolated the impedances.

[5]Isolating $\|A\|_2$ might be slightly influenced by $|\mathcal{V}|$, as this norm is dependent on the number of routes.

2−norm score significantly. This means that any Monti Carlo process we apply to calculate the mean 2−norm score does not need different impedances.

The system of equations

$$\begin{aligned}\mathbf{ZI} &= A^t\mathbf{V},\\ \mathbf{S} &= (B^t\mathbf{V}) \circ \mathbf{I}^*,\\ \mathbf{s} &= -A\mathbf{S},\end{aligned}$$

can be partitioned into a collection of systems, one for each route. It is possible to solve the state estimation for each route individually, because a voltage of node n only influences the powers of its neighbouring nodes and thus never the nodes in other routes. We can produce full and partial matrices for each route and calculate the 2−norm for each route separately. A quick empirical test does however show that this separation does not always significantly improve the last step, which is the most valuable to improve. In Figure 4.12 we show that first calculating Incident Voltage-bounds for each route and then summing them does not improve the original Incident Voltage bound by any significant margin.

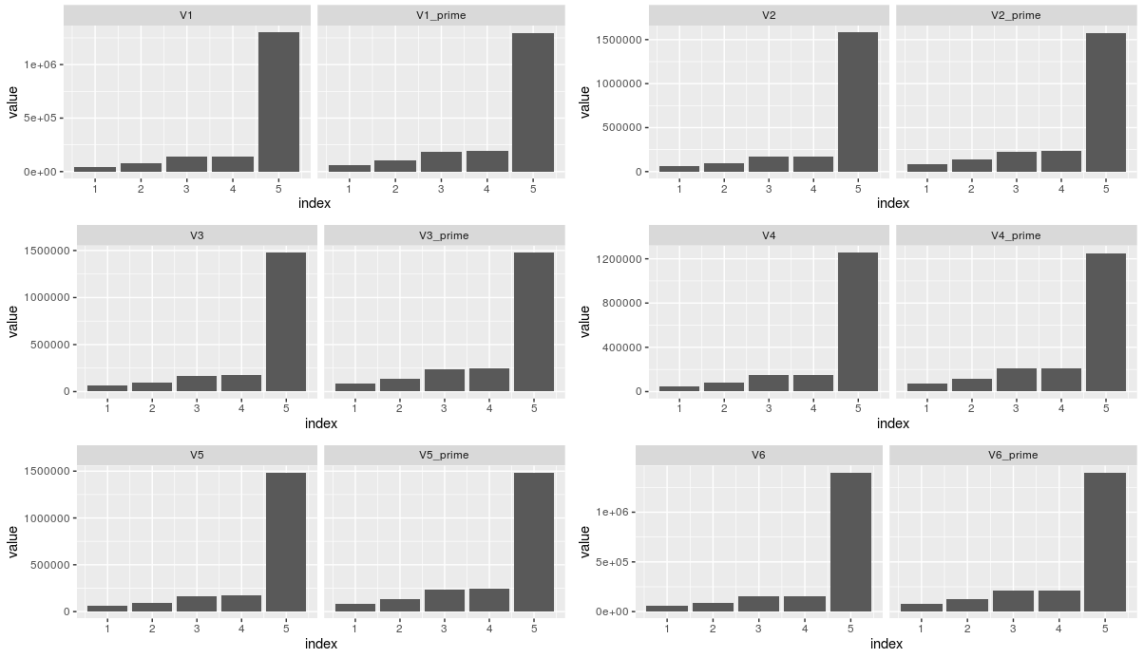

Figure 4.12: Given $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = 100$ and $\deg(1) = 2$, the bounds are calculated for six different predictive states $\mathbf{V_1}, \ldots, \mathbf{V_6}$, the left plots are for the entire graph, the right plots depict the sum of values of both routes.

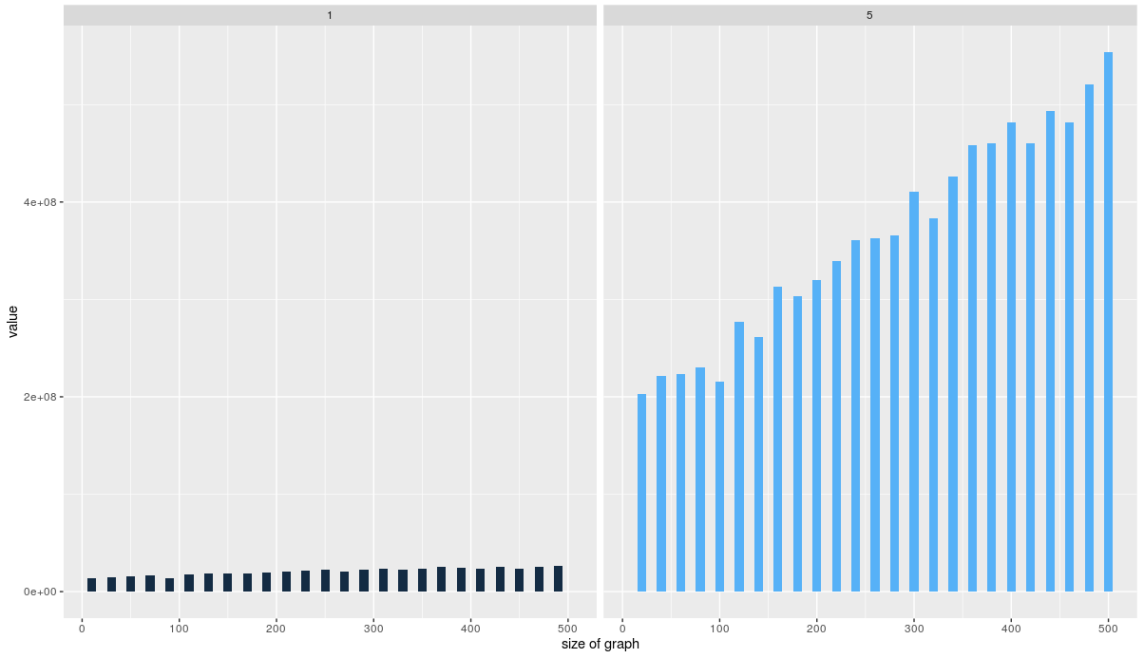

Figure 4.13: The 2−norm (left) and Incident Voltage bound (right) are calculated for different size $|\mathcal{V}|$ with $\deg(1) = 1$.

After the last step it is the first step that is most significant. Figure 4.13 shows what happens when we let the size of $\mathcal{V}$ vary. It shows that the significance of the last step increases as the graph increases in size.

Figure 4.13 indicates that for smaller graphs the Incident Voltage bound is more precise. Let us look at how the Incident Voltage score influences the original 2−norm score and the Incident Voltage bound. Because the Incident Voltage norm increases the bound drastically, it is not certain whether the nodes preferred by the Incident Voltage score are the same nodes preferred by the 2−norm score. We show a case where both scores pick the same candidates. In Figure 4.14 we plotted both the 2−norm score and the Incident Voltage-norm score for the pairs $(\mathbf{V}, \mathbf{V}')$ given measured locations $\mathcal{O}$ with $|\mathcal{O}| = 1$. Figure 4.14 shows that low scores, which are desirable, are attained at the same nodes by both scores. It is near the end of the route where the scores seem to differ. Comparing these scores to the actual Incident Voltage-score as per Definition 4.1.6, not much can be be said just yet, as there is still quite a bit of variance in play. We need to compare many different true states to many different predictive states to see if these two scores also align with the original Incident Voltage-score.This will be done in the next chapter.

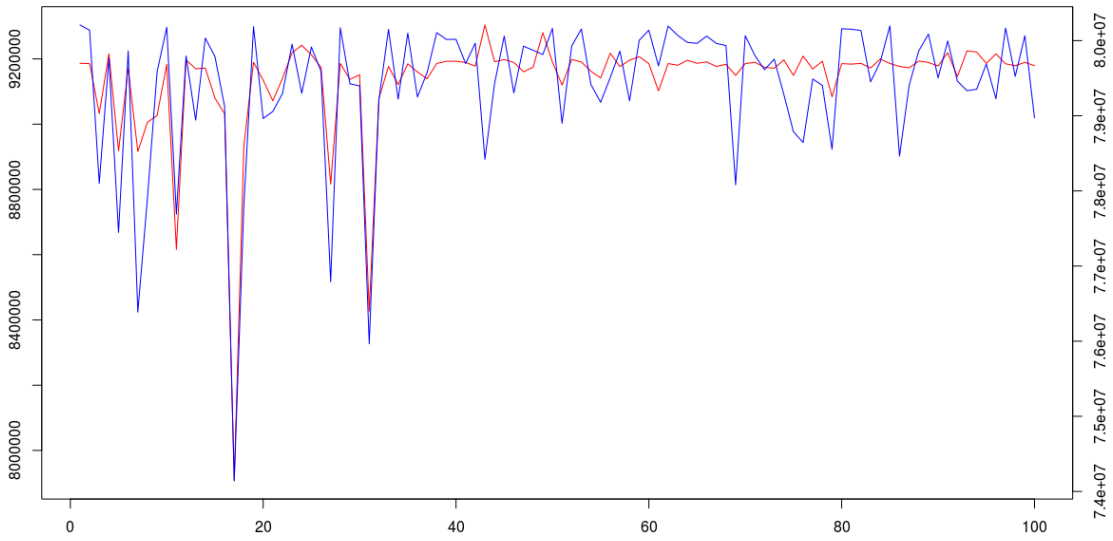

Figure 4.14: For a graph with 100 nodes and $\deg(1) = 1$ a true and predictive state are scored. On the x-axis we have the index of the measured node, on the y-axis we have the $2-$norm score in red, overlapped by the Incident Voltage bound in blue.

We will end this section by noting that the Incident Voltage score can be used on weighted graphs.We will need the weighted (full) incidence matrix.

Definition 4.3.2. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with arrow weights $W = \{W_e \in \mathbb{R} \mid e \in \mathcal{A}\}$, we call the binary matrix $A \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{A}|}$ defined by*

$$A_{me} = \begin{cases} \sqrt{W_e} & \text{if } m \text{ is the source of } e \\ -\sqrt{W_e} & \text{if } m \text{ is the sink of } e \\ 0 & \text{else.} \end{cases} \tag{4.3.1}$$

the weighted full incidence matrix of $\mathcal{G}$.

A similar definition is possible for the partial incidence matrix. More information can be found in [15]. Exploring the effects of weighted incident matrices is not within the scope of this thesis. If the need to compensate for areas which require measuring for reasons other than those explored in this model, a weighted incidence matrix might be the solution. One possible reason would be the existence of congestion areas.

Chapter 5

Results

5.1 General method

In this chapter we try to determine whether the Incident Voltage score is preforms better compared to other methods, given certain forecasters.

Recall that for larger graphs $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ and a non-trivial budget B of at least 3 nodes, the amount of subsets $\mathcal{O} \subseteq \mathcal{V}$ with $|\mathcal{O}| = B$ is quite large. For $n = |\mathcal{V}|$ there are $\binom{n}{B}$ sets of size B, which quickly become an insurmountable amount. We can therefore only calculate the score of a prediction pair $(\mathbf{V}, \mathbf{V}')$ for a select number of sets $\mathcal{O}$, as to not calculate until the end of time. In Section 4.2 we showed that the Incident Voltage score prefers nodes with a high degree, and dislikes nodes with degree equal to one (i.e. end nodes). We shall be comparing the nodes produced by the Incident Voltage score with samples of high degree. First we discuss the general method.

Definition 5.1.1. *For a graph $G = (\mathcal{V}, \mathcal{A})$, a budget $B \in \{0, \dots, |\mathcal{V}|\}$ and Incident Voltage score $\mathcal{IV}(\mathbf{V})$ we can sort the score from highest (best) to lowest (worst): $\{iv_1, \dots, iv_{|\mathcal{V}|}\}$. The Incident Voltage set for budget B is then $\{iv_1, \dots, iv_B\}$, the first B nodes of the sorted score, generally indicated by $\mathcal{O}_{\mathcal{IV}(\mathbf{V})}$ given the budget B.*

We shall calculate a score $f(\mathbf{V}, \mathbf{V}')$ for a given $\mathbf{V}$ and 100 associated $\mathbf{V}'$ for 100 sets $\mathcal{O}$ with $|\mathcal{O}| = 20$ from a certain range $\mathcal{WV}$. With these assumptions there are still a few variables we control. We can vary

i) the score f,

ii) the forecaster F,

iii) the range $\mathcal{W}$,

iv) and the graph $\mathcal{G}$.

There are two scores we shall study. The 2−norm score as in Definition 3.3.3 and the (normed) mean squared error-score.

Definition 5.1.2. *The (normed) MSE-norm score of true state* $(\mathbf{V}, \mathbf{s})$ *and predicted state* $(\mathbf{V}', \mathbf{s}')$ *is defined as*

$$MSE(\mathbf{s}, \mathbf{s}') = \left| \frac{1}{N} \sum_{n=1}^{N} (\mathbf{s}_n - \mathbf{s}'_n)^2 \right|.$$

We take the complex norm (i.e. absolute value) of the mean squared error (MSE), because the MSE might be a complex number, and scoring functions must produce real numbers. We show that even though the Incident Voltage score is derived from the $2-$norm score, it is also a good estimation of the MSE-score.

In this thesis we will not be using actual forecasters, which is not within our scope. Instead we create a statistical procedure to generate $\mathbf{V}'$ from $\mathbf{V}$. This means that we assume $\mathbf{V}'$ to adhere to some distribution around $\mathbf{V}$. We can vary the forecaster by varying the distribution which generates $\mathbf{V}'$. Recall that a prediction $\mathbf{V}'$ is defined such that $\mathbf{V}'_k = \mathbf{V}_k + \mathbf{X}_k$ for some $\mathbf{X}_k \in \mathbb{R}$, so generating $\mathbf{V}'$ from $\mathbf{V}$ is equivalent to generating $\mathbf{X}_k$ through some statistical distribution. We make a distinction between two types of distributions.

i) Smooth distributions. These are distributions that do not have many extreme points and most points cluster nicely around the mean. The most classical example is the normal distribution.

ii) Non smooth distributions. These distributions are such that some predictions drastically mispredicted the actual voltage. This is best represented by the generalized extreme value (GEV) distribution.

In Section 4.2 (in particular Figure 4.5) the GEV distribution has been previously mentioned. For each $k \in \mathcal{V}$ we can choose to give node k either a normal distribution with some bias, or a generalized extreme value distribution also with bias.

Now for the range we can take all nodes $\mathcal{W} = \mathcal{V} \setminus \{substation\}$ or just a selection of nodes with a high degree $\mathcal{W} = \{v \in \mathcal{V} \mid \deg(v) \geq T, v \neq substation\}$, where T is the highest degree such that $|\mathcal{W}| \geq B$. We need T to have this property or else we shall be sampling over a set with less nodes then we have a budget for. We can also give $\mathcal{W}$ some priority nodes. These are nodes that will be in each sample $\mathcal{O}$, giving them priority over the other nodes in $\mathcal{V}$. Again, we should be aware that this decreases the amount of possible samples, so we must choose T low enough to guarantee enough variation. An obvious subset to give priority to is the set of nodes $v \in \mathcal{W}$ that have the actual highest possible degree, that is $\deg(v) \gg T$.

There is also variation in the graph $\mathcal{G}$. A typical network can easily have 1000 nodes and 35 routes. We generate four different graphs of this size with similar distributions in nodal degrees. That is to say, all four graphs inherit the same amount of nodes with degree $5, 4, \ldots, 1$ up to a small difference. The variation mainly lies in the position of the nodes with high degree.

Let $\mathbf{V}$ be a true state and $\mathbf{V}'$ be one of the 100 predicted states. Then for measuring locations $\mathcal{O}$, we can compare the score $f(\mathbf{V}, \mathbf{V}'_{\mathcal{O}})$ to that of $f(\mathbf{V}, \mathbf{V}'_{\mathcal{O}_{\mathcal{IV}(\mathbf{V})}})$, where $\mathcal{O}_{\mathcal{IV}(\mathbf{V})}$ are the measuring locations generated by the Incident Voltage score, i.e. its 20 highest scores. Note that in this scenario we compare the actual voltage used in $\mathcal{IV}(\mathbf{V})$ to different $\mathbf{V}'$, not a profile of different $\mathbf{V}$, which is more realistic. Case studies where we use a profile of different $\mathbf{V}$ are discussed later on. In Figure 5.1 we see the comparisons for the $2-$norm and different $\mathcal{O}$.

We calculate the percentage of locations that generate a lower score then the nodes generated by the Incident Voltage score. For instance, in the case of Figure 5.1, that percentage would be

23%. Each $\mathbf{V}'$ will generate a different percentage in $[0, 1]$ and 100 predicted states will generate a vector of percentages, say *per*. It is possible to do some statistical analysis on this vector. We look at the distribution of *per* and calculate the p value for the following hypotheses test

H_0: $per \geq 0.5$

H_1: $per < 0.5$.

Rejecting the null hypothesis would indicate the Incident Voltage score is better than randomly choosing a set $\mathcal{O}$ in the given range $\mathcal{W}$. We also observe the 95% interval in which the mean of *per* lies, to show the significance of the Incident Voltage score, given the aforementioned choices of score f, forecaster F, range $\mathcal{W}$ and graph $\mathcal{G}$.

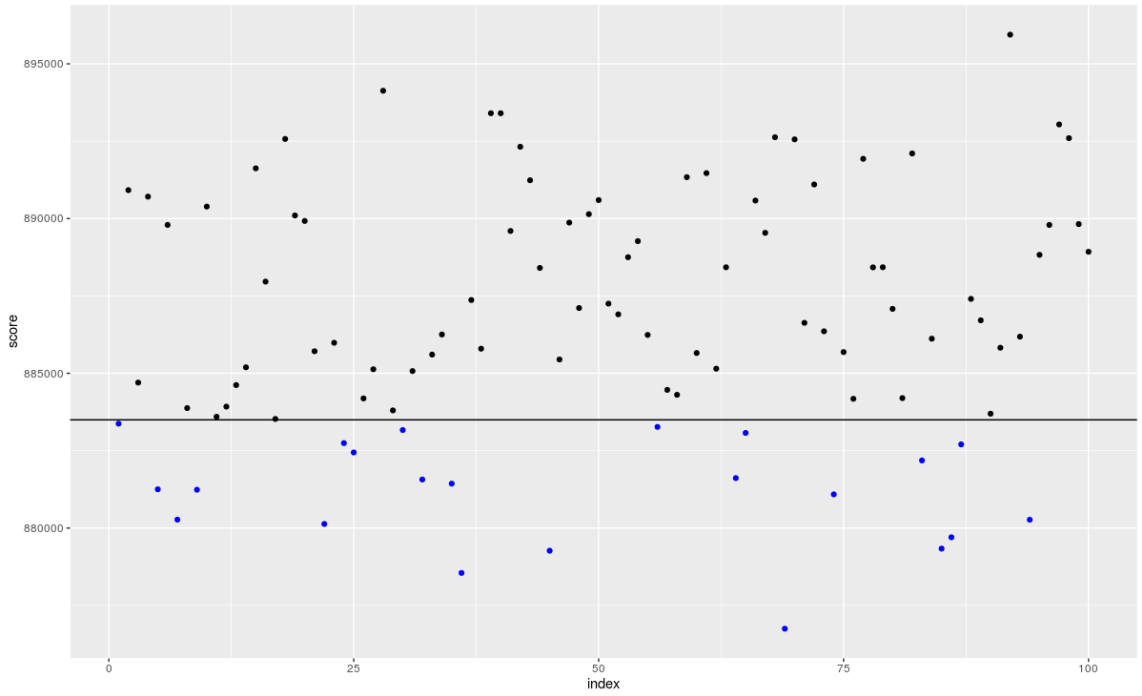

Figure 5.1: The 2−norm score $\|\mathbf{s} - \mathbf{s}'\|_2$ is plotted for 100 different measured sets $\mathcal{O}$ with $|\mathcal{O}| = 20$. In blue are all $\mathcal{O}$ that scored better then $\mathcal{O}_{\mathcal{IV}(\mathbf{V})}$, which is indicated by the horizontal line.

This vector *per* is still dependent on $\mathbf{V}$, and different $\mathbf{V}$ will generate different vectors $per_{\mathbf{V}}$, and different Incident Vector scores $\mathcal{IV}(\mathbf{V})$. This means that a p-value is dependent on $\mathbf{V}$, so we must run the simulation for different $\mathbf{V}$. It was chosen to run the simulation for 15 different $\mathbf{V}$ and use box-plots to show the distribution of all $p-$values.

Preferably we use historical data from Alliander to generate a set of common voltage vectors that represent a typical day in the power distribution grid. If this data is not available for voltage vectors, we must generate $\mathbf{V}$ by a random process.

Calculating a single *per* takes about one and a half hours computing time on a common laptop. So checking all different permutations of score, forecaster, range and graphs is not reasonable. In

the next section we start by showing that the Incident Voltage score scores better if the range is $\mathcal{W} = \mathcal{V} \setminus \{substation\}$, the set of all nodes. This requires more samples, as the range is so large, thus an even larger amount of computing time is needed. In my own coding work I quickly noticed that sampling over nodes with high degrees produces better scores then random sampling. For this reason we shall not look at any case studies in the case where $\mathcal{W} = \mathcal{V} \setminus \{substation\}$.

5.2 Incident Voltage set versus random sets of high degree

In this section we apply the method of Section 5.1 to the range $\mathcal{W} = \{v \in \mathcal{V} \mid \deg(v) \text{ relatively high}\}$, while prioritizing nodes with the highest possible degree, as mentioned in Section 5.1. We do this for six different case studies, each of which takes four large graphs ($\approx$ 1000 nodes) and repeats our method for fiveteen different true states, each with 100 corresponding predicted states and a sample size of 100 different sets in $\mathcal{W}$. The six experiments are displayed in Table 5.1. We take four experiments with the 2−norm score and also compare the Incident Voltage score to the MSE score in experiment five and six.

experiment	1	2	3	4	5	6
score	2−norm	2−norm	2−norm	2−norm	MSE	MSE
V generation	Outward	Outward	GEV	GEV	Outward	GEV
Forecasting	GEV	Normal	Normal	GEV	GEV	GEV

Table 5.1: Six different case studies for comparing the Incident Voltage score to random sets of high degree.

Let us briefly discuss what is indicated by the second and third row of Table 5.1. As previously mentioned, all voltage data has been generated by statistical processes, which is advised if no historical data is available. I have created two different methods for generating such a voltage profile. First up is the GEV generation, which creates the most diverse voltage vectors. Lets say we generate 100 different true state on 10 nodes for a single phase. The distribution might look something like Figure 5.2.

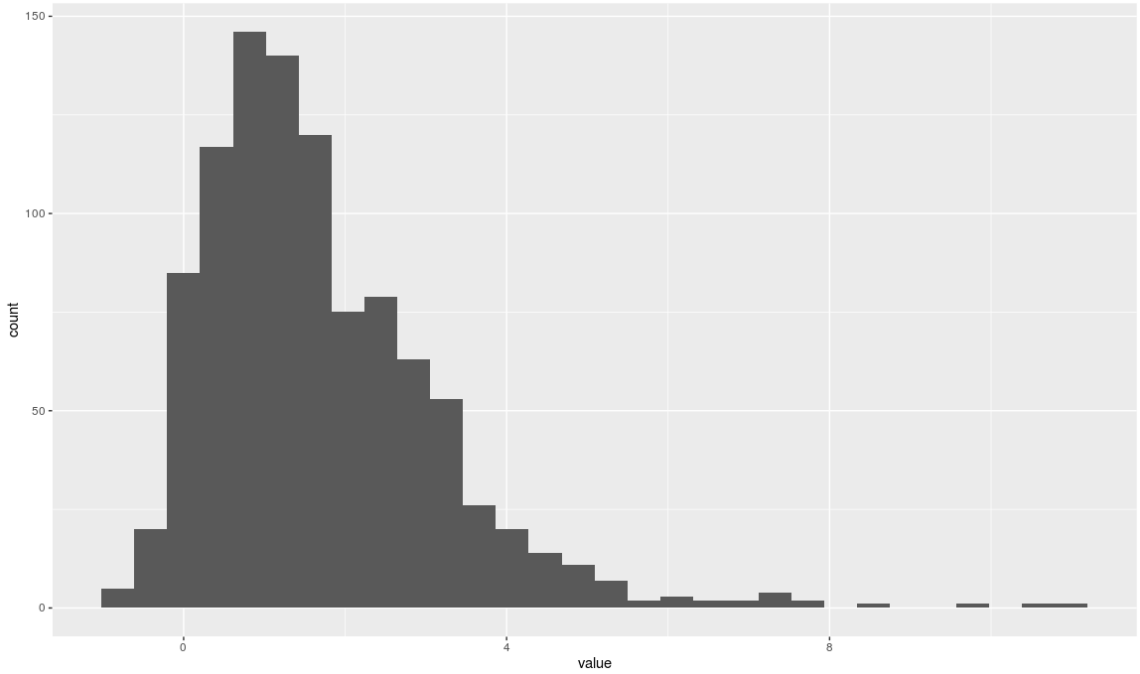

Figure 5.2: Distribution of elements in true states for GEV distribution.

We can also let the values be outwards decreasing. That is to say, the values will be lower the further away they are from the substation and will never have a value higher than its parent. We provide the generation function with an upper a and lower bound b and divide the interval $[a, b]$ into n smaller sections $[a_i, b_i]$ such that $a_0 = a, a_i = b_{i-1}$ for $n > i > 0$ and $b_n = b$. The amount of sections n equal the depth of the tree, i.e. the length of the longest path from the substation. Then a given node m will have $\mathbf{V}_m \in [a_i, b_i]$ for i the depth of node m with respect to the substation. The substation is given $\mathbf{V_0} = a$ and voltage of other nodes are chosen by a uniform distribution. Figure 5.3 shows us what us a possible distribution for $a = 100$ and $b = 130$.

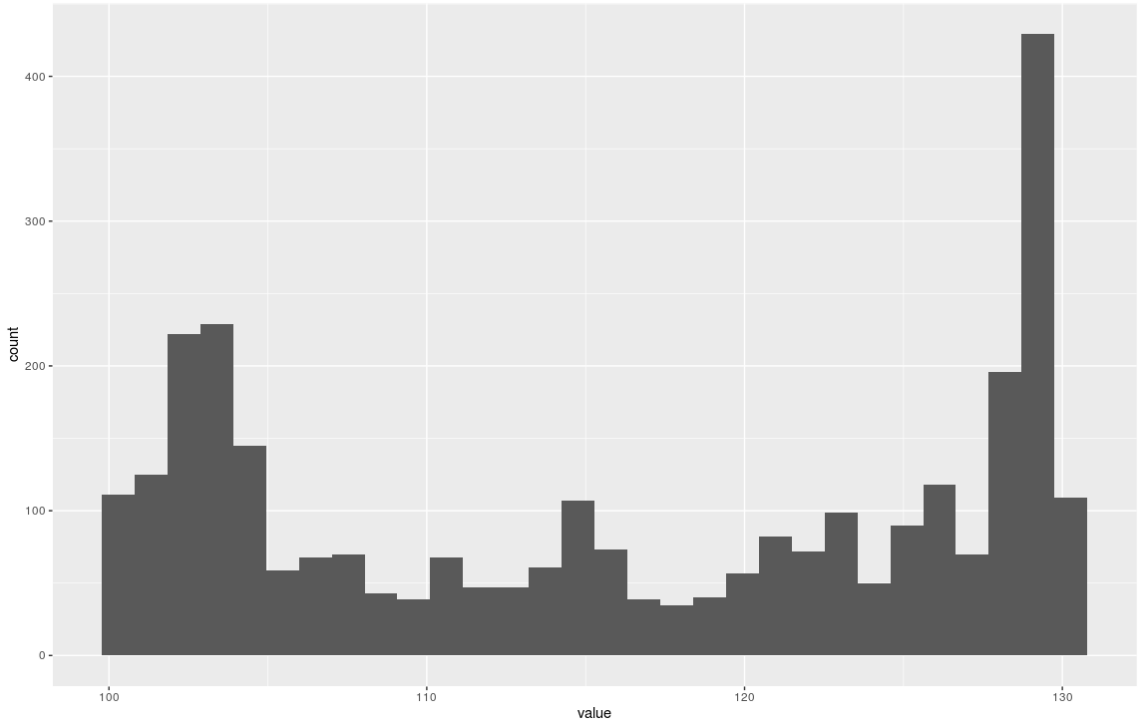

Figure 5.3: Distribution of elements in true states for outwards distribution.

For simulating forecasters we randomly choose $\mathbf{X}_n$ such that $\mathbf{V}'_n = \mathbf{V}_n + \mathbf{X}_n$. We let $\mathbf{X}_n$ either be come from of a normal distribution or a GEV distribution. In both cases the distribution has the same parameters, while also adhering to different bias. That is, $\mathbf{X}_n = \mu + p$ for some translation μ and p out of the mentioned distribution (normal or GEV). The global $p-$values for the hypothesis tests in Section 5.1 are always very low, never reaching above $1.00e-05$ for each of the graphs, often as low as $1.00e-40$. However, displaying the values for each true state $\mathbf{V}$ tells a different story, one with more variance. The global means of *per*, which indicated what percentage of samples scored better then the Incident Voltage score, have been displayed in Table 5.2. Note that there are four values, as we tested four graphs.

experiment	1	2	3	4	5	6
score	2−norm	2−norm	2−norm	2−norm	MSE	MSE
$\mathbf{V}$ generation	Outward	Outward	GEV	GEV	Outward	GEV
Forecasting	GEV	Normal	GEV	normal	GEV	GEV
means of *per*	0.39	0.44	0.36	0.24	0.40	0.38
given graph $\mathcal{G}$	0.33	0.42	0.34	0.12	0.35	0.33
	0.33	0.42	0.30	0.12	0.46	0.36
	0.22	0.44	0.33	0.11	0.26	0.33

Table 5.2: Case study results.

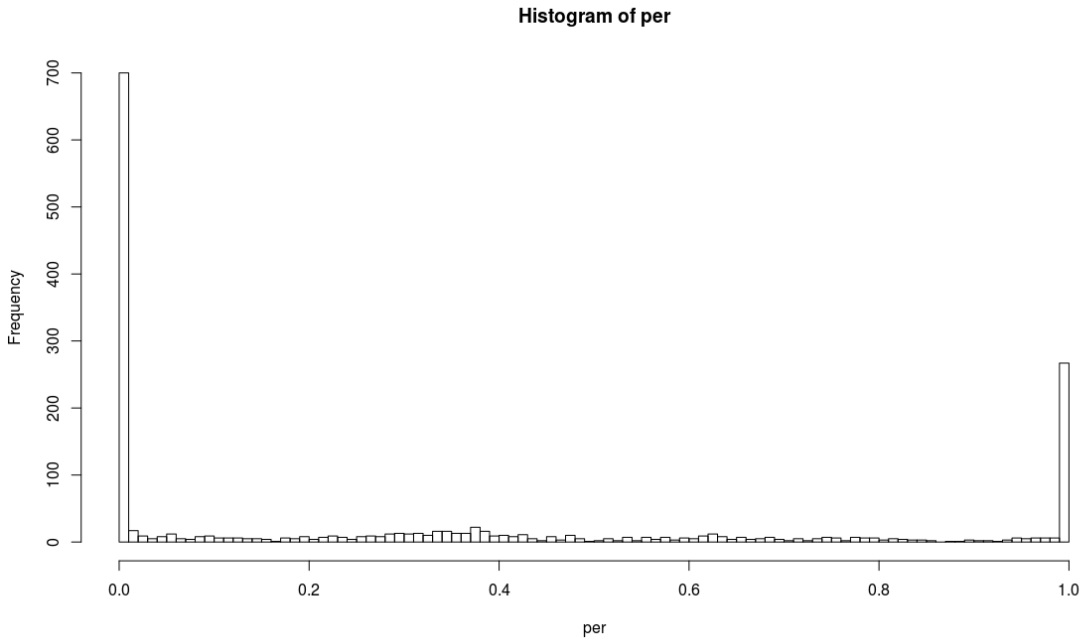

Figure 5.4: Distribution of *per* for experiment 6 over an fifteen different true states, each with 100 predicted states.

Out of these experiments, we see that experiment 4 produces the best results, and experiment 2 the worst. Experiments 5 and 6 show the score also improves the MSE score. Looking at the distribution of *per* for experiment 6 reveals more about the score. In Figure 5.4 we see that in most cases the Incident Voltage score either scores better than any sample, or worse then any sample. I was not able to proof why this is, though I hypothesise that it is due to codependency in measuring locations. Sometimes, when a node is predicted poorly, it affects its neighbours. Its neighbours will

then negatively effect the 2−norm if they are measured, even though in general they will have a positive contribution. Since the Incident Voltage score is not dependent on predicted states, it will not take these poorly predicted nodes into consideration. This is reasonable as in the real world this would also not be possible.

This is not an isolated case. In fact, all four graphs show this pattern for experiment 6. Figure 5.5 shows the histograms for experiment 5. In Figures 5.6 and 5.7 we see the histograms per graph for the 2−norm experiments with the best and worst mean values respectively (namely experiment 4 and 2). They too show the pattern of spikes at $per = 0$ and $per = 1$.

Seeing as different voltage profiles produce difference levels of accuracy of the Incident Voltage score, it would be wise that users of the score provide the functions with a large variation of profiles. This is reinforced by the finding in Section 4.2, which states that different profiles will recommend different nodes, though they will always come from the set of nodes with high degree. It is therefore recommended to provide different profiles. Perhaps one that represents a typical morning in the summer, while a second one represent a typical evening of that same summer. Or perhaps the user wishes to distinguish between seasons, or workdays versus weekends. That is up to the user.

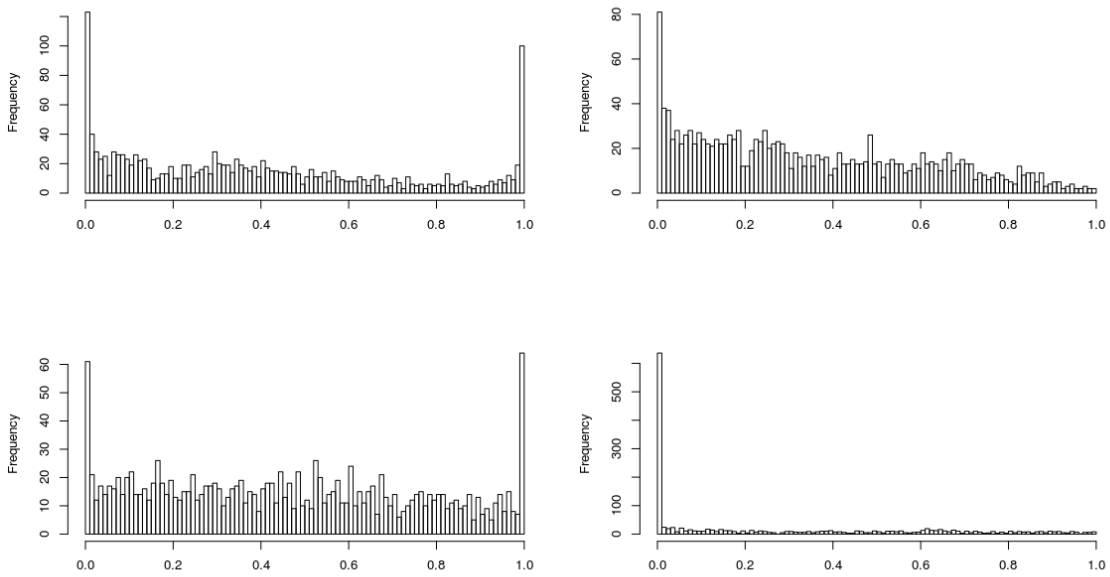

Figure 5.5: From top left to bottom right (in reading direction): the distribution of *per* for the four graphs of experiment 5.

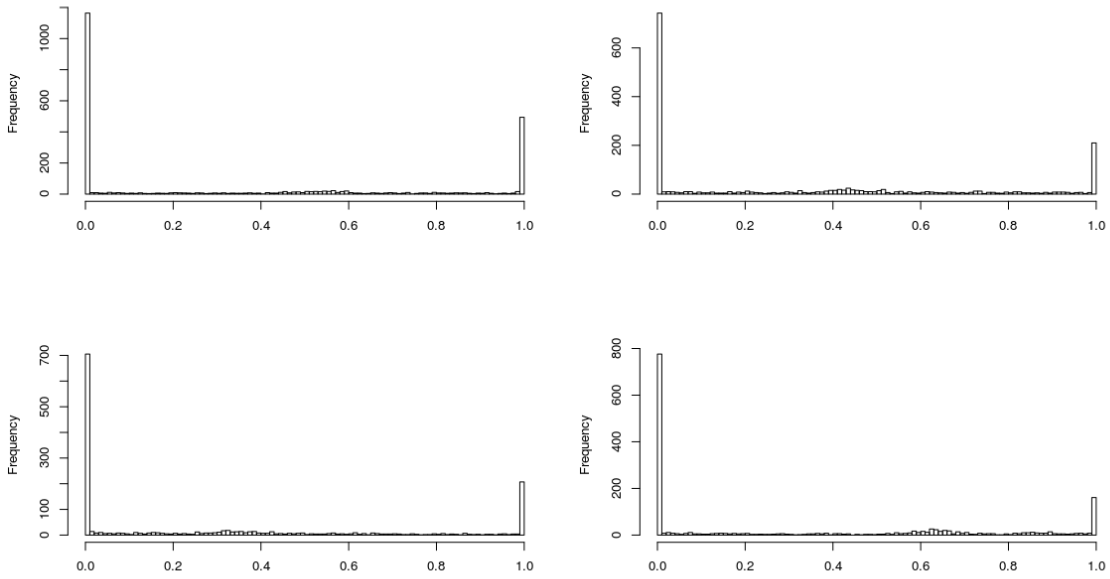

Figure 5.6: From top left to bottom right: the distribution of *per* for the four graphs of experiment 4.

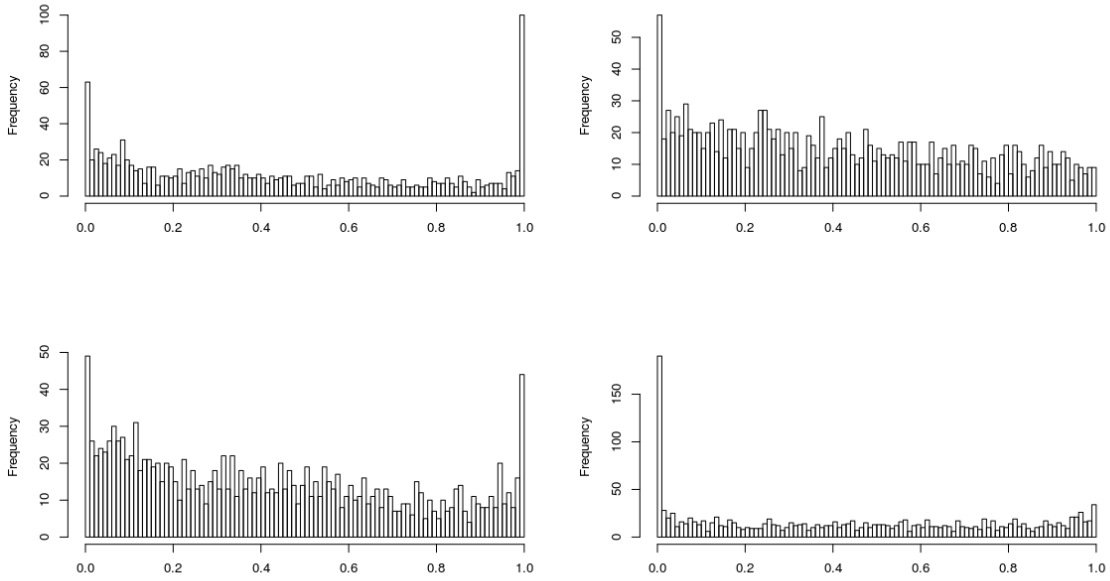

Figure 5.7: From top left to bottom right: the distribution of *per* for the four graphs of experiment 2.

5.3 Incident Voltage set versus Greedy

Another comparison we can make is observing how the Incident Voltage set scores versus nodes chosen by a greedy algorithm. The greedy algorithm is as follows. For a given true state $\mathbf{V}$ we calculate which node produces the best mean score among ten different $\mathbf{V}'$, i.e. we set $|\mathcal{O}| = 1$ and find the best individual measuring locations.

Definition 5.3.1. *For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, a budget $B \in \{0, \ldots, |\mathcal{V}|\}$ and Greedy score G we can sort the score from best to worse: $\{gr_1, \ldots, gr_{|\mathcal{V}|}\}$. The Greedy set for budget B is then $\{gr_1, \ldots, gr_B\}$, the first B nodes of the sorted vector.*

For $B = 1, \ldots, |\mathcal{V}|$ we compare the Incident Voltage set[1] to the Greedy set. This was done for a single graph of 1000 nodes and 35 routes and ten different true states. Given one of the true states, the comparison is visualized in Figure 5.8. In Figure 5.9 the relative distance is plotted. This distance is calculated by

$$\text{relative distance} := \frac{\text{score Greedy} - \text{score Incident Voltage}}{\text{original } 2\text{-norm}}.$$

[1] See Definition 5.1.1.

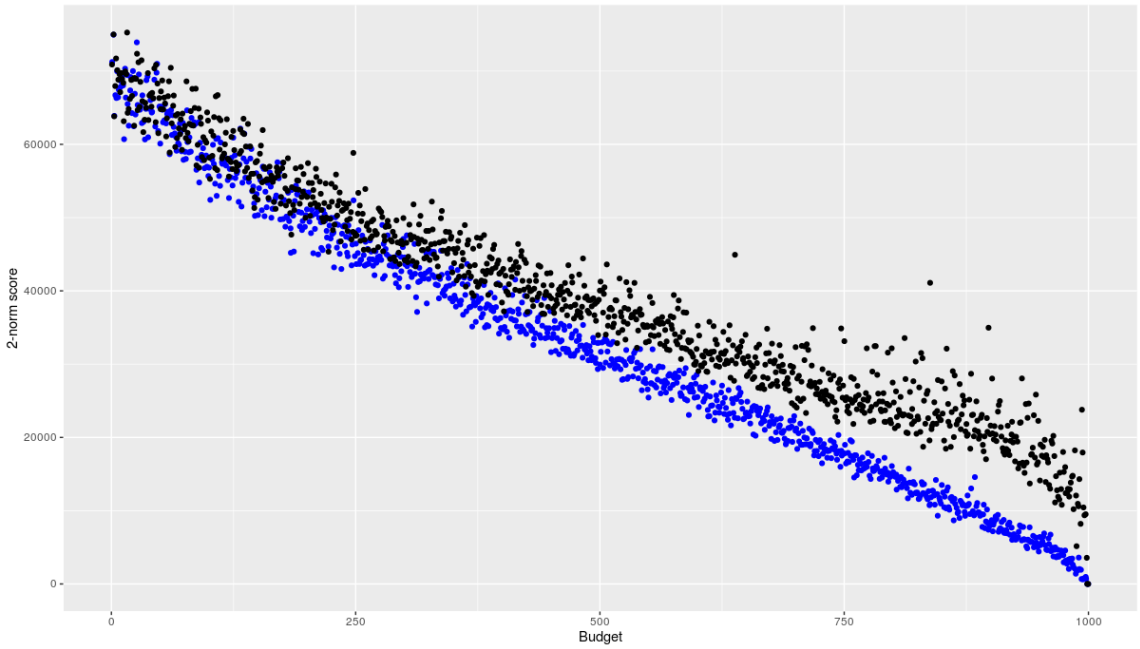

Figure 5.8: For a given **V** the score of the Incident Voltage set (blue) and the Greedy set (black) is plotted versus the budget.

experiment	1	2	3	4
score	2−norm	2−norm	2−norm	2−norm
V generation	Outward	Outward	GEV	GEV
Forecasting	GEV	Normal	Normal	GEV

Table 5.3: The four case studies observed

The same plots can be made by considering all ten experiments at the same time, grouping for the different experiments, i.e. different **V**. This relative distance can be found in Figure 5.10. These plots show that the Incident Voltage sets are better in this scenario then Greedy sets. We used Normal distribution forecasting and GEV **V** generation. This process was done for one graph and four different case studies, as displayed in Table 5.3. The other case studies are displayed in Figures 5.11,5.12 and 5.13. We see that the GEV distribution causes more extreme points, as expected. In the case where both the true and predicted states adhere to a GEV distribution, (Figure 5.13) there is no statistical difference. In the case of Figure 5.12 we see that Greedy is only worse after a large amount (300+) of locations are measured.

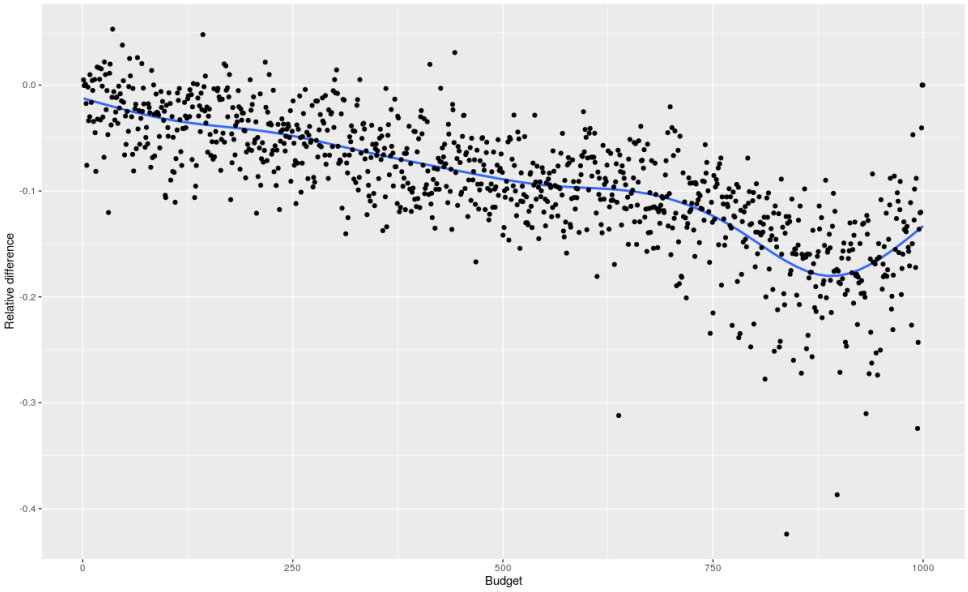

Figure 5.9: The relative difference for different budgets. In blue a function is fitted between the observations using *geom_ smooth* in Rstudio.

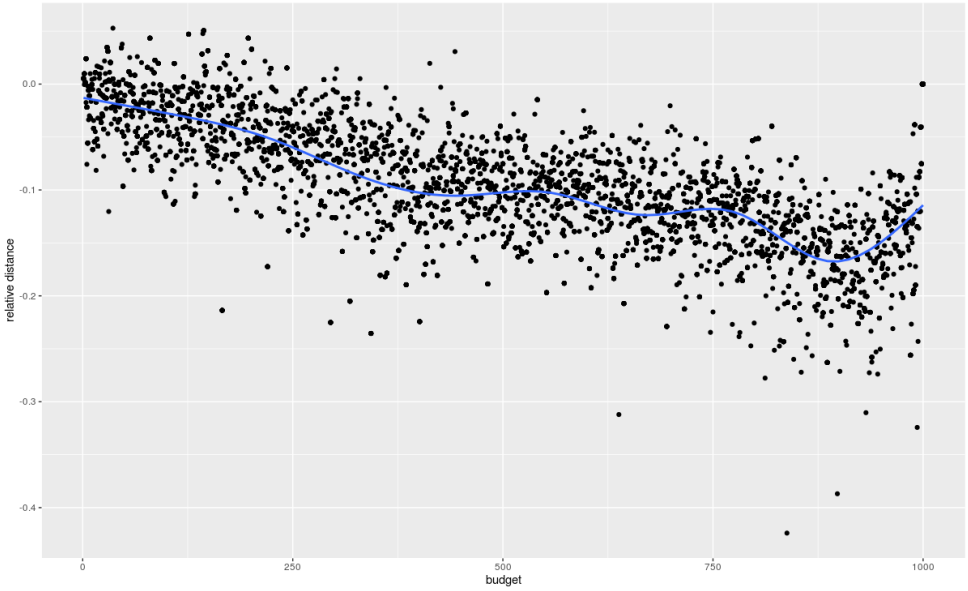

Figure 5.10: The relative difference for all experiments.

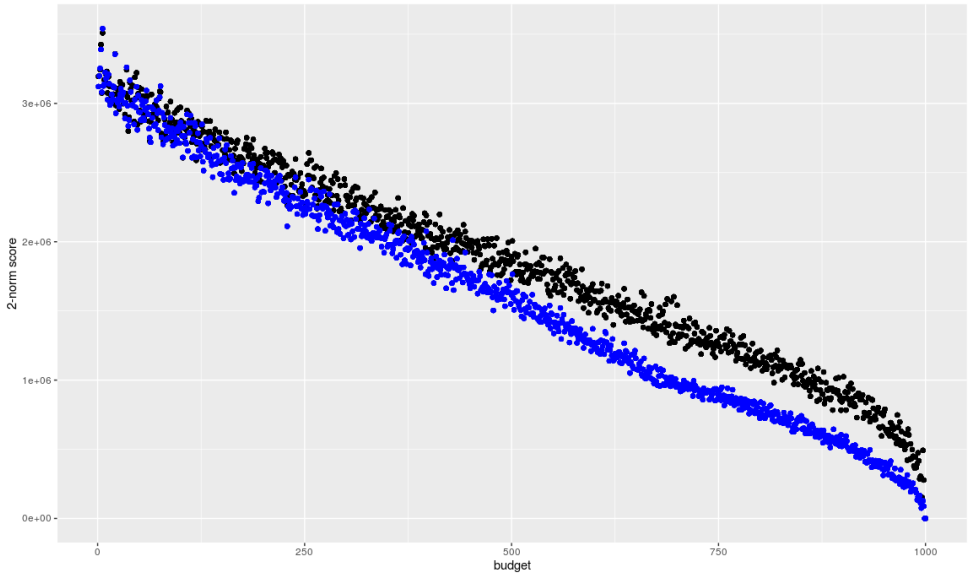

Figure 5.11: Greedy versus Incident Voltage, **V** generation is Outward, forecasting is Normal distributed.

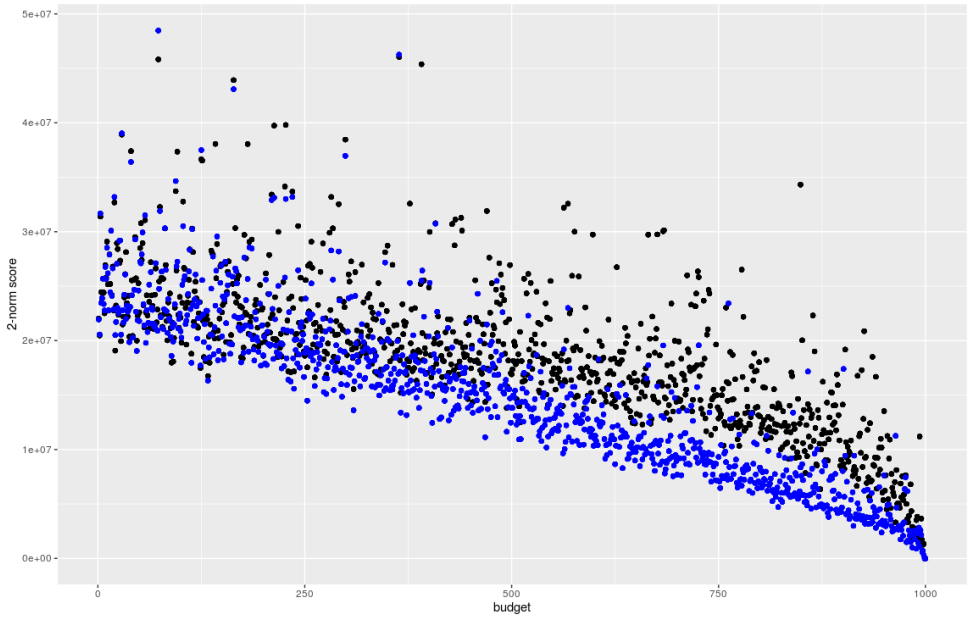

Figure 5.12: Greedy versus Incident Voltage, **V** generation is Outward, forecasting is GEV distributed.

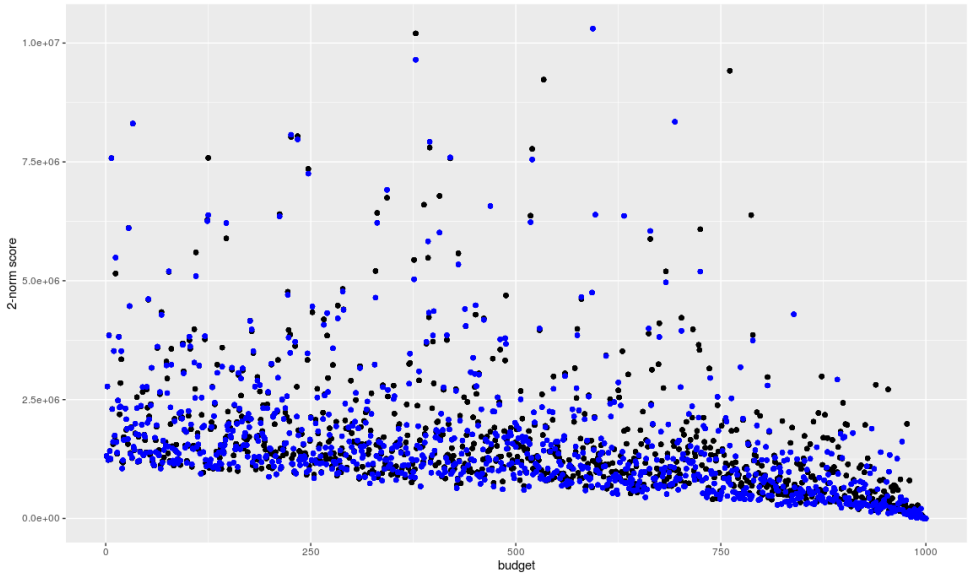

Figure 5.13: Greedy versus Incident Voltage, $\mathbf{V}$ generation is GEV, forecasting is GEV distributed.

Chapter 6

Discussion

6.1 Key findings

In this thesis we proposed a new scoring mechanism to evaluate which locations in a power distribution network should be measured. It does so by providing a forecaster $\mathcal{F}$ with voltage measuring data post priori, i.e. after the forecast was made. When the voltage prediction is altered, the nodes are chosen as to minimize the Incident Voltage bound (3.3.17) which bounds the $2-$norm score.

The proposed score, called the Incident Voltage score, is independent of $\mathcal{F}$ and has been tested on four large networks using different methods regarding generating true states $\mathbf{V}$ and predicted states $\mathbf{V}'$.

In Chapter 5 we showed that the Incident Voltage score is consistently better than considering random samples from the subset $\mathcal{W} \subseteq \mathcal{V}$ with $\mathcal{W} = \{m \mid \deg(m) \text{ is high}\}$, prioritizing nodes with the highest possible degrees. We showed that in most cases

$$per := \frac{\#\text{samples that scored better then } \mathcal{IV}}{\#\text{samples}}$$

equals either zero or one. This means that the Incident Voltage score was either better than all samples, or worse then all samples. Recall Table 5.2, which presents all means of per, given different methods for true and predicted state generation. We saw that if the true states were taken from the generalized extreme value (GEV) distribution, the score preformed best. This means that if voltage states exhibit more extreme values, i.e. moments in time where voltage values spike at some locations in the network, the score is higher recommended as to when this is not the case. Table 5.2 seems to show no indication that there is a difference in GEV and normally distributed prediction errors. We can however infer that the GEV distribution produces more variation in the results mean(per), as to be expected from a distribution known for its extreme points. As such the Incident Voltage score is an significant improvement compared to (randomly) choosing measuring locations with a high degree, given that accurate profiles of the true states $\mathbf{V}$ can be produces by the user.

6.2 Limitations

There are some limitations to the Incident Voltage score and the general method. As previously stated, the Incident Voltages score indicates nodes that consistently score better than picking randomly from the set of nodes with a high degree.

In terms of absolute score, it is only a few percent better than the average taken over a hundred random samples. Recall Figure 6.1 which was also used in Chapter 5. It is an example of the spread in scores given a hundred different samples sets $\mathcal{O}$. The mean of these samples lies around 887500, while the Incident Voltages scores generates a set which has a corresponding score of around 883750. We calculate $100 \cdot \frac{883750}{887500} \approx 99.58\%$, meaning it is less than a 0.5% improvement. Having calculated this value for all the case studies in Chapter 5 this percentages varies between 0% and 5%, with a larger spread if the corresponding forecaster and style of generating true states exhibit a larger spread. A worst case scenario has been detailed in Figure 6.2, which shows the improvement can even be negative at times.

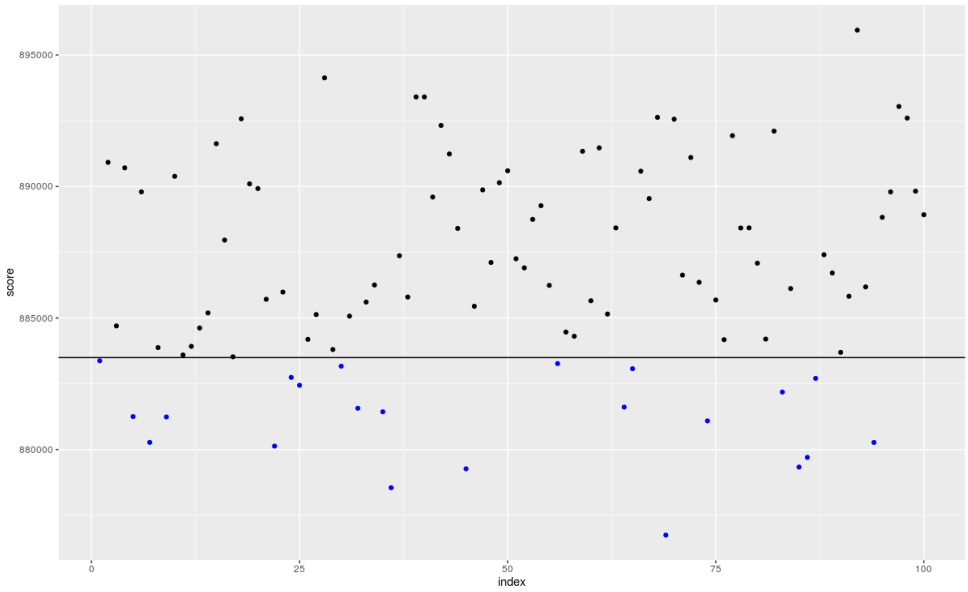

Figure 6.1: The 2−norm score $\|\mathbf{s} - \mathbf{s}'\|_2$ is plotted for 100 different measured sets $\mathcal{O}$ with $|\mathcal{O}| = 20$. In blue are all $\mathcal{O}$ that scored better then the set generated by the Incident Voltage score, which is indicated by the horizontal line.

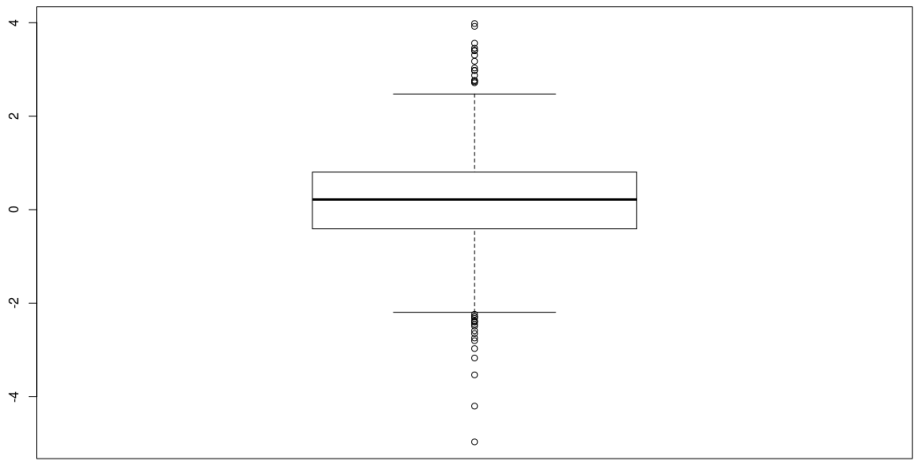

Figure 6.2: A boxplot detailing a worst-case scenario for the improved percentage over a set of 15 different true states and 100 corresponding predicted states for each true state.

Additionally, we can observe how changing the budget restraint influences the value of the $2-$norm score for given true state $(\mathbf{V}, \mathbf{s})$ and predicted state $(\mathbf{V}', \mathbf{s}')$. As expected, plotting the $2-$norm as a function of the budget is an almost strictly decreasing function that equals zero when all nodes are measured. In Figure 6.3 we see this decreasing function for a given pair of true and predicted states. It was normalized because only the relative progression is interesting here, not the initial accuracy of the predicted state. In Figure 6.4 we see the same plot for 25 different pairs, showing the range of possibilities when it comes to the size of the gaps, and the locations with respect to the $x-$axis (budget). We see that large jumps in accuracy are possible and occur often. Determining where those gaps are is most likely dependent on which nodes are predicted the least accurate, meaning we have no control over locating them.

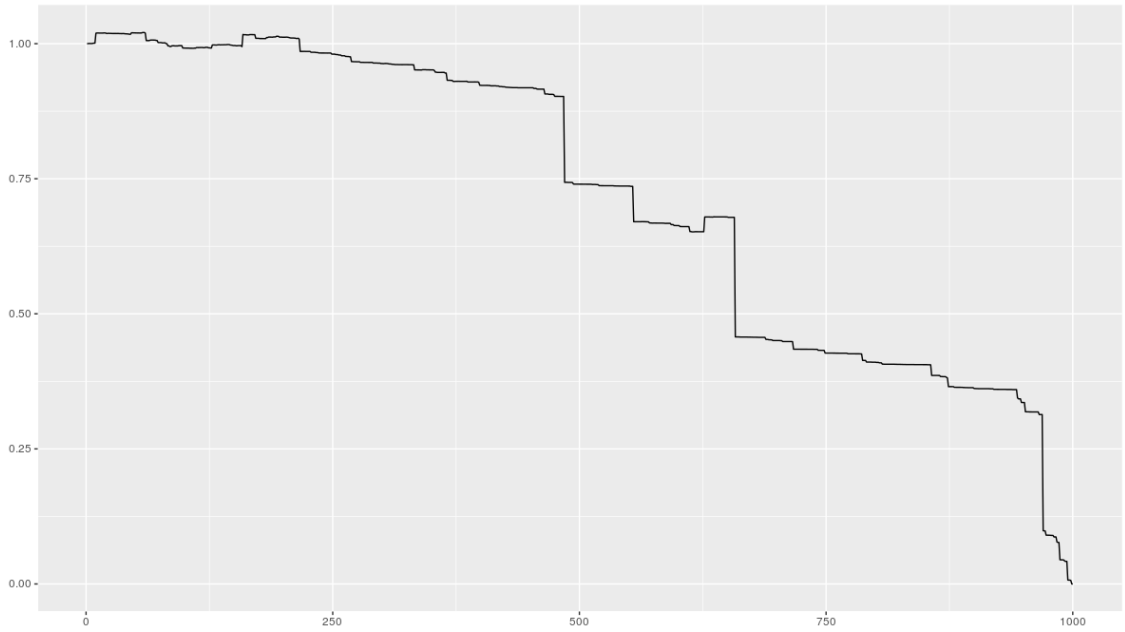

Figure 6.3: The progression in score ($y-$axis) versus budget ($x-$axis) for a predicted state given a true state.

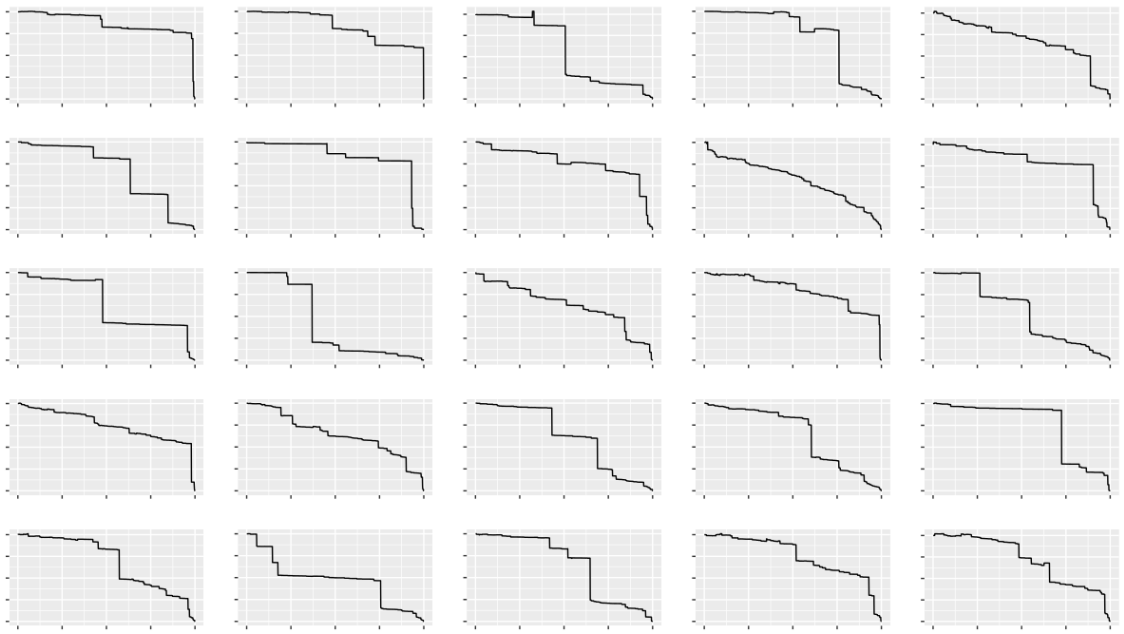

Figure 6.4: The progression in score ($y-$axis) versus budget ($x-$axis) for 25 different predicted states given a true state.

Lastly we briefly discuss realism. How accurately does our model reflect the real world? At the start of this project multiple assumptions were made to make life easier for me, the writer. This was done because my expertise lies in theoretical mathematics, and not in electrical engineering. As such, there is still room for improvement when it comes to portraying the world accurately within the model. Most, if not all, of this thesis comes from a theoretical point of view. Within the model I am convinced of the results, and hoped to have convinced you as well. But there are still some possibilities to represent the physical world more realistically. In Chapter 8 we discuss possibilities for future work.

Chapter 7

Conclusion

During this research, the Incident Voltage score has been introduced and the following questions have been answered:

1. how is scoring single valued forecasters considered viewed Decision theory?
2. how can we describe voltage predictions and calculating nodal power from nodal voltages within linear algebra?
3. how can the $2-$norm for a given true and predicted state be reformulated using this mathematical model?
4. how accurate is the Incident Voltage score at estimating which nodes improve upon the $2-$norm score?

At the start of this thesis we globally explained the underlying problem and laid out a general plan for solving the sensor problem. After that we introduced a method to represent the difference $\mathbf{s}-\mathbf{s}'$ as a function of $\mathbf{V}$ and $\mathbf{V}'$ using linear algebra. Using that representation we provided an upper bound which we called the Incident Voltage bound that generated the Incident Voltage score. This score is dependent on the full and partial incident matrices A and B respectively, and needs a profile of common voltage vectors to preform adequately. The score for a single voltage vector $\mathbf{V}$ at node k reads as follows:

$$\mathcal{IV}(\mathbf{V})_k \coloneqq |A_{(\cdot,e_k)}\mathbf{V}| + \sqrt{\sum_{e\in\mathcal{A}:\ e \text{ incident to } k} |B_{(\cdot,e)}\mathbf{V}|^2},$$

where e_k is the unique arrow with node k as its sink.

The size of the voltage profile is dependent on the variation in possible voltages, as discussed in Section 4.2. When a profile has been created, calculating each $\mathcal{IV}(\mathbf{V})$ will show the variance and then the user can determine if the variance is satisfactory. If not, the profile should be dissected into different profiles so that each profile has a lower variance and represent a certain type of state. This type can be the moment of the day, season, temperature or any other useful distinction.

In Chapter 4 the Incident Voltage upper bound was proven and its properties were discussed. In Chapter 5 we discussed some case studies, provided we simulated the forecaster by a random

process. We showed under which assumptions the Incident Voltage score can significantly improve the 2−norm score and the MSE-score.

When no historical data is available, we can generate data by statistical processes. These processes where explored in Chapter 5 and provided these represent a typical day, we showed that the Incident Voltage score is significantly better then the average sample, even is the sample range was already modified to only chose nodes with high degree. Moreover, we showed that the score works best if large voltage spikes are common and occur everywhere in the network.

Chapter 8

Future work

There are a few aspects that I was unable to fit in this thesis, mainly due to inexperience in the field at the start of this thesis. In this chapter we briefly discuss the subject which can be expanded on to further solidify the usefulness of the Incidence Voltage bound. Most of what has been done in this thesis still lives within the theoretical world, as some assumptions where made to simplify the process. Now that everything is finished I can see how we can generalize certain aspects of the thesis so that the process becomes more realistic. We discuss non-diagonal impedances, local influence of measured nodes, measuring current instead of voltage and a priory measuring.

We start with generalization that impedances are non diagonal matrices. In practice that means that there is interference between the three phases of the system. We used our assumption that each impedance is diagonal to simplify the system from

$$\begin{aligned} \mathbf{ZI} &= A^t\mathbf{V}, \\ \mathbf{S} &= (B^t\mathbf{V}) \circ \mathbf{I}^*, \\ \mathbf{s} &= -A\mathbf{S}, \end{aligned}$$

to

$$\begin{aligned} \mathbf{ZI} &= A^t\mathbf{V}, \\ \mathbf{S} &= (B^t\mathbf{V})\mathbf{I}^*, \\ \mathbf{s} &= -A\mathbf{S}. \end{aligned}$$

Note that the second set of equations is calculated for each phase individually, which can be done so because each $\mathbf{Z}$ is a now a diagonal matrix.

When we do not assume each impedance to be diagonal, then $\mathbf{Z}$ becomes a block diagonal matrix. The system of equations would have to be reformulated. There is one other instance where we call upon $\mathbf{Z}$ being a diagonal matrix. We bound the $2-$norm by the Incident Voltage norm and Theorem 4.1.5, we isolate the norm of $\mathbf{Z}$ using Lemma 4.1.4, which states that the $2-$norm of a diagonal matrix is its largest entry. A similar result is true for block diagonal matrices. In general the $2-$norm of a matrix is its largest eigenvalue. Let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ be a graph with $\mathbf{Z}$ the block diagonal matrix of $\{\mathbf{Z}_e \mid e \in \mathcal{A}\}$. Then

$$\det \mathbf{Z} = \prod_{e \in \mathcal{A}} \mathbf{Z}_e,$$

so consequently we have for the set of eigenvalues $\sigma(\mathbf{Z})$ that

$$\sigma(\mathbf{Z}) = \bigcup_{e \in \mathcal{A}} \sigma(\mathbf{Z}_e),$$

and so

$$\|\mathbf{Z}\|_2 = \max\{\lambda \mid \lambda \in \bigcup_{e \in \mathcal{A}} \sigma(\mathbf{Z}_e)\}.$$

In short, this means that the 2−norm of $\mathbf{Z}$ is the largest eigenvalue in $\bigcup_e \sigma\mathbf{Z}_e$, where each set of eigenvalues can quickly be calculated by the characteristic polynomial, which is only of degree 3 as $\mathbf{Z}_e \in \mathbb{C}^{3\times 3}$ for each $e \in \mathcal{A}$.

This thesis does not consider the local influence of a measured node. There is a possibility that a measured node bounds the possible voltages of neighbouring nodes. For instance, If we measure nodes 2 and 5 as in Figure 8.1, then we might be able in infer the voltage of node 3. In the code written to measure nodes $\mathcal{O}$, each $\mathbf{V}'_n$ changes to $\mathbf{V}_n$ if $n \in \mathcal{O}$. We can change this part of the code to also account for these possible local influence, for instance by lowering the distance $|\mathbf{V}'_m - \mathbf{V}_m|$ if m is a neighbour of n, though not lowering it completely to zero. This way we can account for possible local influences when checking the case studies of Chapter 5.

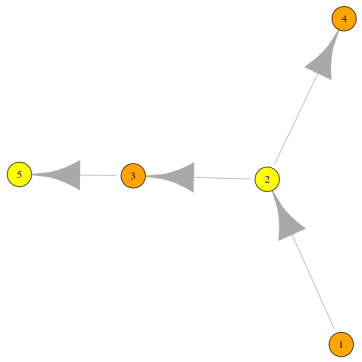

Figure 8.1: A small graph with nodes 2 and 5 measured.

To calculate the nodal powers by

$$\begin{aligned} \mathbf{ZI} &= A^t\mathbf{V}, \\ \mathbf{S} &= (B^t\mathbf{V}) \circ \mathbf{I}^*, \\ \mathbf{s} &= -A\mathbf{S}, \end{aligned}$$

you need two of the three variables $\mathbf{Z}, \mathbf{V}$ and $\mathbf{I}$. Resistances can be assumed to be known, so we can either solve for $\mathbf{V}$ or for $\mathbf{I}$, both of which are possible. When I started this thesis I had planned to develop a method for both measuring $\mathbf{V}$ and measuring $\mathbf{I}$. The latter requires calculating $\mathbf{V}$

from $\mathbf{I}$ which can be done if the voltage of the substation is assumed to be known. If the voltage of the substation is known we can recursively calculate all $\mathbf{V}_n$, starting with the children of the substation. I never got this function working, and due to time restraints decided to forgo this part of the thesis. A bound similar to $\mathcal{IV}(\mathbf{V})$ might be possible when measuring $\mathbf{I}$ and my code can be use to check the case studies described in Chapter 5.

Another possible improvement is weighing nodes based on priorities that are not taken into account by the model. A possible priority would be congestion areas where cables are overloaded and require more attention. This can easily be done if the user knows what weights they want to use. Using weighted incidence matrices for weighted graphs [15] as input for the Incidence Voltage score will provide the desired result. Though this should be used with cation. Adding relatively large weight will quickly overpower the score, rendering it useless. More research is needed.

Let us also discuss a priory measuring. In this thesis the definition of measured was such that a node was considered measured if any prediction of that node would equal the actual voltage. This means nodes are measured post priori. We let a forecaster F predict each node and afterwards set $\mathbf{V'}_n = \mathbf{V}_n$ if n is measured. This does not take into consideration that a forecaster might improve if it is handed more accurate information, i.e. a set of nodes that are already measured beforehand. The methods used and the Incident Voltage score found in this thesis might be applicable to a priori measuring, though rigorous testing would be needed.

All the aforementioned points are still theoretical. The most important future work would be using an actual forecaster and not simulate the predicted states $\mathbf{V'}$, and see if the results obtained in Chapter 5 can be reproduced. The other points are there to explore other options (measuring currents) or to make our theory more realistic. Testing on one (or more) actual forecasters would be the final test.

Bibliography

[1] Jeffrey D. Taft and Paul De Martini, *Sensing and Measurement for Advanced Power Grids*

[2] Ananth Narayan Samudrala, M. Hadi Amini, Soummya Kar and Rick S. Blum, *Optimal Sensor Placement for Topology Identification in Smart Power Grids*

[3] Roel Dobbe, Werner van Westering, Stephan Liu and Daniel Arnold, *Linear Single- and Three-Phase Voltage Forecasting and Bayesian State Estimation with Limited Sensing*

[4] Siddharth Bhela, Vassilis Kekatos, Liang Zhang, and Sriharsha Veeramachaneni, *Enhancing Observibility in Power Distribtion Grids*

[5] Francesco Fusco and Jonas C. Villumsen, *Sensor Placement for Optimal Estimation in Power Distribution Grids*

[6] Xiao Li, Anna Scaglione and Tsung-Hui Chang, *A Framework for Phasor Measurement Placement in Hybrid State Estimation via Gauss-Newton*

[7] A. Phadke, J. Thorp, and K. Karimi, *State Estimation with Phasor Measurements*

[8] A. Meliopoulos, G. Cokkinides, C. Hedrington, and T. Conrad, *The Supercalibrator-A Fully Distributed State Estimator*

[9] R. Nuqui and A. Phadke, *Hybrid Linear State Estimation Utilizing Synchronized Phasor Measurements*

[10] S. Chakrabarti, E. Kyriakides, G. Ledwich, and A. Ghosh, *A Comparative Study of the Methods of Inclusion of PMU Current Phasor Measurements in a Hybrid State Estimator*

[11] R. Avila-Rosales, M. Rice, J. Giri, L. Beard, and F. Galvan, *Recent Experience with a Hybrid SCADA/PMU Online State Estimator*

[12] Xiao Li, Anna Scaglione and Tsung-Hui Chang, *Optimal Sensor Placement for Hybrid State Estimation in Smart Grid*

[13] Qiao Li, Rohit Negi and Marija D, *Phasor measurement units placement for power system state estimation: A greedy approach*

[14] Giovanni Parmigiani and Lurdes Inoue, *Decision Theory, Principles and Approaches*

[15] Jean H. Gallier, *Graphs and Graph Laplacians*. https://www.cis.upenn.edu/ cis515/cis515-14-graphlap.pdf