NVE
Reguleringsmyndigheten
for energi – RME

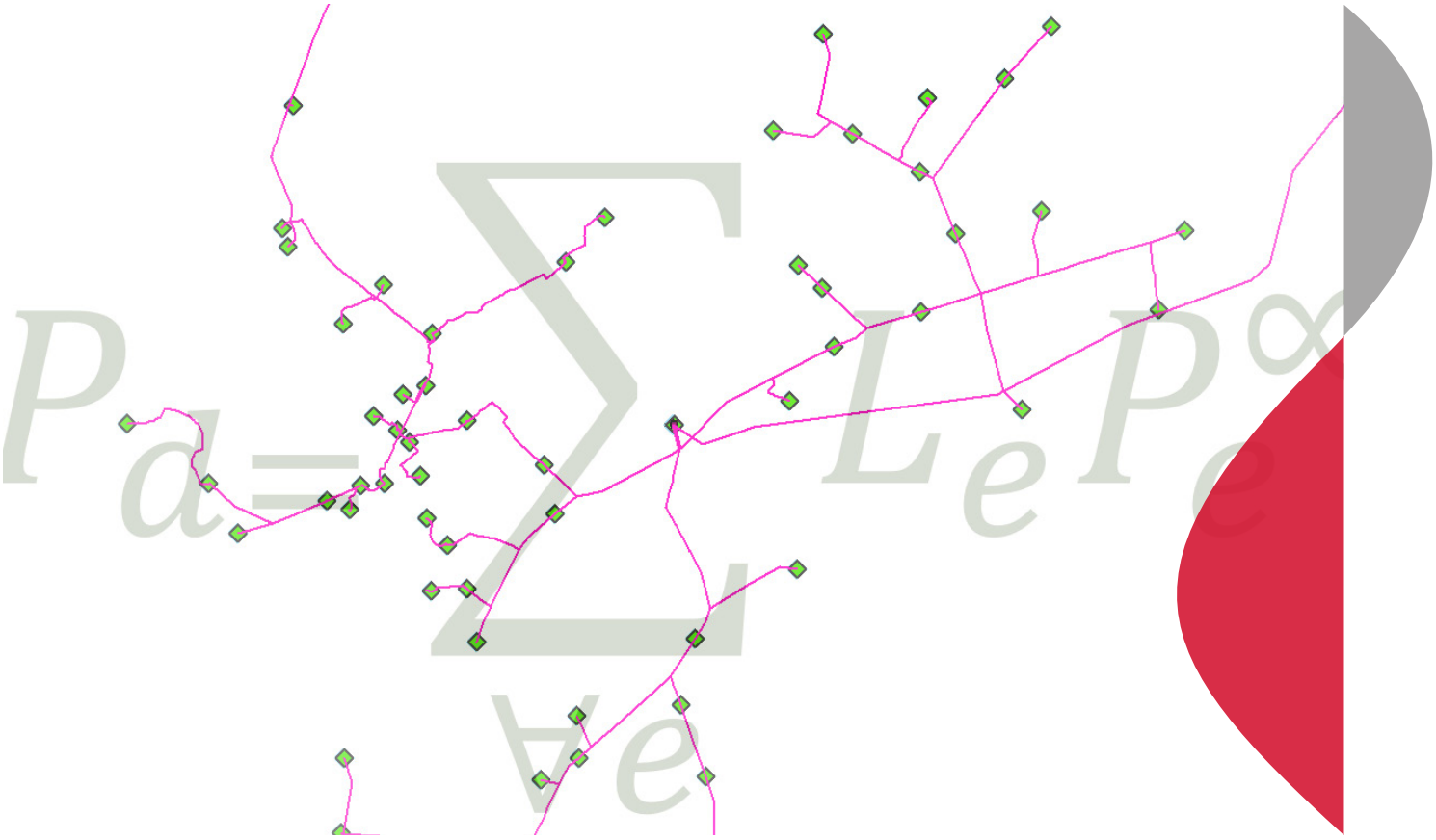
RME EKSTERN RAPPORT

Nr 01/2019

Power distance as an output parameter for grid companies

Formulation, application and comparison of minimal, power-flow based and grid-free power distance parameters

THEMA Consulting Group



RME Ekstern rapport nr 1-2019

Power distance as an output parameter for grid companies

Utgitt av: Norges vassdrags- og energidirektorat
Redaktør: Tore Langset
Forfatter: THEMA Consulting Group

Trykk: NVEs hustrykkeri
Forsidefoto: Ole-Petter Kordahl
ISBN: 978-82-410-1962-3
ISSN: 2535-8243 (online)

Sammendrag: Denne rapporten utforsker metoder for å beregne nye oppgavevariabler som kan brukes for å sammenligne nettselskapene. THEMA Consulting Group har utviklet tilnærminger for å beregne variablene effekt- og energiavstand. Metodene er testet ut på små datasett fra enkelte nettselskaper.

Emneord: Effektdistanse, energidistanse, power-flow, sammenlignende analyser, inntektsregulering.

Norges vassdrags- og energidirektorat
Middelthunsgate 29
Postboks 5091 Majorstua
0301 OSLO

Telefon: 22 95 95 95
Epost: nve@nve.no
Internett: www.nve.no

Preface

The Norwegian Energy Regulatory Authority uses a DEA-model when determining the allowed revenue of Norwegian electricity distribution operators (DSOs). The DEA-model calculates each DSOs relative efficiency by comparing all DSOs input/output ratios. The input is a measure of the DSOs' costs related to their tasks. The outputs are substitutes for the different tasks the DSOs have to solve when building, maintaining and operating the grid. The measured efficiency has a direct influence on the returns of the DSO.

The task of the DSO is changing due to increased use and dependency of electricity. Electrification is a prerequisite for achieving the climate goals and the DSOs play a key role in this transition. It is important that the output parameters are describing the task –or the cost drivers – of the DSO within this context.

One parameter that effectively captures the distribution of demand in a grid system is the power distance. The power distance is the product of the distance to each substation and the transferred power, scaled by a factor that reflects how the investment cost of power lines increases with higher capacity. This parameter takes into account that customers are different in their demand and in their location to the substations.

In this report, THEMA Consulting Group (THEMA) investigates power distance parameters for the DEA-model. THEMA introduces four methodologies to calculate power distance and applies them to a selection of test cases. The methodologies have different data requirements and varies in terms of computational complexity. THEMA discusses the results for hourly power distances and annually aggregated energy distance parameters in the context of applicability and regulatory implications.

We will use the findings from this report to launch a follow-up project in 2020. This project will investigate in more detail the pros and cons of the proposed methods. In this project, we will also evaluate the applicability of the power distance methods in larger grid-systems and ultimately the entire Norwegian high-voltage distribution grid, and conclude on which parameters are best suited for the purpose.

The findings, analysis and recommendations of this report are those of THEMA and do not necessarily reflect our position.

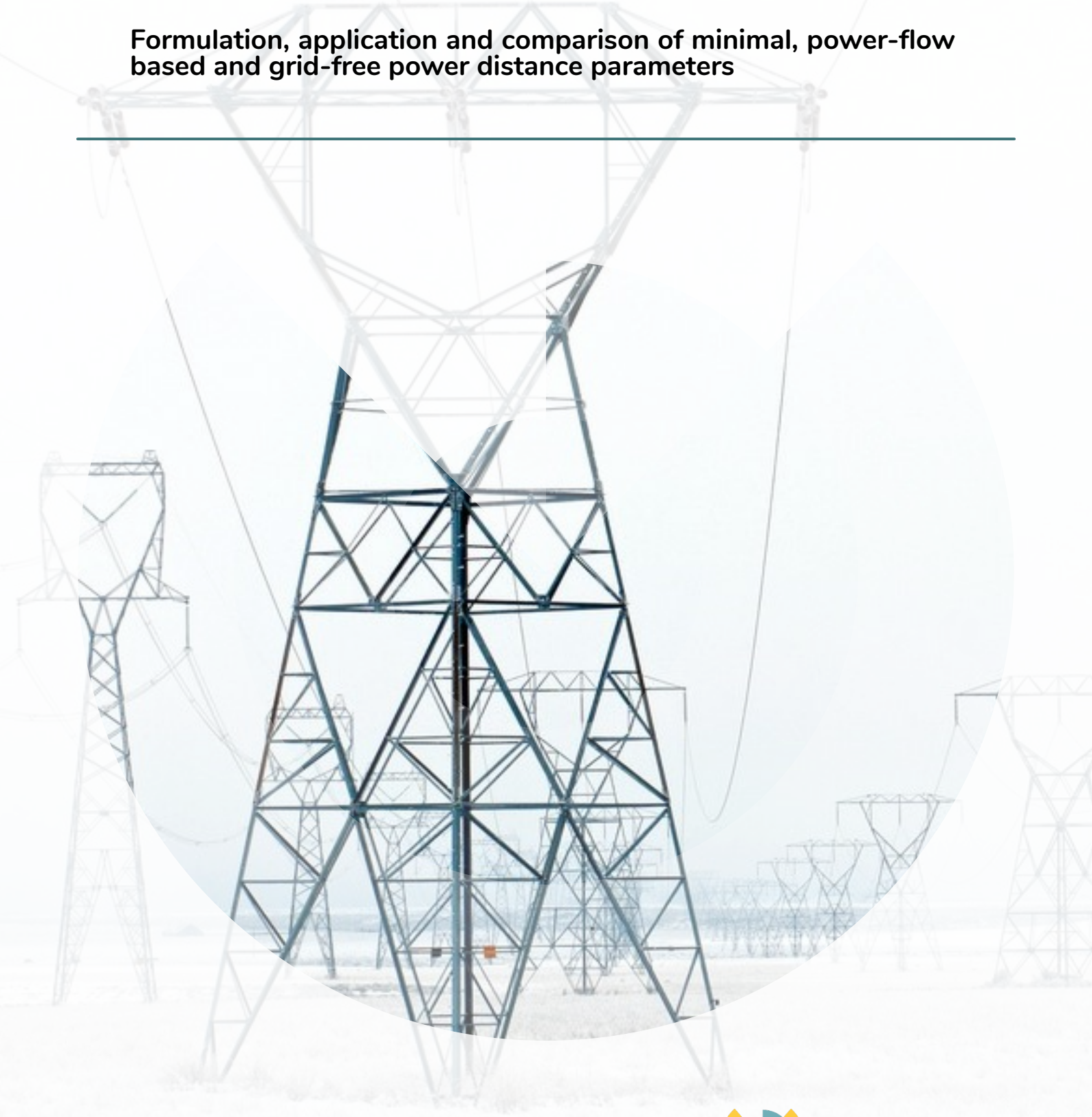
Oslo, December 2019

Ove Flataker
Director,
Energy Regulatory Authority

Tore Langset
Head of Section,
Section for Economic Regulation

Power distance as an output parameter for grid companies

Formulation, application and comparison of minimal, power-flow based and grid-free power distance parameters



THEMA
CONSULTING GROUP

Project info

Client
RME

Project number
NVE-19-01

Project title
Power distance as an output parameter for grid companies

Report info

Report availability
public

THEMA report number
2019-18

ISBN number
978-82-8368-059-1

Date of publication
25th November 2019

About the project

This report investigates alternative output parameters for the DEA model in Norwegian DSO income regulation. The power distance parameter is introduced as a variable that reflects the task of grid companies by capturing the geographical distribution of demand in the high voltage distribution grid. Four methodologies to calculate a power distance, with different requirements on data quality and computational complexity, are introduced and applied to a selection of test cases. The results for hourly power distances and annually aggregated energy distance parameters are discussed in the context of applicability and regulatory implications.

Project team

Project manager
Theodor Borsche, Åsmund Jenssen

Contributors (alphabetically)
Kristin Arnesen
Kristine Fiksen
Lisa Zafoschnig

Executive Summary

Reguleringsmyndigheten for Energi (RME) is the Norwegian regulatory authority for energy. Among its responsibilities is the regulation of income for Distribution System Operators (DSOs) and the resulting grid tariffs. In this context, Norway was among the pioneers of introducing a performance benchmarking model to compare the efficiency of grid companies. The so-called Data Envelopment Analysis (DEA) benchmarking model evaluates the performance of each grid company based on a number of indicators and determines the allowed income based on the resulting relative efficiency and the company's actual annual cost. Currently the output parameters of the benchmarking model are the number of customers, the total length of lines and the number of substations in the high-voltage distribution grid. As more data is made available from smart-meters and centralised infrastructure databases, new output parameters for the benchmarking process can be considered. Ideally, such parameters should represent the task of the DSO (not its effort) and provide incentives for efficient grid reinforcement while being highly exogenous, comparable and easy to compute from available data.

The power distance as an output parameter

THEMA Consulting Gorup has been commissioned by RME to investigate different methods to define and compute new output parameters for the DEA model that take the distribution of demand into account and thereby provide a more objective representation of the task of a grid company. A parameter that effectively captures the distribution of demand in a grid system is the *power distance*. The power distance parameter is the product of the distance to each substation and transferred power, scaled by a factor that reflects how the investment cost of power lines increases with higher capacity. The distance to each substation can be defined through the real lines in the grid, a synthetically created grid or a geographical length.

We investigate four methodologies for computing a power distance parameter:

minimal power distance: The power distance in the minimal radial grid configuration based on the existing grid.

power-flow based power distance: The power distance based on physically optimal flows in the existing grid where all line segments have the same specifications.

artificial grid based power distance: A power distance computed for an artificially constructed grid that connects all substations based on the minimal increase in power distance.

demand distribution based power distance: A power distance obtained from the statistical distribution of demand around each transformer station without considering a grid.

Analysis of the power distance based on real grid areas

To analyse the computation methods on real distribution grid systems, six Norwegian DSOs provided power demand and generation data for selected grid areas. The metering data, aggregated per substation, were combined with data from RME's extensive database on grid infrastructure to obtain a complete picture of each test case.

The provided data was generally of high quality though some manual adjustments were required to compile the grid data and match demand data to

the substations. For this project most challenges related to data availability could be overcome, however, in a regulatory setting data quality is a crucial factor as any manual handling will affect comparability and may increase administrative costs for the regulator.

Minimal power distance

Due to the high computational complexity the **minimal power distance** could not be computed for the provided test cases. A real grid system is typically characterised by a large number of meshes, allowing power to flow to one substation over different paths. The optimisation problem to compute the minimal power distance would have to iterate over all possible combinations of removing line segments until a radial grid remains - a task that was shown to be infeasible for real grid systems.

Power-flow based power distance

To compute the **power-flow based power distance** it was necessary to compile the infrastructure data per element into a complete grid system before matching the demand data to each substation. This process, which was implemented with the use of standardised algorithms, could be performed for all grid companies in the reference group except for one, for which the line segments where stored in a different format.

The resulting hourly power distance showed a linear relation to the demand in each node scaled by the cost-parameter. Differences in the ratio between the demand and the power flow based power distance are a result of the line lengths in the system. More urban areas result in a larger ratio of power flow based power distance to demand, whereas rural areas are characterised by longer average line length to each substation which leads to a lower ratio. This finding underlines that the power distance accounts for challenges in supplying demand in different grid topologies.

Artificial grid based power distances

The lower data requirements for the grid-free parameters – the artificial grid based power distance and the demand distribution based power distance – that only use the location and demand per substation as input, made it possible to compute the output for all test cases. For the **artificial grid based power distance** a synthetic radial grid is created based on an algorithm that connects substations according to the minimum increase in total power distance. The resulting power distance for each hour is computed from the power flow in the resulting radial grid.

Demand distribution based power distances

The **demand distribution based power distance** does not use a real or synthetic grid at all. Consequently, the requirements for data and computation time are low, which makes it comparably easy to compute the power distance for all test cases. Both grid-free power distance results showed a linear correlation to the power-flow based power distance, indicating that they can be used as suitable proxies. The ratios between the power distances obtained from different computation methods varied between test cases due to differences in topography in each area. While the power flow based power distance accounts for obstacles in the grid and other area specific limitations, such factors are not considered in the grid-free methods. In further work we suggest to consider potential geographical and topographical corrections that can be applied to the grid-free parameters to improve comparability.

Concluding from the results of the different computation methods, the **power-flow based power distance** is identified as the most suitable method to represent the task of a DSO, with improved exogeneity compared to the current output and high comparability between grid areas. At the moment, the use of grid data poses challenges that can impede its use in a regulatory setting.

Therefore, we suggest that the improvement of data quality and consistency should be a focus area for further work. Should it not be possible to achieve a sufficiently high data quality, we propose to apply the **artificial grid based power distance** with an ex-post geographical adjustment.

Any computation of a power distance in the high-voltage distribution grid needs to be coupled with a way of accounting for the demand in the low-voltage distribution grid. The location of customers around a substation could be accounted for by using a method similar to the demand distribution based power distance or initially through the number of substations.

Aggregation to an annual benchmarking parameter

To move from hourly power distance values to a comparable annual benchmarking output we investigate how cost drivers can be represented by different yearly aggregation methods. We refer to any annual parameter as an energy distance and apply the following methods:

1-norm: The sum or average over the full year – can be applied to the demand prior to computing a power distance or to the hourly power distance results. The 1-norm reflects the standard operation of a system and is therefore a proxy for the operational cost of a grid company.

2-norm: The sum of the squares of the difference between hourly values and mean – can be applied with hourly and mean demand or hourly and mean power distance. The 2-norm reflects variations in the system by emphasising hours of large peaks, thereby potentially reflecting the cost of losses.

∞-norm The maximum value in a system – can be applied to the total system demand, the maximum demand per substation/maximum flow per line or the respective power distances. The ∞-norm reflects the maximum load situation for which the grid needs to be equipped. It is a measure for the investment cost.

For the **1-norm** we conclude that using the average demand or the average power distance will not impact comparability between DSOs. Either of the measures will be a suitable parameter to reflect the operational costs of a grid company. For reasons of computational effort, we recommend to use the average system demand and compute the energy distance from it.

In the case of the **2-norm**, we investigate different approaches to correctly account for variations in the system. We conclude that the 2-norm should not be used in the DEA model as the cost of losses is already reflected by the computation of a power distance as transferred power and line length are accounted for.

For the situation of maximum load, we propose to use the maximum demand per substation or, in grid-based computations, the maximum flow per line, to compute a **∞-norm** energy distance. In large grid areas the probability of all customers experiencing their maximum load at the same time is lower than in a smaller area with less consumers. Due to these simultaneity effects, using the hour of maximum total demand or power distance will not affect grid areas of different sizes to the same extent and may underestimate the required investment.

In addition, we propose to keep the number of customers in the DEA benchmark to reflect the administrative costs that a grid company is facing.

Implications of the power distance in DSO income regulation

By using a power distance computation and performing different yearly aggregations, we can represent the main cost drivers in the distribution grid more effectively than the currently used DEA output parameters. Through the choice of the power distance parameter a compromise between the comparability of grid areas, the data requirements, the computational complexity and the exogeneity can

be found according to the main regulatory requirements. Future work on this topic should focus on investigating the applicability of the power distance methods in larger grid systems and ultimately the entire Norwegian high-voltage distribution grid to conclude which output parameters are best suited for the purpose.

Contents

1. Introduction **9**
1.1. The DEA benchmark model . . . 9
1.2. Other relevant models . . . 10
1.3. The power distance for DSO benchmarking . . . 12

2. Formulation and computation of power distance parameters **14**
2.1. General concept of power distance . . . 14
2.2. Minimal power distance . . . 16
2.3. Power-flow based power distance . . . 17
2.4. Artificial grid based power distance . . . 18
2.5. Demand-distribution based power distance . . . 21
2.6. Power distance in the low-voltage distribution system . . . 23
2.7. Alpha parameter . . . 23

3. From power distance to energy distance **27**
3.1. Cost drivers in the distribution grid . . . 27
3.2. From hourly values to a yearly parameter . . . 28
3.3. Input from the reference group . . . 33

4. Test cases **37**
4.1. Description of test cases . . . 37
4.2. Data quality . . . 40

5. Results per test case and per parameter **44**
5.1. Minimal power distance . . . 44
5.2. Power-flow based power distance . . . 44
5.3. Artificial grid power distance . . . 45
5.4. Demand distribution based power distance . . . 47
5.5. Comparison of yearly output parameters - energy distance . . . 48
5.6. Comparison of different alpha parameters . . . 50

6. Evaluation **54**
6.1. Discussion of key factors . . . 54
6.2. Feedback from the reference group . . . 59

7. Recommendations and conclusions **61**
7.1. Discussion of applicability and incentives per parameter . . . 61
7.2. Recommendations . . . 65
7.3. Outlook . . . 66

A. Acronyms **72**

B. References **73**

1. Introduction

Reguleringsmyndigheten for Energi (RME) is responsible for regulating Norwegian electricity network companies (Distribution System Operator (DSO)[1]). A key element of the regulation of distribution grids is RME's Data Envelopment Analysis (DEA) model. RME's DEA models are designed to benchmark the costs of a network company given a set of outputs that describe the tasks of the given company. In the distribution grid, the outputs are the number of customers, kilometres of lines and the number of substations[2] in the high-voltage distribution grid [3]as proxies for customer demand for each DSO.

RME is now exploring the possibility to replace the existing output measures by exogenous measures that better reflects the distribution of demand. A parameter that considers the transferred power and the distance to each demand node is the electric power distance as an hourly value, or the energy distance as an annual aggregation. In this report, THEMA provides a method for estimating measures of electric power distance and energy distance, as well as a discussion of the applicability of the proposed method. The main part of the report concerns the calculation of the power distance using different methodologies, and we also discuss how to move from the electric power distance to an annual measure of the energy distance as a benchmarking parameter.

[1]The Norwegian term *nettselskap* is translated to grid company or DSO in this report.

[2]In this report we use *substation* to refer to the Norwegian *nettstasjon*, i.e. a transformer station from the high-voltage (10 to 22 kV) distribution grid to the low-voltage (220 to 400 V) distribution grid.

[3]high-voltage refers to a voltage level of 1 to 22 kV in the Norwegian distribution grid. In this report we will use *high-voltage* and *low-voltage* (220 to 400 V) according to Norwegian grid standards.

1.1. The DEA benchmark model

The DEA model determines how much income each grid company can collect from their customers grid tariffs. The idea behind the design of the regulation is that each grid company incurs costs in solving their tasks. Costs are related to operational costs, capital costs, cost of losses, and CENS costs (Cost of Energy not Supplied, KILE in Norwegian). The task of the grid company is currently represented by number of customers, number of substations, and kilometres of lines, often referred to as the outputs.

As a sum, the costs incurred by all grid companies in the distributional level are covered (with the RME reference interest rate representing an expected rate of return), but the regulation aims to reward the most efficient companies with higher income, and thus create incentives for operational improvement and socioeconomic investments. The DEA model is applied to benchmark the companies against one another and determine which ones are the most efficient. The most efficient companies are called "front" companies.

In short, the average company is classified as 100 % efficient, while the most efficient companies are above 100 % and less efficient companies have an efficiency below 100 %. We will not describe the income regulation of Norwegian companies in detail in this report, but simply state how a company can collect its costs given its allowed income. A grid company's allowed income is given as

$$IR_t = (1-\rho)\cdot K_t + \rho \cdot K_t^*$$

Where K_t is the company's actual costs, K_t^* is the cost norm, and ρ is a factor defining to what extent the income of a grid company is benchmarked. If a company is an average company

(100 % efficient), $K_t = K_t^*$. If the company has an efficiency above 100 %, then $K_t \geq K_t^*$, and has a rate of return higher than the RME interest rate. A less efficient company will have a rate of return that is less than the RME interest rate – providing an incentive to become more efficient. Currently, ρ is set to 60 %, meaning that 40 % of the cost base can be directly passed on to consumers, while 60 % are based on the benchmarked cost norm.

As a grid company is evaluated on how cost efficiently it covers its tasks, it is important that the output parameters are describing the task – or the cost drivers – of the grid company in a relevant way. The task is to cover demand of all customers at all times, and the main costs drivers are investment costs (CAPEX), OPEX such as cost of delivering power and cost of losses, O&M costs and administrative costs.

1.2. Other relevant models

There is no standardised framework for income regulation of grid companies and, in general, each country has developed its own approach to determine the allowed income. Despite the large disparities between the existing frameworks some common methods and tools can be identified.

Centralised planning An approval for each new project is required before it can be added to the asset base, based on which the grid company is remunerated at a specified rate of return. Similar approaches are more common in the transmission grid where investments are typically larger and less frequent.

Cost plus In this traditional model for regulating monopoly revenues the network company is allowed to recover its actual costs including a return on invested capital. In a distribution grid the revenues would typically not depend on approval for individual investments but rather be based on the companies' reporting of their own costs, subject to regulatory approval of total spending and admissible costs.

Revenue caps A cap on revenues is set based on allowed costs per DSO plus additional factors. A common approach is the so-called RPI-X model where the annual regulated revenue may not be increased by more than the retail price index (RPI) minus a fixed adjustment factor (X). Another example is the RIIO (Incentives + Innovation + Output) approach used in the UK which includes allowances for innovation funding and incentive rewards for reliability, customer service and reduced losses.

Performance benchmarking The performance of each company relative to its task is benchmarked against other grid companies or against a synthetic grid to define a revenue cap. The allowed income is determined by a grid company's relative efficiency. The current DEA benchmarking model in Norway is an example of a performance based regulation.

In the context of the Norwegian DEA benchmarking and potential adjustments to its output, we consider it relevant to comment on performance benchmarking models from other countries. Two noteworthy approaches are the Reference Network Model (RNM) in Spain and the Network Performance Assessment Model (NPAM) in Sweden.

1.2.1. Spain

In Spain a methodology was developed to compare the real grid of DSOs against an idealised radial grid that is created by the so-called RNM. The framework, which is applied to all voltage levels in the distribution grid, determines the efficiency of each DSO individually relative to an ideal grid instead of comparing the performance of grid companies against one another.

The RNM is described as a large-scale planning tool which plans the electrical distribution network using the GPS coordinates and power of every single customer and distributed energy resource. In the process of creating an idealised grid the model considers both technical constraints – such as voltage limits and capacity constraints – and geographical limitations – such as topography or

water bodies. The cost of individual components in the idealised grid are based on a standardised equipment library.

There are two set-ups of the RNM which are used in the benchmarking process. A greenfield model that builds the entire grid system, including transformer stations, from scratch based on the distribution of demand points, and a brownfield model that considers extensions to the existing grid based on new demand or production points. The greenfield model is used to determine the efficiency of the existing grid relative to an idealised grid. While the grid from the greenfield model serves as an indicator as to which grid configuration would be best suited if all demand points were to be connected at the same time, it does not account for the historic development of the grid. The brownfield model is applied to determine which new investments in grid extensions and reinforcements are most efficient based on the existing grid. [1]

Both efficiency measures are considered in the annual benchmarking that is performed before the tariffs of the next year are determined. [2]

Beyond its use as a regulatory tool, the RNM is also used by DSOs for planning purposes and to evaluate the impact of the integration of distributed generation.

1.2.2. Sweden

In Sweden a benchmarking model based on the so-called NPAM was introduced in 2004 and later suspended. The NPAM was a newly developed model that compares existing grids to a standard asset. The aim was to evaluate the benefits to the consumer from an artificial radial network without excess capacity. In parallel to the development of the model a new legislative act was introduced to accommodate for the changes in regulation.

The NPAM creates a radial network that connects all nodes in a system on four voltage levels. Additionally spare capacity is considered to reflect the need for grid enforcements beyond a pure radial topology to ensure security of supply. The capital cost of the fictive network is referred to as the cost of connection which is based on standard cost functions per asset that take into account the physical properties of cables and transformers. The radial grid is also used to determine the cost of delivery, which describes the losses in the system and the cost of reliability, which is linked to the spare capacity. Other cost components in the model are the cost of administration, which scales with number of customers, and the cost of services, which are reported costs for connections to the transmission grid. The total cost from the NPAM is then compared to the revenues per DSO. Overall a DSO would be allowed to collect revenues that correspond to the customer value as calculated by the model. [3]

Despite the innovative approach with a focus on the DSOs task and the resulting customer benefit, the NPAM faced large challenges and was later suspended from the Swedish regulation. Generally, the complex model setup was not well documented and therefore hard to understand for anyone outside of the regulatory authority. Tests on the NPAM after its release also raised concerns on the robustness and fairness of the model. Geographical differences of few meters, within the uncertainty of data quality, resulted in entirely different synthetic networks and therefore different benchmarking premises for the same company.

It was also viewed critically that the model did not account for the historic development of the grid or geographical constraints. Inherited assets can seem inefficient compared to a synthetic grid, even though they were an efficient solution at the time of construction.

The main concern, however, was that the framework was designed as an ex-post model, meaning that the benchmarking was applied the following year after tariffs had been determined and paid. Results of the NPAM indicated that *"the Swedish network companies are overcharging their subscribers with approximately 20%"* [4]. This could lead to DSOs being obliged to pay back part of its revenue to the customers.

In 2008 the major criticism on the benchmark-

ing approach followed by the appeal of several DSOs and a lawsuit led to its suspension from the Swedish regulation.

1.3. The power distance for DSO benchmarking

The aim of this report is to define methodologies to calculate an alternative benchmarking parameter for DSO income regulation. The current benchmarking based on number of customers, substations and length of lines has two main draw-backs: it does not necessarily reflect the challenges of supplying power over long distances and the output is partly under control of the DSO through their investment decisions. As the consumer patterns in the distribution grid change through technological advancements such as distributed generation, local storage and EV charging infrastructure, a benchmarking parameter should be able to capture the resulting changes in the task of a DSO. Additionally, a higher quality of available data makes it possible to account for information that was not previously available. In [5] a so-called power distance parameter was suggested, that takes into account the demand and the distance over which power has to be transferred to each customer. In this report we will focus on defining and comparing different approaches to calculate such a power distance parameter. We further discuss how hourly power distance parameters can be aggregated to a yearly value - an energy distance.

The suitability of different power distance methodologies as benchmarking parameters depends on several factors. For a parameter to be an objective benchmark we deem the following factors to be most important:

Representation of the DSO's task The power distance should represent the task of the DSO rather than its effort. A DSO has the obligation to cover the demand of customers in the grid at all times. To fulfil this task a DSO needs to equip the grid accordingly to ensure security of supply and cost effectiveness.

Data Requirements In terms of needed input data for each methodology two factors are crucial: availability and ease of handling. The required data for a computation should be available and easily accessible for each DSO from centralised sources. Additionally, handling large amounts of data can be challenging both for those supplying it and those handling it in the final computation. With an increased need for data, the consistency and quality can be a further constraint.

Computational Effort The complexity of the calculation method can be a major limitation. Ideally, it should be possible to calculate the benchmarking parameter with low computational effort.

Exogeneity The chosen benchmarking parameter should be as neutral as possible. It should not be possible for the DSO to influence the output of the power distance calculation through their decisions on investments or operation. In the current regulation the length of lines is under direct control of the DSO which may incentivise building longer lines than necessary.

Incentives The choice of benchmarking parameter may create different incentives. An ideal parameter should not disincentivise any of the following: building the shortest/most efficient connections, investment in assets that improve security of supply, providing input data of high quality, among others.

Intuitiveness The method to compute a power distance should be fairly intuitive and easy to communicate to stakeholders. Any entity affected by the regulation should ideally be able to comprehend the computation steps and their implications.

While one methodology may give the most realistic representation of the task of a DSO, computation time may be a limitation and higher data

requirements may pose challenges towards exogeneity. On the other hand, a simple approach with minimal need for data and computational effort may incentivise unnecessary investments. The choice of benchmarking parameter will be based on a compromise between the criteria defined above. The individual implications for each of the proposed methods will be discussed in more detail later.

In this report we will introduce four methodologies to compute a power distance parameter. Each approach for computing hourly power distance values is described in detail in Section 2 before possible aggregations to an hourly benchmarking parameter are described in Section 3. The methodologies are tested on real grid data from six Norwegian DSOs and compared in Section 5. As a result, we can evaluate and compare the different methodologies and their implications in Section 6 before we conclude with final recommendations in Section 7.

2. Formulation and computation of power distance parameters

The power distance is identified as a parameter that captures the distribution of demand in a grid area, by including both the length over which power needs to be transported and the demand per node. In this chapter we describe the general concept and mathematical formulation of the power distance before providing a detailed overview of the different approaches to compute such a parameter. For simplicity, each computation method is introduced based on an exemplary test case.

2.1. General concept of power distance

The general idea behind the power distance is to have a more meaningful parameter describing the task of a grid company. Currently, the length of lines is one of the parameters used. Previously, the demand served over a year was an indicator as well, benefiting grid companies with large consumers close to transformer stations. The power distance improves on these two metrics by combining demand served and distance per consumer into the product of the two indicators. Transporting more power over a certain distance gives a higher power distance than transporting less power over the same line. In turn, serving a large load close by is considered much easier than serving the same load if it is further away. The power distance should constitute a more informative output parameter and thus lead to a fairer benchmarking of grid companies.

Mathematically, the power distance is defined as the sum of the products of the transported power and line length over all line elements. The general equation for the power distance P_d is

$$P_\mathrm{d} = \sum_{\forall e} L_e \, P_e^\alpha \quad , \tag{2.1}$$

with L_e the length of line element e, P_e the (absolute) flow along that line element, and $e \in \mathcal{E}$[1] the set of all lines owned by the grid company. The externally set parameter α describes the scaling effect of costs, i.e., a line able to carry twice the power is typically less than twice as expensive.

The main question related to the power distance parameter is how the flows, i.e. the transferred power, over each line P_i are determined and how the distance is defined. There are two main challenges: how to handle complex grid topologies? And how to compute the power distance if grid data is not available?

In a purely radial or tree-shaped grid, the flows are determined by the demand. While computing the power distance requires care, the result is uniquely determined. If, however, there are meshes in the topology, the power may be distributed to the consumer along different paths. In turn, the power distance is no longer uniquely determined. If grid data is altogether missing, one needs to pursue an alternative approach based on the location of the demand and possibly on local geographical limitations to define the distance. To address these challenges, we are proposing different algorithms to compute the power flow on grid lines, and hence arrive at different power distance parameters. These are the minimal power distance (Section 2.2), the power-flow based power distance (Section 2.3), the artificial grid power distance (Section 2.4), and the demand distribution based power distance (Sec-

[1] e as in edge in a graph

tion 2.5). All of these parameters have different requirements on data availability, computation time and exogeneity – meaning the extent to which the parameter can or cannot be influenced by the grid company.

For simplicity, all our algorithms aggregate the demand per substation (*nettstasjon*). Section 2.6 discusses how this could distort incentives against using 3-phase systems on the low voltage level. Section 2.7 discusses different approaches and results for estimating the α parameter needed in the power distance calculation.

Exemplary Test Case

In the following sections each of the proposed methodologies is described based on a simplified test case. The following points are critical in defining a representative test case for this purpose:

- The test case should be simple enough to explain each methodology but complex enough to highlight challenges with each method.
- The grid lines should not always be a direct connection between two points, to reflect obstacles that occur in a real grid such as buildings, water bodies or steep terrain.

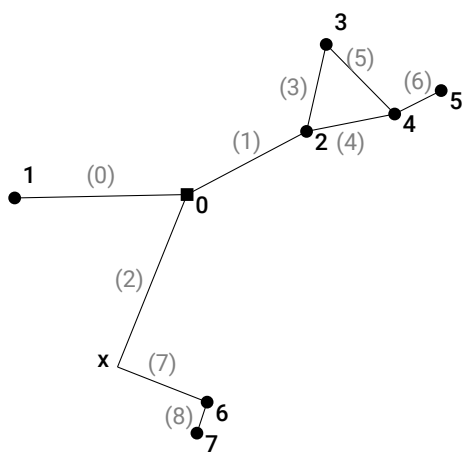


Figure 2.1.: *Test Case for the power distance calculations*

- The grid should have meshes to reflect that there are different lines over which power can be distributed to the same node.
- The demand per node should be chosen in a way that makes the effect of the cost scaling parameter alpha visible when constructing an artificial grid.
- The demand point should be distributed geographically to better visualise the demand distribution based power distance.

Based on the above-mentioned criteria we have selected a grid topology that reflects the properties of each parameter, shown in Figure 2.1.

Nodes [2] are numbered in bold where node **0** is the transformer station that connects the distribution grid to the regional grid. For simplicity the test case only contains one transformer station [3]. In a case with multiple transformer stations, each algorithm assigns nodes to one transformer either through the closest distance or the existing grid connections. In the test case, there are nine straight line segments, numbered with an index in brackets e.g. (1), that connect the nodes and create one mesh (between lines (3),(4) and (5)) and one corner (lines (2) and (7)) in the grid. The demand at each node is summarised in Table 2.1

Node	1	2	3	4	5	6	7
Demand	200	20	50	100	40	60	50

Table 2.1.: *Demand per node in the test case in kW*

The test case was created based on real geographical coordinates at a selected location in Norway. The resulting line lengths per segment are summarised in Table 2.2. Additionally the start and end point of each line is given in accordance with Figure 2.1.

Note that demand is given in kW and line lengths in m for the test case while the power

[2] *Node* refers to each point in the graph. In the case of this report a *node* is equivalent to a *substation*

[3] *transformer (station)* refers to the transformer between the regional grid (*R-nett*) and the high voltage distribution grid

Line	from node	to node	length [m]
(0)	0	1	456
(1)	0	2	357
(2)	0	x	496
(3)	2	3	237
(4)	2	4	238
(5)	3	4	258
(6)	4	5	137
(7)	x	6	254
(8)	6	7	86

Table 2.2.: *Details on lines in the test case, x refers to the connection point between lines (6) and (7)*

distance uses the unit MW$^\alpha$ km. Due to the scaling of the alpha parameter the unit conversion is:

$$P_\mathrm{d}\ [\mathrm{MW}^\alpha\ \mathrm{km}] = 1000^{(\alpha+1)} \cdot P_\mathrm{d}\ [\mathrm{kW}^\alpha\ \mathrm{m}] \tag{2.2}$$

$$P_\mathrm{d}\ [\mathrm{MW}^\alpha\ \mathrm{km}] = 1000^{\alpha} \cdot P_\mathrm{d}\ [\mathrm{kW}^\alpha\ \mathrm{km}] \tag{2.3}$$

$$P_\mathrm{d}\ [\mathrm{MW}^\alpha\ \mathrm{km}] = 1000 \cdot P_\mathrm{d}\ [\mathrm{MW}^\alpha\ \mathrm{m}] \tag{2.4}$$

2.2. Minimal power distance

The minimal power distance $P_\mathrm{d}^\mathrm{min}$ refers to the minimal power distance that can be achieved given a certain demand distribution and the existing grid topology. Mathematically, this can be written as

$$P_\mathrm{d}^\mathrm{min} = \min_{p_e} \sum_{e \in E} L_e\, |p_e|^\alpha \ . \tag{2.5}$$

As was shown in [5], finding the minimal power distance will always result in a radial system, but iteratively scanning for the minimal grid configuration is computationally expensive if the topology has meshes. We were able to describe a globally optimal[4] algorithm for finding the minimal power distance, but it scales exponentially with the number of meshes.

In this approach the upper bound on the complexity is

$$\mathcal{O}(E^C)\ , \tag{2.6}$$

with E the number of lines in the graph and C the number of meshes.[5]

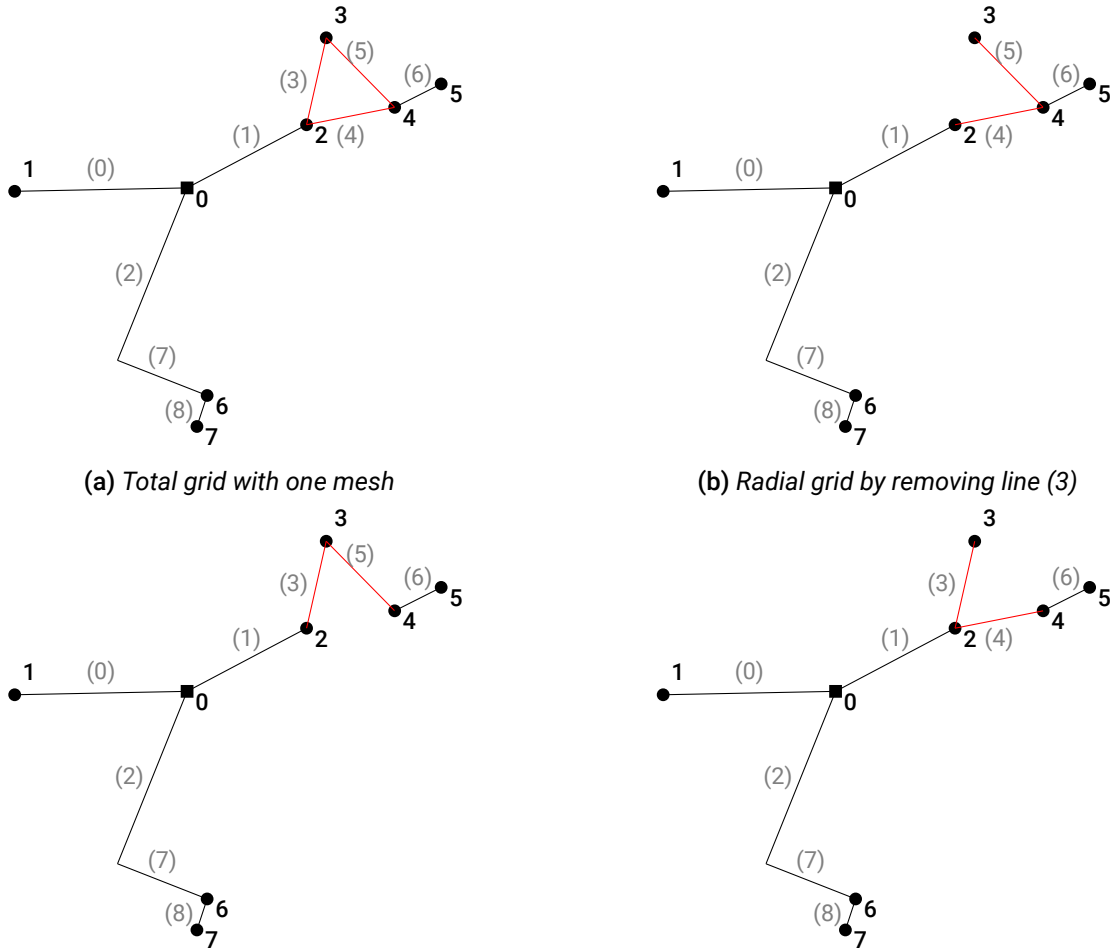
The algorithm uses the finding that an optimal cost minimising topology is always tree-shaped. The approach is to effectively test all possible combinations between removing lines, and searching for the one tree within the given grid topology that minimises the power distance. If there is one mesh, the algorithm iterates over all lines that constitute the mesh. If there are two meshes, the algorithm has to iterate over all combinations of removing two lines, one from each mesh. If there are C meshes, the algorithm needs to iterate over all combinations of removing C lines from the grid. This is why the computational complexity grows exponentially with the number of meshes. For more details and alternative, sub-optimal algorithms, see [5].

On the test case the algorithm iterates over each line segment ((3),(4),(5)) that constitutes the mesh in the system. At each combination of removing a line a radial grid remains for which the power distance can be calculated. The resulting grid systems are shown in Figure 2.2 where the mesh elements are marked in red. Removing line (4) gives the highest power distance (16.54 MW$^\alpha$ km) for the system, while removing lines (3) gives a lower power distance of 15.93 MW$^\alpha$ km). The minimal power distance for the system is 15.60 MW$^\alpha$ km for a radial system when line (5) is removed.

For the test case, computing the minimal power distance is a simple task with few iterations through a single mesh. However, each additional mesh in a system will greatly increase the complexity. In the case of real grid systems this reaches an extent where the minimal power flow calculation is

[4] "Global optimality" means in simplified terms that the algorithm always finds the best solution to a problem

[5] As shown in [5], this bound can be formulated somewhat more tightly as $\mathcal{O}(\prod_{c \in C} E_c)$, with E_c the number of lines constituting each mesh. However, this does not affect the exponential scaling property.



(a) *Total grid with one mesh*

(b) *Radial grid by removing line (3)*

(c) *Radial grid by removing line (4)*

(d) *Radial grid by removing line (5)*

Figure 2.2.: *Illustration of the iteration for finding the minimal power distance based on the test case*

too complex to be performed.

2.3. Power-flow based power distance

In [5], we also proposed an alternative to the minimal power distance. Instead of finding the tree which minimises the power distance, the full, meshed topology is used. The flows are determined by a power flow computation, with equal impedance per km for each line.

This power-flow based power distance $P_{\mathrm{d}}^{\mathrm{flow}}$ has two main strengths: the p_e are easily computed using existing tools for power flows, and the approach is far more intuitive and more easily communicated to stakeholders affected by the regulation.

The power distance is computed as in (2.8),

$$P_{\mathrm{d}}^{\mathrm{flow}} = \sum_{\forall e} L_e \, P_e^{\alpha} \tag{2.7}$$

where P^e is computed by an optimal power flow algorithm.

An optimal power flow model is a common tool for the analysis of electrical system operation and component sizing for system expansions. A power flow analysis determines the optimal steady-state operation of a grid system based on generation,

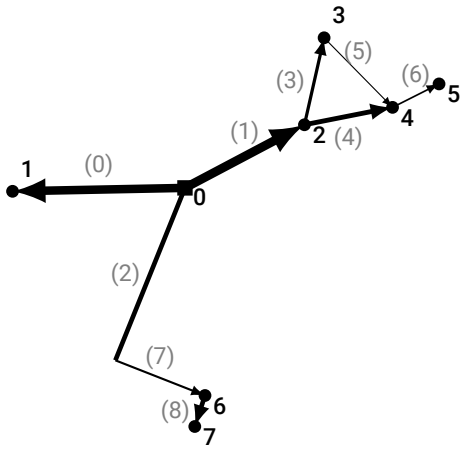


Figure 2.3.: *Power flow based power distance, thickness of lines defines flow on line, arrow defines direction of flow*

demand, and line specifications. In the case of constant line specifications and known demand per node, the power flow computation determines the flow per line that minimises the real and reactive losses in the system within component constraints.

For the power-flow based power distance the full meshed grid is used. In the test case all nodes and lines are considered in the algorithm which allows power to be distributed to one node over multiple lines. For the radial part of the grid the power flow is defined by the demand (and generation) at each node. In the meshed part of the grid (lines (3),(4),(5)) power can flow towards node 4 via line (4) and via the lines (3) and (5). Alternatively power could flow to node 3 via the line (3) or via line (4) and (5). The power flow algorithm calculates the flow on each line segment based on the ideal physical flow. The resulting flows for the test case are visualised in Figure 2.3, where the arrows define the flow direction and the thickness indicates the intensity of flow on each line. From the results of power flow per segment and the length of each line a power distance can be calculated. For the test case the power-flow based power distance is 16.68 MW^{α} km .

2.4. Artificial grid based power distance

Using the existing grid has two drawbacks: 1) it requires an extensive amount of high-quality data, and 2) it affects the exogeneity of the parameter, as the grid companies, theoretically, have the possibility to affect the power distance benchmarking parameter with their investment decisions. The latter issue may lead to undesired investment incentives or disincentives, as was shown in [5].

Alternatively, one may attempt to construct an ideal grid given the demand and the location of the closest transformer station. The benchmarking models used in Sweden and still in use in Spain do exactly this. However, constructing a realistic grid as a direct benchmark is by no means a trivial task. Many restrictions would need to be accounted for (streets, buildings, bodies of water, steepness of terrain, other zonal restrictions), and an extensive cost and infrastructure data base would need to be used.

We do not attempt to construct a realistic grid layout, but rather aim for an idealised grid. This grid should take into account the demand distribution and the alpha parameter, but it should not be seen as *the best achievable topology*. Rather, it should be used as a relative benchmark in the sense "how close does each grid company come towards the idealised grid?" and then benchmarking the grid companies against each other. Complicating factors such as the terrain, building density or coastal obstacles can be taken into account in additional correction factors in the benchmarking.

To compute the idealised grid, we follow this algorithm:

1. We identify the closest substation to any transformer station, and connect it to its transformer station.
2. The closest unconnected substation - the free substation - is identified.
3. All demand except the demand from the free substation is allocated to the closest node

(transformer station or connected substation). The power distance is computed.

4. The free substation is connected to the node where it increases the power distance by the least amount.

5. The algorithm repeats Step 2 to 4 until all substations are connected.

Step 3 is essential to incorporate the α-Parameter into the algorithm. The algorithm itself is a deterministic heuristic to find a good approximation for the ideal grid connecting all substations to the transformer station(s), given an α. Since it is deterministic, it will always yield the same result for a given set of data. Being a heuristic, it is one approach to obtaining a result for which there is no guarantee on optimality.

Based on the artificial grid a power distance is computed using a power flow method. The method determines the flow on each line element of the grid and calculates the power distance as follows:

$$P_{\mathrm{d}}^{art} = \sum_{\forall e} L_e \, P_e^{\alpha} \quad , \tag{2.8}$$

with L_e the length of line element e, P_e the (absolute) flow along that line element, and $e \in \mathcal{E}$ the set of all lines in the artificial radial grid.

Note, that there are two possibilities to use the artificial grid methodology. One can either compute a new grid for each hour and demand distribution or build an artificial grid once and populate the nodes with hourly demand values for each calculation. We consider it the most objective to create one artificial grid per DSO based on a representative hour, e.g. the hour of highest demand, and use the same grid for each hourly calculation. Creating an artificial grid for each hour not only drastically increases computation time but would also not reflect the real situation where one grid should be equipped to handle demand in any hour.

For the test case, shown in Figure 2.4, the algorithm uses the coordinates and demand of each node as an input. Starting from the transformer station the nodes in the system are iteratively connected through direct lines based on which connection increases the total power distance by the least in each step. The order in which lines are added by the algorithm is reflected by the line numbering, i.e. line (0) is the first connection.

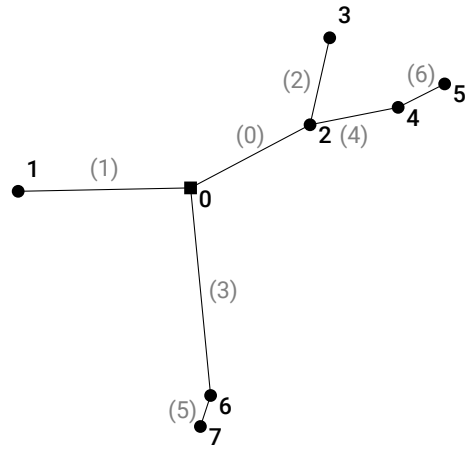


Figure 2.4.: *Artificial grid based on the test case, numbering of lines refers to the order in which they are connected*

Each connection step performed by the algorithm will be described in the following:

1. **Line (0)**
   - The algorithm will first identify node 2 to be the closest substation to the transformer station.
   - Line (0) is built between the transformer station and node 2.
2. **Line (1)**
   - The closest free substation to the transformer station or an already connected node is identified at node 1.
   - All remaining demand is allocated to the closest node, meaning that the demand from nodes 6 and 7 are allocated to the transformer station and the demand from nodes 3, 4 and 5 will be projected to node 2 leading to a total demand of 210 kW at node 2.

- The power distance for this system is computed and the free node 1 will be connected where the power distance increases by the least amount. In this case, node 1 is connected directly to the transformer station because the large distance to all other nodes would increase the total power distance more.

3. **Line (2)**
   - In the next step, node 3 is identified as the closest free transformer station.
   - Again, the demand of all non-connected nodes is allocated to connected substations and the power distance is computed.
   - Node 3 is connected to node 2, to which the demand of node 4 and 5 are allocated, via line (2).
4. **Line (3)**
   - The next free substation is node 6,
   - Line (3) is built with the same approach as for previous lines. Since the distances to the remaining connected nodes would increase the power distance more, node 6 is directly connected to the transformer station via line (3).
5. **Line (4)**
   - Node 4 is identified as the closest unconnected node.
   - In this step the allocation of the remaining demand becomes crucial to the decision whether it should be connected to node 3 or node 2.
   - The demand at node 5 will be allocated to the closest connected node, node 3. As a result, 210 kW of power will need to be distributed to node 2 and at least 130 kW to node 3.
   - The demand in node 7 is allocated to node 6 and does not impact the branch towards node 4.
   - If node 4 is connected to node 3, then the total flow to node 3 needs to be 190 kW of which 90 kW cover the demand in node 3, 40 kW are transferred to node 5 and 100 kW to node 4.
   - Should node 4 be connected to node 2, the flow to node 3 will be 130 kW, and the flow to node 4 will be 100 kW.
   - When calculating the power distance for both alternatives, one can see that the connection of node 4 to node 2 increases the power distance by less than a connection between node 3 and 4.
   - The artificial grid thus establishes a connection between node 2 and 4. Intuitively, by looking at the test case connecting node 4 to node 3 would create a longer path for power to flow in a radial system.
6. **Line (5)**
   - In the next step, node 7 is the closest unconnected substation.
   - Because of its short distance to node 6 and no close substation from where demand is allocated towards this branch, node 7 is directly connected to node 6.
7. **Line (6)**
   - The last node to be connected is node 5
   - Node 5 is connected to the closest substation as there is no remaining unconnected demand that needs to be taken into account.

For the resulting radial system, the power flow is uniquely defined by the demand at each node. From the power flow and the length per line the artificial grid based power distance can be calculated. For the test case it is 14.85 MW^{α} km. Note that the power distance on the artificial grid is lower than the minimal and the power flow based power distance. This is a result of shorter lines in the artificial grid which will always create direct connections between nodes and does not account for obstacles or meshes.

2.5. Demand-distribution based power distance

The last approach to compute a power distance parameter avoids using a grid - real or synthesised - altogether. Instead, the power distance is computed from the distribution of demand around the closest transformer station.

The reasoning behind this is to define a highly exogenous measure with minimal data requirements and limited computational effort. Computing the demand distribution and ultimately the demand-distribution based power distance is a purely statistical approach that only takes the geographical coordinates and demand per customer as input. The resulting parameter is neither related to nor benchmarked against the real grid, thereby providing an exogenous measure of the DSOs task rather than their effort.

The demand-distribution based power distance is computed in the following steps:

1. We identify the closest transformer station to every substation and isolate areas with k substations around one transformer station as origin.
2. For every substation i within one area the distance (r_i) and angle (θ_i) to the origin is calculated. The radius is the euclidean distance between node and transformer station. The angle is defined in degrees where 0 degrees points North.
3. The distance between substation and origin is multiplied with the substation demand (D_i) to obtain a linear power distance: $P_{d_i} = r_i^{1/\alpha} \cdot D_i$. The radius is scaled down by the cost-scaling parameter to avoid double counting in step 6.
4. For each substation a normal distribution $\mathcal{N}(\theta_i, \sigma)$, where $\sigma = 1/\alpha$, is created and scaled with the linear power distance P_{d_i}. Here the alpha parameter defines the width of the distribution and thereby the distance between two demand points at which building one line becomes more economocially favourable than building two lines.
5. The individual distributions per substation are added to create a continuous demand distribution around the transformer station $\mathcal{D}(\theta) = \sum_{i=0}^{k} \mathcal{N}(\theta_i, \sigma)$
6. The integral over the distribution (all angles from 0 to 360 degrees) to the power of α gives the demand distribution based power distance around the transformer station

$$P_{\mathrm{d}}^{\mathrm{stat}} = \int_{\theta=0}^{360} \mathcal{D}(\theta)^{\alpha} d\theta$$

In the initial and most simple approach, topographical constraints such as terrain steepness, water bodies and densely inhabited areas or differences due to weather conditions are not accounted for. However, correction factors can be applied to the calculated parameter at the expense of higher computation time. Examples for such correction factors are a topographical correction based on the proportion of area types in the considered surrounding or a weather correction based on historic weather statistics per area.

The approach can be described in more detail when the test case is considered. Figure 2.5 shows the input for the statistical methodology. For each

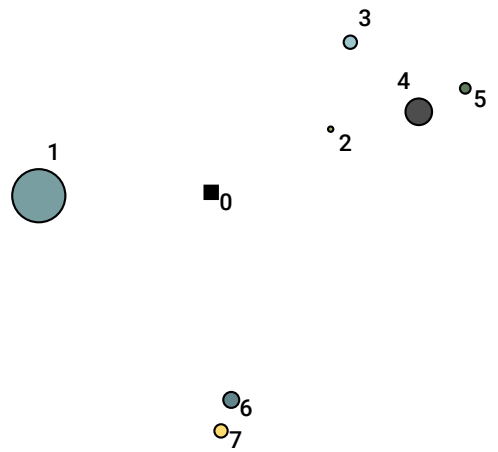


Figure 2.5.: *Input for the statistical power distance calculation based on the test case. Size of bubbles shows demand per node*

node the geographical coordinates define the position and the demand per node determines the size of the bubbles. For each node the radius and angle to the transformer station can be calculated as summarised in Table 2.3.

Node	radius [m]	angle [deg]	demand [kW]
1	456	267	200
2	357	62	20
3	542	43	50
4	589	69	100
5	726	68	40
6	552	175	60
7	632	178	50

Table 2.3.: *Input per node for the statistical power distance distribution*

Following step 4. in the process description a scaled normal distribution of the power distance is calculated around the angle of each node. The individual distributions are visualised in Figure 2.6. The colours of the curves correspond to the colour coding of nodes in Figure 2.5 and serve as a guide to the eye. Overlaps in the individual distribution symbolise angular proximity between two nodes. In this approach, we argue that it is more efficient to serve nodes that are in angular proximity – i.e. show overlaps – with one strong line rather than two individual lines.

Consequently, when the individual distributions are summed to a continuous distributions, as described in step 5., larger peaks are visible at angles where multiple nodes are in close angular proximity. The continuous distribution for different alpha parameters is shown in Figure 2.7. As alpha represents the cost-scaling of lines with transferred power, the economic efficiency of building one or multiple lines depends on its value. For large alpha values building individual lines becomes more efficient, resulting in less overlapping of peaks in the distribution, while small alpha values suggest building strong lines to supply multiple nodes is more cost efficient, resulting in stronger overlaps in Figure 2.7. The α-parameter is defined by the actual

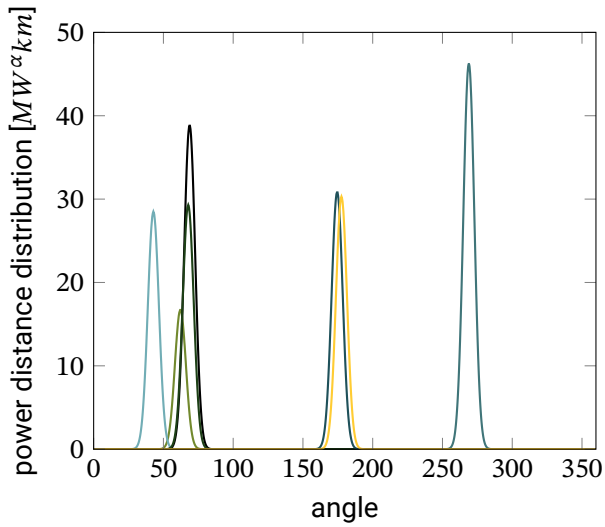


Figure 2.6.: *Normal distribution at each node adjusted for the cost-scaled demand per node (P^{α}) and the radius*

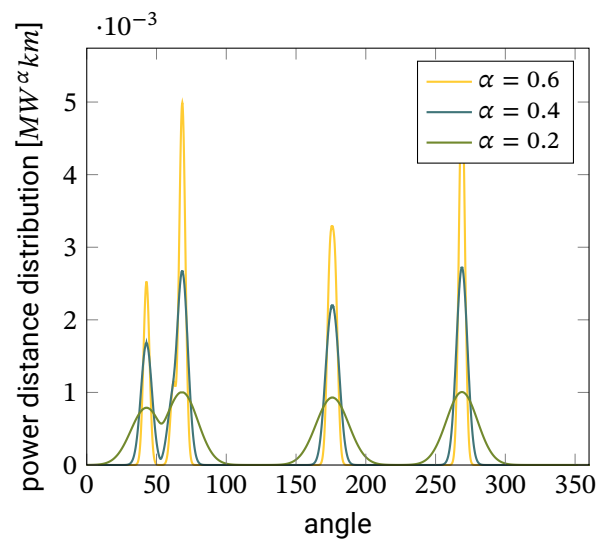


Figure 2.7.: *Total power distance distribution for the test case for different alpha parameters (normalised)*

financing conditions for assets in the distribution grid and will be discussed in more detail in section 2.7 The resulting power distance in this system is 20.55 MW^{α} km .

Table 2.4 summarises the power distance results for all four introduced methods. As discussed, the artificial grid based power distance is

the lowest, as all grid connections are direct and no meshes occur in the system. For the full system the minimal power distance is lower than the power flow based power distance, which supports the hypothesis that the minimal required grid will always be a radial system. Compared to the artificial grid the minimal radial grid has longer line lengths due to obstacles in the real grid as simulated by lines (2) and (7). As a result, the artificial grid based power distance is lower than the minimal power distance.

Table 2.4.: *Comparison of power distance result for the test case*

Parameter	Method	Power distance [MW$^{\alpha}$ km]
$P_{\mathrm{d}}^{\mathrm{min}}$	Minimal	15.60
$P_{\mathrm{d}}^{\mathrm{flow}}$	Power Flow	16.68
$P_{\mathrm{d}}^{\mathrm{art}}$	Artificial grid	14.85
$P_{\mathrm{d}}^{\mathrm{stat}}$	Statistical	20.55

2.6. Power distance in the low-voltage distribution system

So far and for the purpose of this study, we compute the power distance parameter in the high-voltage (1 to 22 kV) distribution system only. The last mile from the substation to the consumer is not taken into account.

In Norway, two systems for the low voltage distribution are used: 220 V / 1-phase or 3-phase lines and 400 V / 3-phase lines. Generally speaking, three-phase distribution is the technically preferred option, while 1-phase is the historically wider used system. 3-phase systems are needed for fast-charging of electric vehicles or other high-power applications, and are better suited to ensure equal loading of all phases. While 3-phase systems are used in both 400 V and 220 V systems, the main advantage of higher voltage lines is that they also allow for longer lines due to reduced losses.

Not taking into account the low-voltage distribution grid can lead to distortion of the incentives towards the technically inferior 220 V system: since the 220 V lines are usually shorter than those in with 400 V, more of the distance from a transformer station to the consumer is covered by high voltage lines. If only high-voltage distribution is considered in the benchmarking, a grid company would be considered to have a higher output with the 220 V system, than an otherwise equal grid company using 400 V lines. Nota bene, the same effect is evident in the current output parameter "length of all high voltage distribution lines" and only partly corrected by the use of "number of substations".

How can this be addressed? Using the explicit line length or power distance for the low-voltage distribution grid may not be feasible due to a lack of detailed grid data on this voltage level. However, the location and demand profile of each consumer is known thanks to the Smart Meter (SM) infrastructure. It should be feasible to take the low-voltage distribution into account through a power distance computation, possibly similar to the demand-distribution based power distance. By doing so, a grid company should not have a disadvantage, or might ideally even have an advantage from using longer 400 V lines.

However, a more detailed discussion and analysis of how a low-voltage power distance could be defined and how it could be integrated into the regulatory framework is beyond the scope of this report.

2.7. Alpha parameter

In all proposed methods to define and calculate the power distance, the cost scaling parameter α is used to represent the relation between cost and capacity of a power line. To realistically capture the needed investments to build and maintain lines at different rated power, α should be between 0 and 1. For $\alpha = 0$ the power distance is equal to the sum of the length of lines and independent of the transmitted power, for $\alpha = 1$ the relation between cost and power is linear and suggests that building

two lines at half capacity will incur the same costs as building one line at full capacity.

In the previous study [5] the impact of different values for α was investigated, but the exact value was not determined. One aim of this project is to define methodologies to determine the relation between cost and capacity and consequently calculate the alpha parameter. Three methodologies that were investigated are described below.

After testing three methodologies we can not give a definite answer to which value for the α-parameter should be used in calculations. Instead, the use of different methodologies has allowed us to identify a range in which the α parameter is likely to lie. The choice may depend on the selected computation method for the power distance parameter.

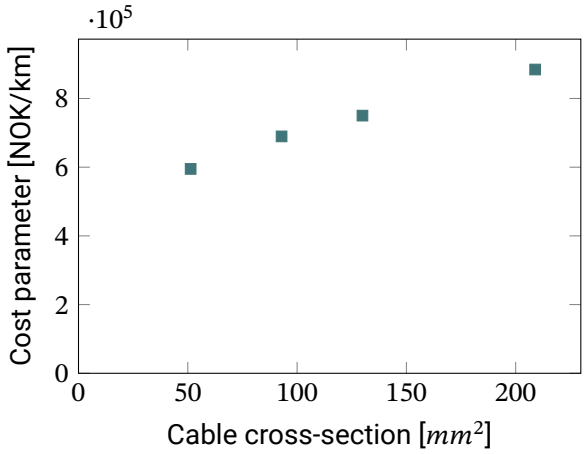


Figure 2.8.: *Relation of cost per km to the cable cross section.*

2.7.1. Alpha based on the cost catalogue

To verify the cost difference between building lines at different voltage levels, data from the cost catalogue [6] can be used which specifies cost data per cable type. In this approach we argue that the main driver for cost increases related to rated power is the cable diameter. Fundamentally, the transmitted power in an electric cable is defined by the voltage multiplied by the current, or after applying Ohm's law, by the square of the voltage divided by the resistance.

$$P = V \cdot I = \frac{V^2}{R} \tag{2.9}$$

If we assume that the voltage on a given line is fixed, the power scales with the inverse of the resistance. The resistance R of an electric cable is defined by

$$R = \frac{\rho \cdot L}{A} \tag{2.10}$$

where ρ is the material specific resistivity, L is the length and A is the cross-section area of the cable. For a given length and material, the power on a cable will consequently be proportional to the cross-section area.

$$P = \frac{V^2}{L \cdot \rho} \cdot A \tag{2.11}$$

The cost catalogue provides detailed information on unit costs per cable type. Additionally, the material specific resistivity of the used materials and the cable cross section can be extracted from a database on cable characteristics. A limitation of this method is, that only the capital cost for cables is considered. The total incurred costs per line length will also depend on other required hardware, land preparation and operation costs which may vary significantly for different projects. However, generally we assume that overall cost of building lines of different voltage levels in the high-voltage distribution grid is not expected to differ greatly as large CAPEX components and operational costs are similar. Thus, the relation between cable cost and rate power can be used to determine the α-parameter.

The calculation was performed for a FeAl overhead line without insulation. Other cable types are described in the cost catalogue as well but we decided to focus on the cable type with the most available data. The obtained relation between the cable cross-section and the cost per unit length is shown in Figure 2.8. The cost parameter shows a logarithmic increase with cable cross-section, and consequently with transferred power. By using a logarithmic fit on the obtained data points the α-parameter can be calculated to $\alpha = 0.35$.

2.7.2. Calculating alpha from the asset register

A second approach to determine the alpha parameter is to investigate the relation of total cost per DSO versus the length of lines per voltage level. The resulting linear equation system gives a solution vector which is expected to show an exponential trend with higher line voltage from which alpha can be obtained.

$$\begin{bmatrix} L_1^{1\ kV} & \dots & L_1^{11\ kV} & L_1^{22\ kV} \\ \dots & \dots & \dots & \dots \\ L_n^{1\ kV} & \dots & L_n^{11\ kV} & L_n^{22\ kV} \end{bmatrix} \cdot \begin{bmatrix} x^{1\ kV} \\ \dots \\ x^{1\ kV} \end{bmatrix} = \begin{bmatrix} C_1 \\ \dots \\ C_n \end{bmatrix} \tag{2.12}$$

where $L_i^{1\ kV}$ is the total length of lines with voltage 1 kV of DSO i, C_i is the total incurred cost until the present year for DSO i and $x^{1\ kV}$ is the cost parameter for lines of voltage 1 kV. The size of matrix L will be determined by the number of DSOs n and the number of voltage levels 1 to 22 kV.

As part of the current regulation the expenses of all DSOs are collected in an asset register. The asset register reports on the total incurred expenses per DSO and year including new investments, reinvestments and depreciation for each grid level and line type. For the calculation of the alpha parameter the total cost of all high voltage lines in the local distribution grid is considered. From the RME grid database the total length of lines per voltage level for each DSO can be obtained (within the constraints of data availability). The lines are aggregated into three voltage levels: 5, 11 and 22 kV, while reported voltages may vary.

Data from all Norwegian DSOs in the asset register with costs > 0 are used. Therefore, the resulting problem becomes an overdetermined linear equation system which can be solved using a least-squares approach.

For the three analysed voltage levels an increase in cost per km can be observed between 5 kV and 11 kV (see Figure 2.9). However, for a higher line voltage of 22 kV the cost seems to drop to the lowest level. Based on the obtained results we deem the results from this approach non-conclusive.

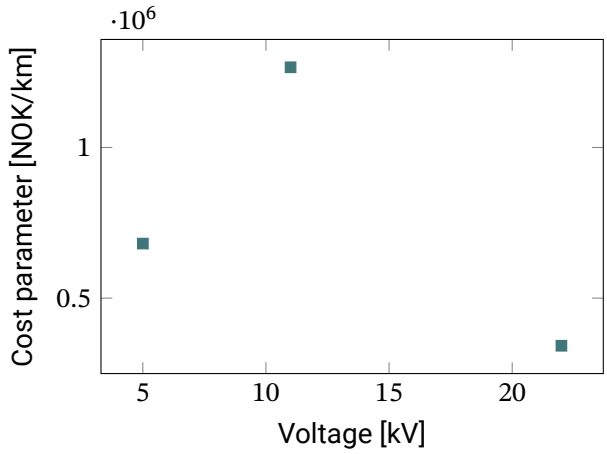


Figure 2.9.: *Relation of cost and line voltage from the asset register calculation*

Limitations of this approach are that only total investment costs are considered per voltage level. In the distribution grid the main expenses for building a power line are location and topography specific, rather than voltage dependent and the cost of cables scales with the cable cross-section which is independent of the voltage. Costs for land preparation and general hardware components (masts, insulation, etc.) will not differ greatly between an 11 kV and a 22 kV line. Additionally, the line voltage is constrained by the existing grid and does not necessarily reflect the actual capacity on the line. As accumulated costs until the present year are used, historic cost developments and different financing structures can not be taken into account.

2.7.3. Estimating alpha from the idealised grid

When calculating the power distance based on an idealised grid the α-parameter is taken as an input. Depending on its value the connections between substations, which the algorithm builds based on the minimal increase in power distance, will differ.

With the assumption that additions to the real grid are always made under ideal economic conditions, the idealised grid which resembles the structure of the real grid the most should represent the existing investment structure. For the investigated test cases we found that an α-parameter in the range of 0.3-0.5 will give the closest match to the real grid.

2.7.4. Further suggestions for calculating the alpha parameter

Beyond the three investigated methods, further approaches were considered and will be discussed briefly. The implementation of these methods is beyond the scope of this project and would require additional data that was not available during this project.

Real cases The α-parameter could also be estimated from project specific data of different DSOs. In this methodology, the sample size of available data is crucial to avoid bias from geographical differences. If the sample size is small, cost differences are more likely to be caused by location specific constraints. Some examples are that the cost for land preparation and digging will differ in rural and urban areas, installation on steeper terrain is expected to incur higher costs.Thus an additional correction for topography would need to be performed. We consider this approach the most challenging due to limitations of data availability and bias.

Current DEA benchmarking Introducing a new benchmarking parameter in the DEA model should provide a more objective comparison of the performance of grid companies. At the same time a change in the regulatory structure should not majorly change the rating of individual companies and their positioning with respect to other DSOs. Based on computations for all grid areas in Norway and a comparison of the power distance results to the current benchmarking output, it could be possible to scale the alpha parameter to fit a similar benchmarking structure as obtained from the current method. It will be interesting to investigate how the alpha parameter impacts the rating of grid companies based on the power distance compared to the current output parameters. However, we do not believe that using the current benchmark as a starting point for defining the alpha parameter is an approach that should be pursued.

3. From power distance to energy distance

The power distance is computed for a specific hour, and can be calculated individually for all hours within the year. For the DEA approach, we need a selection of one or more parameters describing the task of the grid companies throughout the whole year. Selecting one individual hour might not be representative of or describing the task of the grid company, while it on the other hand would not be efficient or necessary to account for all 8760 hours in the year.

In this chapter, we will discuss how the power distance can enter the DEA benchmarking used for the income regulation of grid companies in Norway through one or more annually aggregated parameters. We revisit some of the challenges with aggregating the hourly power distance into a yearly energy distance as described in [5], and introduce additional concerns raised by a reference group of grid companies participating in this work.

The conclusion of this chapter is a selection of recommended designs for the energy distance parameter, that will be tested and evaluated in the continuation of this report.

Important questions that we address in this chapter are:

- What is the actual task of a grid company?
- (How) is the task of grid companies with very different demand profiles comparable?
- What are the main cost drivers in building and maintaining the high voltage distribution grid?
- Which aggregation methods can be used to define energy distance parameters that accurately reflect the cost drivers?

3.1. Cost drivers in the distribution grid

The choice of energy distance parameter(s) should reflect the cost drivers that a grid company is exposed to in the distribution grid. A DSO has the obligation to cover the demand of customers in the grid at all times. Consequently, the grid company has to build and maintain a grid that is able to serve its customers/demand at all hours - even the hour with the highest load. This applies even though a grid can operate above limits in select hours if they are few enough and far enough apart. In the future, a grid company serving the distribution grid needs to accommodate changes in customer patterns that are likely to lead to higher peak loads due to new types of demand, feed-in from distributed residential generation and electrification of several sectors such as onshore and offshore transportation.

On the other hand, the grid companies can expect new consumers to provide flexibility to the grid and relieve strain on the electric infrastructure through demand response and flexible demand with the increased introduction of smart homes.

The required grid infrastructure to supply customers in the distribution grid incurs high costs for the DSOs. The costs that a grid company faces can be broadly divided into four categories: Investment costs, cost of energy delivered, administrative costs, and cost of losses. The overall cost is directly related to the power and energy supplied but each cost component has different driving factors.

Investment costs: These costs represent the instantaneous power that a grid company needs to cover with its infrastructure. If an area has a high

peak demand, the power grid has to be built to cope with that demand regardless of the total energy delivered within a year. If an area has a high density of vacation homes, they might be faced with high demand in weekends and holidays, while the total annual energy consumption is quite low. A parameter that captures peak demand and investment costs is the maximum power distance – described by the ∞-norm.

Cost of energy delivered: The cost of energy delivered, i.e. the operational cost, is driven by total energy delivered in a grid. A comparable measure in this context it the 1-norm – the sum of (or average) kWh over the year.

Administrative costs: This includes costs related to performing administrative tasks, such as customer service, invoicing, etc. These are costs related to the number of customers served in the grid area, that comes in addition to the cost of delivering energy. This cost is described separately from the power distance approach as two companies who deliver the same amount of energy, over similar distances to different amount of customers will have different tasks to cover.

Cost of losses: The losses in a power system are quadratic to the energy delivered in a grid. As a result, losses increase quadratically with power transferred, and two grids facing the same total or average energy within a year might have entirely different losses. To illustrate this we describe an exemplary case of grid A and B, with 4 hours of demand:

Example: cost of losses

First, keep in mind that $Power = U \cdot I$ and $Losses = I^2 \cdot R$, so that with constant voltage, the current I is increasing linearly with increasing power, and losses are increasing quadratically with power. The hourly demand of company A and B for four arbitrary hours is given in Table 3.1

The sum (or average) of power is equal for the two companies, but the amount of losses they face are quite different, as seen in Figure 3.1. In fact, company B is suffering almost 40 % more losses than company A, due to its variable consumption.

Company	h1	h2	h3	h4	sum
A	5	5	5	5	**20**
B	5	10	2.5	2.5	**20**

Table 3.1.: *Hourly power consumption*

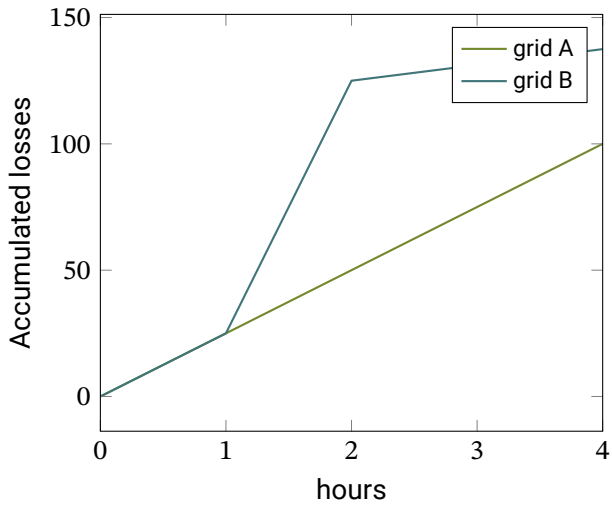


Figure 3.1.: *Accumulated losses for company A and B.*

Such challenges should be accounted for in the selected of annual benchmarking parameters.

3.2. From hourly values to a yearly parameter

The tasks of the grid company are slightly more complex than building a grid that covers the hourly peak demand, and that can provide for a flat or a variable profile throughout the year. The main challenges that grid companies are facing will be discussed in the context of different energy distance parameters. The following section gives an overview of possible yearly aggregation methods and their suitability as benchmarking parameters. In [5] we discussed four main options for moving from hourly power distance to yearly energy dis-

tance:

- aggregating hourly power distances,
- selecting representative hours,
- averaging demand over several hours,
- formulating an optimisation problem directly for the energy distance.

The first three are briefly revisited here, and we refer to [5] for further details.

3.2.1. Selecting parameters to describe the task of the grid company

Before we go ahead selecting parameters to describe the task of the grid company, we need to decide how to account for the different concerns raised by a number of grid companies in a related workshop. In order to identify challenges that the grid companies face in performing their task - specifically challenges that incur direct costs - we invited a selection of six grid companies to a workshop with RME, where we asked them to elaborate on how they would describe their task as a grid company. The received input was used as a knowledge base to define aggregation methods and will be discussed in more detail in Section 3.3.

A brief summary of the main concerns and the possible integration in the benchmarking process is given in Box 3.1.

First of all, we aim to select parameters that both account for investment costs and the cost of losses, meaning that we will study a collection of more parameters rather than a single one. We will also consider combining new and existing parameters in order to describe the grid companies and their tasks.

Second, there are several concerns that we will not recommend to reflect in the chosen parameters - either due to complexity, or because we suspect similar ramifications for all grid companies, as described in Box 3.1. In the following chapter we will present a set of alternative methods of how to aggregate power distances to energy distances describing and comparing the tasks of the grid companies over the full year.

Box 3.1: Concerns raised by the grid companies - and recommended actions

- **Maximum local vs. total demand:** Considered in selecting the parameters, as it can be directly handled in how we aggregate the energy distance.
- **Installed capacity vs. actual demand:** Assuming this is an issue that is relevant for all grid companies and potentially creates systematic differences, this aspect requires further investigation.
- **Accounting for historical demand distribution:** Not to be accounted for by the parameters describing the task of the grid company. We have stated some suggestions in how to resolve this issue in Section 3.3.3.
- **Costs vary with type and area of construction:** Not directly handled by the parameters, but recommended to be adjusted for after the fact, as is already practised in the DEA.
- **Demand peaks of less than one hour:** Could be handled outside of the DEA by including specific cases in the regulation for higher grid levels. It could also be accounted for in the parameter by reporting maximum load for the specific demand locations (given that they are few enough).

3.2.2. Aggregating hourly power distances

The simplest approach to arrive at an energy distance is to use a summation, or in generalised terms a norm over all hourly values. Other possibilities are considering the maximum value or accounting for variations by using a quadratic measure. Typical norms would be:

1-norm This is the sum or mean over the absolute values.[1]

[1] As all power distances are positive, it is the same as the

2-norm This is the square root of the sum over the squares.

∞-norm The infinity norm, or largest absolute value in the set.

We also suggest to keep "number of customers" as a parameter, serving as a proxy for operational and administrative tasks that are proportional to number of customers, but we do not study this parameter in more detail.

The three norms have different interpretations. The 1-norm considers the total energy, but is independent of the distribution over time. The demand profile might be flat or uneven, but the 1-norm would be the same[2]. Applying a 1-norm could be interpreted as representing the expected task of the grid operator with even consumption in its supply area. It generally provides a measure for the mean task of the grid company.

A 2-norm allows outliers to have a higher impact than values closer to the average. Hence, the 2-norm addresses the problem of variability in demand. When using the 2-norm, hours with a high power distance value are, in theory, weighted more than those with a low power distance.

The ∞-norm only looks at the largest power distance in the year. This is meaningful in the sense that this is actually the largest power distance on which the grid company has to dimension its grid. But at the same time it could give the grid company an incentive to promote increasing the peak demand, in order to have a higher output in the DEA.

Note that the input for all norms can be interpreted differently. An average value can be interpreted as the power distance that is calculated from the average demand per node or the average of all hourly power distances. Similarly, the maximum value may be reflected by a real situation in the grid, i.e. the maximum demand or power distance in one year, or by a "worst-case" for which the maximum demand at each node irrespective of whether they occur at the same time is used.

sum of hourly power distances. If one divides by the number of hours, it is the mean.

[2]Strictly, this holds only for $alpha = 1$, however, the core of the argument also holds when considering $\alpha < 1$.

Generally, aggregating the data based on the three new norms can be performed either before or after calculating the power distance, and the aggregating can either be performed on an annual, seasonal or nodal level. We elaborate on what this means for the different norms.

3.2.3. 1-norm

The 1-norm reflects the average situation that the grid is exposed to. It can be expressed by the average or the mean value, since all benchmarking is performed on annual basis they will only differ in magnitude with the proportionality to the number of hours. Aggregating the 1-norm before calculating the power distance means summing or averaging over all hours for each node. As we are studying the sum or average for the system, it makes sense to perform the aggregation on an annual, not seasonal level. There are two possibilities of performing the 1-norm with given hourly demand data per node:

average demand: The power distance calculated for the average demand in the system. The power distance is only calculated once.

average P_d: The average power distance in the year. Determined after calculating the power distance for each hour.

Average demand

The result of the first method is a set of sums or averages per node, and the next step is to calculate the power distance for this summed up or average hour.

$$D1_i = \frac{\sum_h D_{ih}}{H}$$

The power distance will be one value for the whole system, which serves as an energy distance - or a parameter describing the cost of supplying the average demand to the system, i.e. the operational costs.

Average power distance

The other option is to calculate all the hourly power distances based on actual demand first, and then take the sum or average of the 8760 resulting power distances. The latter method is more demanding in terms of computation time but may reflect the task of the DSO more accurately.

The 1-norm parameter will be studied and compared for all test cases for annual aggregation both prior to and after calculating the power distance in Section 5.5.

3.2.4. The 2-norm

The 2-norm can account for variability in demand. As describes before, uneven demand can lead to larger losses in a system with similar maximum and average demand. In short, it should be able to represent the difference between two systems that have similar average demand, but where one of the systems has a very unevenly distributed demand over time. If we aggregate the hourly demand by the 2-norm method prior to calculating the power distance, we end up with one value per node. The mathematical definition of the 2-norm is

$$D2_i = \sqrt{\sum_h \left(D_{ih} - \overline{D_i}\right)^2}$$

However, the losses increase quadratically with increasing demand, meaning that losses below average also decrease quadratic. The above definition of the 2-norm equally rewards any demand that is different from the average, regardless of direction. If we apply the 2-norm (as is) to represent increased losses due to variability in demand, we reward the grid company for having low demand and less losses.

Another property we want the parameter to hold is that it should only account for the variation around the mean, not the level of demand, as the 1-norm is already comparing the grid companies on the overall average demand level. If the average level is also included in the 2-norm and both are used for benchmarking, this would result in double counting.

Generally, the cost of losses in a system are already indirectly accounted for in the power distance calculation. The main assumption behind defining the power distance which scales with line length, is that transferring power over long distances incurs higher costs. The increased cost is related to investment and operation but also to the losses. Additionally, applying a 2-norm aggregation would introduce the cost of losses in the benchmark twice, which should be avoided.

We do not deem the 2-norm to be an ideal benchmarking parameter, as it may conflict with the 1-norm and would need additional adjustments to only reward demand peaks and not variability that leads to lower demand. As a result, we did not further investigate the 2-norm on the obtained test cases. A consideration of an aggregation parameter that appropriately accounts for the cost of losses can be included in future work.

3.2.5. The infinity-norm

One of the concerns mentioned by the grid companies was the difference between maximum total demand in the system, and maximum local demand. In short, a grid company can serve areas in which the demand peaks never occur at the same time. The hour with the largest total demand might be an hour where a large area that usually have high demand is experiencing quite low demand. When applying an ∞-norm, we therefore need to carefully consider the level of aggregation. As briefly mentioned above using a total maximum or a nodal maximum can impact the output and the way in which cost drivers are represented in the benchmarking. In this report, four alternatives for the infinity-norm are discussed, namely:

max demand: The power distance calculated for one hour with average demand per node. The power distance is only calculated once.

max P_d: The maximum power distance in the

year. Determined after calculating the power distance for each hour.

max nodal: The power distance calculated for a system where each substation is assigned its maximum yearly demand. The power distance is only calculated once.

max flow: The power distance calculated for a system where each line element is assigned its maximum yearly flow. The power distance is only calculated once.

Each aggregation method and its implications will be discussed below:

Maximum demand

If we apply the ∞-norm prior to calculating the power distance, we use the hour of maximum demand as an input to the power distance calculation. The results will require only one computation of the power distance output and therefore greatly reduces computational effort. However, considering the hour of maximum demand in a system may underestimate the situation of maximum load for which the grid needs to be equipped. The infrastructure will have to be able to supply the maximum demand at every node, irrespective of when it occurs. For larger grid systems, the probability of the maximum demand at each node occurring at the same time will be lower than for smaller systems, while both grids theoretically need to be equipped to handle such a situation.

Maximum power distance

For the infinity-norm based on the maximum power distance, a computation of the output is performed for every hour of available demand data. After the computation of the power distance for each hour, the maximum value is identified. The maximum power distance more accurately represents the hour in which it is most costly to supply power to all customers, compared to using the hour of maximum demand. The maximum power distance will always be at least as high as the power distance in the hour of maximum demand. Similar limitations as for the maximum demand apply in terms of representing the actual required investment to cover the maximum demand per node at any time.

Maximum nodal demand

The third aggregation option is nodal aggregation per substation demand. We find the highest demand in each substation, regardless of hour, and calculate one power distance value for the maximum local demands. This method accounts for the local maximum demand, but it could potentially overestimate the necessary investment level. In this project, only the aggregated demand per substation is considered, which means that simultaneity of demand for each customer is not considered. It could occur that the individual customers under a substation experience their peak demand at very different times, thereby impacting the maximum hour of demand per substation. Ideally, the grid should be equipped in a way that if every customer has its annual peak demand at the same time, power will still be supplied without interruption. In the real case, such extremes are unlikely to occur and may lead to shortages which are covered by the CENS-cost.

Maximum power flows

This method is only relevant for the power distances that consider a real or constructed grid. It is described in detail in the Appendix under Algorithm 1, and allows us to consider the maximum flow on each line, regardless of hour.

The idea is to calculate the optimal power flow for the grid in each hour, and find the maximum flow endured by each line segment, irrespective of the hour in which is occurs. This is equivalent to a nodal aggregation level per line segment. The power distance is then calculated for the system with maximum flows on all lines.

A nodal aggregation based on maximum power flow on each line prior to calculating the power

distance considers the maximum flow on each line, rather than the maximum demand in each node, and is therefore less likely to be an overestimation of the (optimal) investment costs incurred by each grid company. Similarly to the maximum nodal demand, this aggregation method accounts for a maximum situation in the high voltage distribution grid but does not account for differences in peak load in the low-voltage distribution grid.

3.2.6. Selecting representative hours

To simplify any computation, one could aim at using representative hours rather than all 8760 hours of the year. For example, one could use the hour with the highest net demand, winter peak, summer valley or the first Wednesday at 15:00 of each month. Dependent on how the representative hours are defined, the approach might make the task of two grid companies less comparable. One grid company could be faced with high demand in the relevant hours, while the other have a low demand. It may also be possible for the grid companies to affect demand in the representative hours.

Alternatively, one could attempt to identify specific demand situations and compute one representative hour for each situation, potentially weighted by the number of hours with a certain situation. Such identification would rely either on manual identification or on clustering algorithms. Another option is to perform seasonal aggregations rather than only using one annual value per DSO to comapre different companies. We refrain from computing representative hours or seasonal output in this report as the data made available for the test cases covers different times periods. Hence, any selection of a representative hour or seasonal output would no longer be directly comparable.

3.3. Input from the reference group

This section summarises the feedback received from the reference group of Norwegian grid about their interpretation of the task of a DSO. The major concerns that were mentioned with respect to accurately representing cost drivers and the possible implementation in an energy distance parameter will be discussed.

3.3.1. Maximum local vs. total demand

The power distance, irrespective of the chosen computation method, is defined as the distance power travels from point of production to point of consumption, multiplied by the instantaneous amount of power transferred. We have earlier in this chapter argued that the maximum load that a distribution grid has to cover is directly accounting for the investment costs. However, a grid company could serve areas in which the demand peaks never occur at the same time. For example, there could be a grid serving business districts, summer houses, and winter cabins. These areas will experience peak demands at completely different times. It is therefore important to consider the difference in modelling maximum power demand on nodal level, and maximum total power demand for a full grid area. This concern is addressed by investigating the hour of maximum total demand and the nodal maximum demand.

3.3.2. Installed capacity vs. actual demand

In some situations there is a gap between the power the grid companies are obliged to deliver to a certain consumer and the power they actually deliver. This means that the considered fuse size for building a grid extension would not be reflected by the actual supplied demand. In these situations, the local annual or seasonal peak loads will not be representative for the investments incurred by the

grid company.

As all grid companies are subject to the same regulations and laws, it is likely that this issue is as relevant for one as the other. In the DEA, the power/energy distance parameter will merely serve as a means to compare the efforts made by each grid company in relation to its costs, and the actual value of the parameter is therefore only relevant in ranging the companies from least to most efficient. As long as we are confident that this is an issue that is faced in an equal matter by all grid companies, it is not necessary to account for the gap between installed capacity and actual demand. If it is the case that the gap is notably more relevant for one grid company than another and creates systematic differences between DSOs, it would be relevant to account it, inside the benchmarking parameter or separately.

3.3.3. Accounting for historic demand distribution

Grid infrastructure is a long-term investment, and once built it usually serves the system for several decades. Most lines are built to serve the demand that the system is currently facing with some respect to projected development plans in a more long-term horizon. As time passes by, demand centres move, old industry closes down while new industry is established in another location. This dynamic could result in a system with sub-optimal infrastructure for the present task of the DSO. It often makes sense to build new grid based on old grid rather than to redesign the entire grid. How do we account for grid that has been built based on demand that no longer exists, such that its location/sizing is not optimal today?

An example of how extreme the consequences might be is a grid company being obliged to build a line serving a charging point for electric ferries. The required capacity is usually quite high, and so are the investment costs. It is also a high risk investment, as the charging point may be the only consumption served by the line. If some years later, the ferries switch from electricity to hydrogen, and the line becomes redundant, the grid company is stuck with a large investment and no corresponding demand that is accounted for in the power distance calculations.

It is likely that most grid companies have areas with sub-optimal grid infrastructure due to shifting demand centres. If this holds true, it is not a nuance that needs to be captured by the power distance. However, the magnitude of impact is likely to vary with the size of a grid area, which potentially creates higher risk for smaller companies. If it can be established that some grid companies are facing consequences due to old and sub-optimal grid infrastructure, we would recommend that the issue is accounted for by an adjustment factor outside the DEA calculation, as is done with weather or topography today.

The extreme case that a grid company has undertaken a large investment in a line that becomes redundant within a few years is not likely to be equally affecting all grid companies. Even if in many cases the majority of these investments are covered by the customer, the full investment is considered when calculating the RME efficiency. One way to account for these high-risk investments, is that RME could allow the grid companies to report the lines separately and allow the network companies to recover the full costs of these lines without implications for the benchmarking (similar to the way some regional grids are regulated). However, this is also clearly an administrative issue as it would require extra control and monitoring on the part of RME.

3.3.4. Costs vary with type and area of construction

In the power distance calculation the investment costs in new assets is considered through the α-parameter, that accounts for the scaling of costs with line capacity/transferred power. We argue that installation and operation and maintenance costs will be similar for assets of different capacity.

However, the initial investment cost and the operational expenses are also strongly location dependent. Location specific conditions, such as topology, existing settlements and ease of access will affect the choice of line and therefore the total cost. An underground line in an urban area will not incur the same costs as an overhead line of the same capacity in a sparsely populated area. In this situation the costs for the cable, which scale exponentially with the capacity, are the same but construction and maintenance costs will vary greatly.

This issue can be addressed through a topology correction that is either introduced ex-post or in the computation of the power distance. An ex-post approach could resemble the current geographical adjustments used in the DEA benchmarking. A possibility to include these constraints directly in the power distance calculation could be the use of different α-parameters for different line types. However, as described in Section 2.7 computing the α-parameter is a challenging task and defining different values per asset type may lead to a less accurate result than an adjustment after the initial benchmarking.

3.3.5. Demand peaks of less than one hour

This is a situation that is relevant for a lot of new demand entering the system. The most familiar example is fast charging of electric vehicles, particularly public transportation, and ferries. These charging stations require high instantaneous power, but the demand cycles are less than one hour. This is an issue as the demand data available for the power distance calculation is reported as hourly values in kWh/h, for the time being.

One example is a fast charger for an electric ferry shuttle leaving port every 20 minutes. Once arriving at the port, it charges for 5 minutes, and then leaves for the next destination. 15 minutes later, the next ferry arrives, charges for 5 minutes, and moves along. Within one hour, the fast charger has only been active for 15 minutes. The maximum power demand is 4 times as high as the reported kWh/h, and does not reflect the size of the investment in grid capacity.

As long as the demand data is reported on an hourly basis, it is difficult to account for the actual peak demand at locations such as the one described above. It could be possible, as we have already suggested for high capacity grid serving high-risk demand, to report these lines differently, and not include them in the DEA-calculation for the distributional grid. Another option is to somehow flag the relevant demand location, and report the maximum output as it should be a quite constant and known value. If metering data of higher resolution is available the infinity norm could be computed on the maximum demand for each node.

The two concerns of short demand peaks and high risk investment describe changes in the energy system that calls for new parameters describing the task of the grid company. In the current regulation, these investments rarely lead to an increase in the output parameters (number of customers, km of line, number of substations), while they have a large impact on the costs incurred by the grid company.

3.3.6. Demand decreases

Another issue is related to demand decrease, whether general or due to a specific customer reducing its power consumption due to e.g. bankruptcy, relocation or switch to other fuels. With the current DEA model the output would remain unaffected (except in the cases where the number of customers is reduced), while with a power distance variable the output is likely to be reduced, perhaps significantly.

In some cases, connection charges may reduce the risk for the network company as the total cost of an investment is partly covered by the charge. However, the benchmarking may still be affected negatively with a power distance measure instead of line length. Also, the connection charges generally do not apply to investments in

the meshed grid, although changes to the connection charging rules have been introduced from 2019 to include a share of the costs in the meshed grid as well.

It is not immediately clear how demand decreases can be accounted for within a power distance-based output measure or whether it has significant impact overall. One may suspect that there are systematic differences between network companies with respect to underlying demand trends, but this has not been investigated as part of the current project. One option could be to treat singular cases separately in line with the asset-based regional grid DEA model, but this may on the other hand entail large administrative costs for RME and the network companies.

3.3.7. CENS and losses

Finally there is the question of whether Cost of Energy Not Supplied (CENS) and/or losses should be considered in the benchmarking. While these cost elements are a part of the total cost input included in the DEA model, it may also be discussed whether they should be considered on the output side as well. Networks with long line lengths and high power distances will naturally have higher losses, and they will be more exposed to failures and resulting outages that incur CENS costs.

Higher CENS and losses result in lower efficiency as measured by the DEA model, which is necessary for the incentives of network companies to weigh costs of network investments against reductions in CENS and losses optimally. In the current model, the line length may be an acceptable proxy for losses and outage probability in grids with long lines, and a power distance measure may have similar or even better properties. It is outside the scope of this project to consider solutions to this, but RME should consider whether further parameters or adjustments are necessary. It should also be noted that the geography correction in the current model may reflect CENS and losses related to line lengths as well as other factors.

4. Test cases

To investigate the different computation methods for a benchamrking parameters, the grid companies participating in the reference group of this project have provided data from real cases in their distribution grid. For reasons of data privacy and security all grid companies are kept anonymous and each test case will be assigned a letter. All cases are limited to substations under up to three transformer stations from the regional to distribution grid level. For each substation hourly demand and generation values are provided. The demand and generation per customer in the low-voltage distribution grid is aggregated at the connected substation. The grid data is made available by RME from a centralised database.

4.1. Description of test cases

The following sections give a short description of each area and provide an overview of the grid topology per test case. The grid maps show all stations and lines for which data was available from RME. Note that some lines and stations, marked in light grey, could not be connected to the grid system either because they are served from another transformer station or because the grid data was missing information on connections. A general overview of the number of substations and transformer stations as well as the total length of lines per test case are given in Table 4.1.

Table 4.1.: *Summary of grid details per test case*

Test case	# of trafo stations	# of sub-stations	total line length
Test Case A	1	268	426 km
Test Case B	1	321	202 km
Test Case C	1	148	109 km
Test Case D	1	106	149 km
Test Case E	1	132	86 km
Test Case F	3	398	540 km

Test Case A - Large geographical area with more production than consumption

Test Case A is a large geographical area served by one transformer station from the regional grid level. The test case is characterised by both central and rural areas. It holds several small scale hydro power plants, and there are hours within the year with net higher production than consumption in the area.

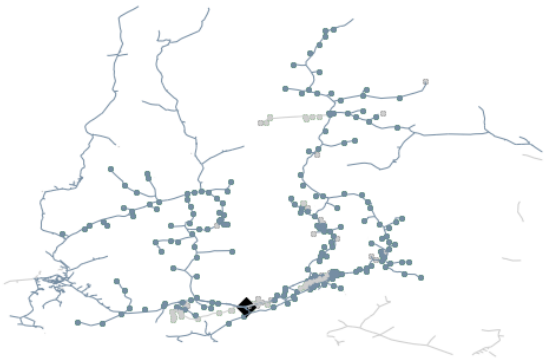


Figure 4.1.: *Grid topology for Test Case A*

The area covers around 270 substations with a total of 4700 metering points. Demand is mostly residential, holiday homes, offices and service buildings. There are no larger industrial consumers in the area.

Compiling this test case from the available data required some manual handling of missing lines to the transformer station. The missing data or mismatch of coordinates requires the addition of line elements which do not necessarily represent

the real system. Even after adding lines, unconnected line elements remain in the test case, this is visualised with different colours in Figure 4.1.

The provided demand data covers the full year 2018 in hourly resolution for all nodes in the test case.

Test Case B - Combination of dense demand centres and rural areas

Figure 4.2.: *Grid topology for Test Case B*

Test Case B is an area served by one transformer station from the regional grid level. It covers both central areas with higher demand density, and more rural areas. There is no large-scale production within the selected case area, but there are customers with excess production and a small scale hydro power plant. Hourly demand data per node was provided by the DSO for a time period from May 2018 to June 2019.

The area covers 320 substations and one transformer station. The large number of substations required more computation time for each parameter compared to other test cases.

Test Case C - Dense urban demand centre

Served by one transformer station from the regional grid level, this area is a dense demand centre with residential buildings, a sports stadium, and several office and service buildings.

The area covers around 2600 metering points

Figure 4.3.: *Grid topology for Test Case C*

that are aggregated to approximately 150 substations. The DSO provided two extensive datasets of hourly data; one with a full year of metering data for 2018 and another for three months from March 2019 to end of May 2019. For the calculations in this project we chose to use the data for three months in 2019 as the dataset had consistent data for all substations and hours whereas there was some data missing in the full 2018 data which could have affected comparability.

In the dense demand centres the grid has a large number of meshes and parallel lines. Despite these challenges the grid data could be compiled into a complete system. Certain lines, marked in grey were not connected to the test case grid, either because they are served from another transformer in the region or because of missing lines in the data.

Test Case D - Rural with decentralised production and planned locations for shore power

Test Case D is served by one transformer from the regional grid level. It is a rural area that is subject to typical challenges due to electrification of the transport sector and decentralisation of power production. The area is characterised by:

- A distributional grid that is mainly serving production sites
- Power production on lines that are built to serve end consumers

Figure 4.4.: *Grid topology for Test Case D*

- Several planned locations for shore power
- Power exchange with other grid companies

The area covers 150 substations and 1100 metering points and the demand data covers a year from June 2018 to June 2019 in hourly resolution.

Test Case E - Suburban region

Figure 4.5.: *Grid topology for Test Case E*

This case covers a suburban area served by one transformer station from the regional grid. Demand is distributed densely around the transformer stations with some outliers in all directions. The area covers residential areas, office and retail buildings without distributed generation or large demand centres. The area covers 130 substations and 5000 metering points. A full year of hourly demand data was provided for this test case from January 2018 to end of December 2018.

Test Case F - dense rural areas and power intensive industry

The test case represents a geographically bounded area in the company's grid. It is mainly served by one transformer station from the regional grid level, and partially served by two additional substations. The area is rural, with small densely populated areas, holiday homes as well as some industry.

The area covers 390 substations and three transformer stations. The provided hourly demand data covers a full year from July 2018 to July 2019. The data for grid lines in the RME database differed from the other test cases as each line was saved as a composition of individual short line elements. The resulting large number of lines and lack of information of connection points posed major challenges when compiling the grid data.

4.1.1. Demand patterns

Each grid company in the reference group provided demand data on hourly level for each substation in the test cases. The aggregated hourly demand profiles are shown in Figure 4.6 on page 43. Note that the time period of provided data varies from test case to test case. All test cases cover a full year of demand but have different start and end dates, except for Test Case C for which a dataset with three months of demand data was used. The resulting profiles show seasonal patterns of high demand in winter and lower demand in summer. Minima in demand can either be related to generation in the system that outweighs demand (as for Test Case D) or errors in measurement data (likely the case for Test Case E). Without further information on the provided demand data we choose to directly use the input data without adjustments of possible errors.

4.2. Data quality

The quality of available data and the ease of the necessary data handling are essential foundations for any approach that is to be used for regulatory purposes. In the following, we discuss our observations and learnings gathered from importing the test cases. We discuss grid data and demand data from Advanced Metering Systems (AMSs) separately.

4.2.1. Grid data

RME has collected a database of all grid elements in Norway. The database includes the location of metering points, substations, transformers, and lines and cables of all voltage levels.

This database was the main source of grid data for the project. There are big advantages of using data from one source, as it allows to use the exact same algorithms to import data for each of the test cases. Despite some of the grid companies offering to provide grid data in their own format, we found it preferable to rely on the RME database to ensure consistency and ease of handling.

The grid data from the RME database does not have information about which line is connecting a certain substation to a transformer station. We therefore received a selection of line data from the high voltage distribution grid covering the area of each test case. We were then faced with the task of identifying which of the lines are relevant for the power distance. While compiling the test cases, we came across a number of challenges related to the data. Handling these issues involved significant manual work and was highly time-consuming.

The following gives an non-exhaustive overview over typical issues:

Missing lines Missing line elements, even if they are short, are hard to solve. It is often not evident where, or indeed whether, an open line end is connected to some other location. In individual cases, the missing connections could be identified on a map and manually connected. Such an approach is not applicable in large test cases or for the entire Norwegian distribution grid.

Lack of connection information The line data only consists of the line geometry, but does not have any information on connection points. It is often easy to assume that two line ends are connected if they are in the same location. This approach is challenged in a number of situations. 1) Substations and transformer stations are represented by a point, and the endpoints of connecting lines are often registered in some distance. 2) T-junctions of lines are only recognised if all lines have an endpoint in the junction location.[1]

Line length The length of each line is determined from the start and end coordinates. The resulting length is the euclidean distance between two points and does not consider differences in length due to topology or line type. Lines that are not a direct connection between two substations are stored as individual line segments that need to be connected as previously described. In the current DEA benchmarking the length of lines is reported by each DSO which is not necessarily the same as the length extracted from the RME database that is used for the power distance calculation.

Voltage level definitions The database provides the voltage level for most (not all) assets. While this is generally useful, the voltage definitions differ and are not clearly communicated. For example, one line might be defined as 21 kV, another as 22 kV depending on whether typical operation voltage or maximum admissible voltage is given.

Different meta data Even though all data comes from the same database, it does not always hold the same meta data, or uses different tags for the same meta-information. For example,

[1] checking for any connection between all lines is computationally very challenging, whereas checking closeness of endpoints of lines is reasonable

object IDs are 1) called different (`objektID`, `lokalID`, `globalID`, `objekt_ID`, ...), and 2) hold different types of IDs across data sets.

Differences in segmentation The data in some grid areas was much more segmented than in others, meaning that a line connecting to points would be split into individual line elements, each stored as an individual line. In other grid areas, the line would be stored as one long geometrical object. Where lines are split in many small elements, this may overload the algorithms for connecting them.

Different data management systems Even though the grid data was extracted from a common database, the original data was provided to RME by each grid company from their data management system. There are currently several different data management systems in use that can comply to RME's data requirements of geographically accurate grid data. However, the line data may vary between systems leading to discrepancies between line segmentation, object IDs and line lengths. By adapting the requirements for the provided grid data to the power distance computations, some challenges that were experienced in this project could potentially be overcome.

4.2.2. AMS and station data

Data on demand and generation per substation was provided by the grid companies for each test case. To our knowledge, the data was extracted from Elhub or internal measurement systems and aggregated per substation.

The provided data was generally of high quality and easy to handle. Challenges, however, occurred in matching the demand data with the RME grid data. The main findings are

Elhub The centralisation and standardisation provided by the metering data hub Elhub facilitates efficient data extraction. Some companies in the reference group explicitly pointed out that they found it much easier extracting the required profiles for the period covered by Elhub than for the preceding period.

Good data quality Where smart-meter data was available, the data quality was generally very good, with only few missing data points, usually at the beginning or end of the data period which could correspond to old, pre-Elhub data or very recent data. Some extreme outliers were also identified and only removed if it could be justified.

Data privacy Data privacy laws give strict rules on data handling and data provision. Aggregating data per substation alleviates the issue, and generally best practice and regulation on data handling, non-disclosure, and limitation of data access and data storage have to be followed. Considering that metering data hold much information about the customers, these measures are unavoidable. We did not find them to impede the project.

Comparability The demand and generation data was provided individually per DSO for each test case. Consequently, data from different time periods was used as input in the calculations. This may impact comparability between the test cases as the peak demand can differ between years, and datasets that cover less than one year will not capture all patterns and extremes.

Station ID The ID for substations differed between the RME grid data and the internal data from DSOs. In some cases ID matching was provided by the DSOs whereas other cases required coordinate matching which will generally be less accurate.

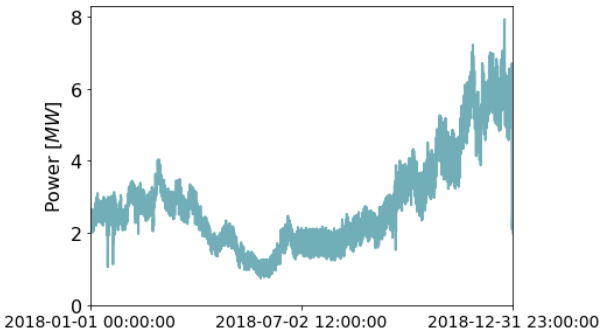
Matching substations To obtain a full dataset of geographical coordinates, connected lines and demand per substation the AMS data has to be matched with the RME grid data. Even when challenges of identifying station IDs were overcome not all stations from the AMS datasets could be found in the RME grid data. This could imply that new stations were added to the grid

that are not yet included in the RME database.

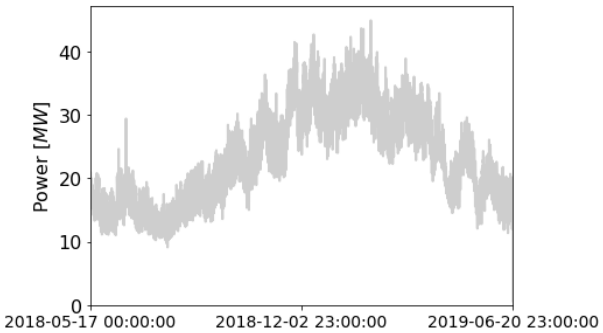
4.2.3. Implications of data quality

The quality of input data for the benchmarking methodology is crucial to ensure an objective comparison between grid companies. Data quality issues may even influence which power distance parameter is most promising for regulatory purposes. Some considerations that should be taken into account are:

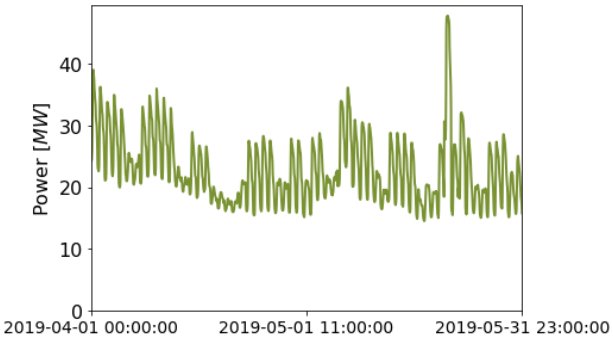
- We found that in many cases, grid data is missing or requires extensive manual work. This may pose serious challenges for any grid-based power distance parameter.
- Where line data is missing, there might be situations where provision of this data is detrimental for the benchmarking performance of a grid company. This would constitute a disincentive to provide high-quality grid data.
- Where station IDs cannot be matched, the associated demand might not be considered in the power distance computation. Here grid companies have an incentive to provide consistent station IDs and to ensure that all data is up to date.
- Should demand data due to privacy consideration only be accessible on substation level, this might imply an advantage for 220 V low-voltage distribution grids – as discussed in Section 2.6.



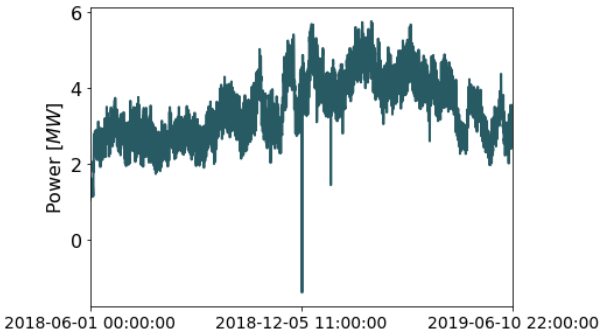
(a) *Hourly Demand Test Case A*



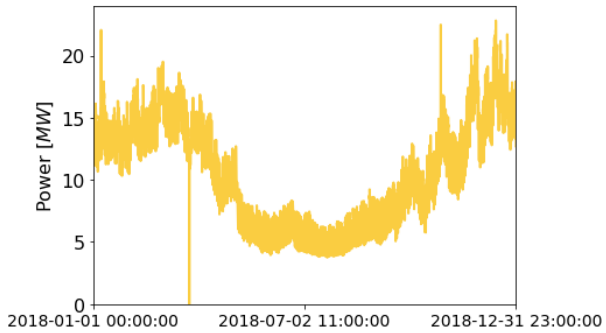
(b) *Hourly Demand Test Case B*



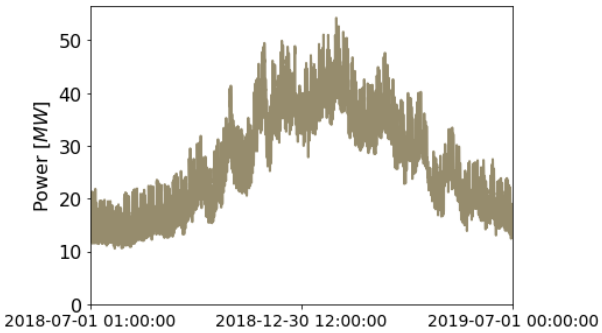
(c) *Hourly Demand Test Case C*



(d) *Hourly Demand Test Case D*



(e) *Hourly Demand Test Case E*



(f) *Hourly Demand Test Case F*

Figure 4.6.: *Hourly demand for the six test cases*

5. Results per test case and per parameter

The following section presents the results per power distance computation method for all investigated test cases. For all calculations per test case an α-parameter of 0.4 was used. An analysis of different values for α is given in Section 5.6.

5.1. Minimal power distance

In the provided test cases the number of meshes in the test cases was too high to allow for a computation. As a result, we were not able to compute the minimal power distance for any of the test cases.

As mentioned in the description of the method, the computational complexity grows exponentially with the number of meshes. This means that for our test cases it is not a question of parallelisation or improvements in computer technology – with over 10×10^{110} required iterations it is simply not possible to find a solution.

Some room might remain for reducing the number of meshes, such as more precise algorithms to recognise parallel lines, a consistent split of the problem along voltage levels, etc. However, it is unlikely that even such approaches will lead to a computable solution in all cases and for all distribution grid topologies found in Norway.

We therefore have to conclude that the minimal power distance cannot be used as a benchmarking parameter, and will not further investigate it in this study.

5.2. Power-flow based power distance

The power-flow based power distance is calculated based on the real grid, including all meshes in the system. The computation determines the flow on each line segment and finally calculates an hourly power distance parameter. The power-flow based power distance could be calculated for all test cases except for Test Case F. The use of the real grid in the computation requires data of high quality and a significant amount of data handling to account for unconnected substations or missing lines. In the case of Test Case A several straight line segments had to be added manually to obtain a closed grid system. In the case of Test Case E the transformer station needed to be manually connected to the grid. For the urban grid in Test Case C simplifications of parallel lines could be applied but required additional efforts. For the case of Test Case F the grid data could not be compiled into a complete grid system. Each line in the Test Case F grid was saved as a number of individual line segments without information on connections between lines and substations. The algorithms used for all test cases to establish connections of

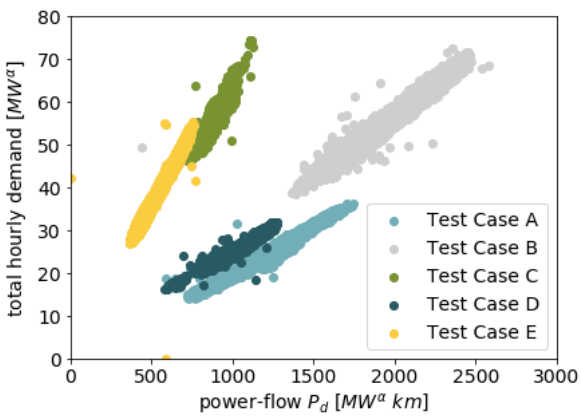


Figure 5.1.: *Comparison of power flow based power distance and the total hourly demand to the power of alpha, for $\alpha = 0.4$*

lines and stations in geographical proximity could not compute a complete grid in this case.

Based on the demand data per substation and the system compiled from the grid data, power-flow based power distance computations were performed. The hourly results for all test cases is shown in Appendix Figure 7.1. Generally a flattening trend can be observed compared to the hourly demand data. This flattening effect is a result of applying the alpha parameter which is below 1 and therefore reduces the intensity of demand in the calculation. Overall, the seasonal trend can still be observed and hourly variations are visible according to demand peaks.

The power-flow based power distance can be compared to the hourly demand in each system. This is illustrated in figure 5.1 with the power-flow based power distance on the x-axis and the demand to the power of alpha (D^{α}) on the y-axis. For each test case a linear relation can be observed, which is to be expected from the formulation of the power distance in Equation 2.8. The differences in slope is due to the differences in line lengths in the system. For systems with longer lines, the power distance, that scales with the length of lines to each substation, will generally be larger than for cases with shorter line length and meshes that offer multiple paths for power to flow. As a result, the more urban areas, Test Case C and Test Case E show the steepest relation between demand and power flow based power distance. Test cases with less meshed grids and overall longer lines show a less steep relation, as for example Test Case D and Test Case A.

For the remaining methods we will compare the results to the results of the power-flow based power distance calculation. We consider the power-flow based power distance to be the most realistic representation of the DSOs task based on the existing grid infrastructure. Ideally other power distance parameters should show a linear relation to the power-flow based power distance to qualify as a good proxy.

5.3. Artificial grid power distance

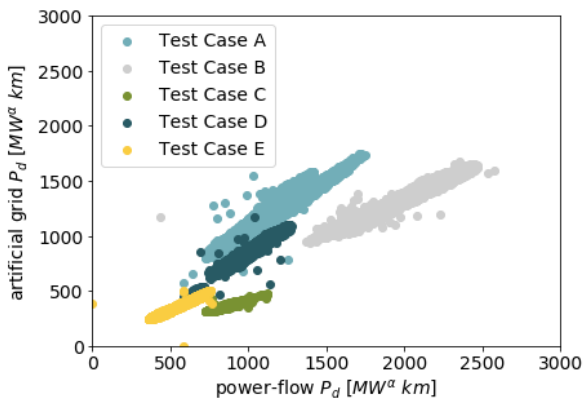
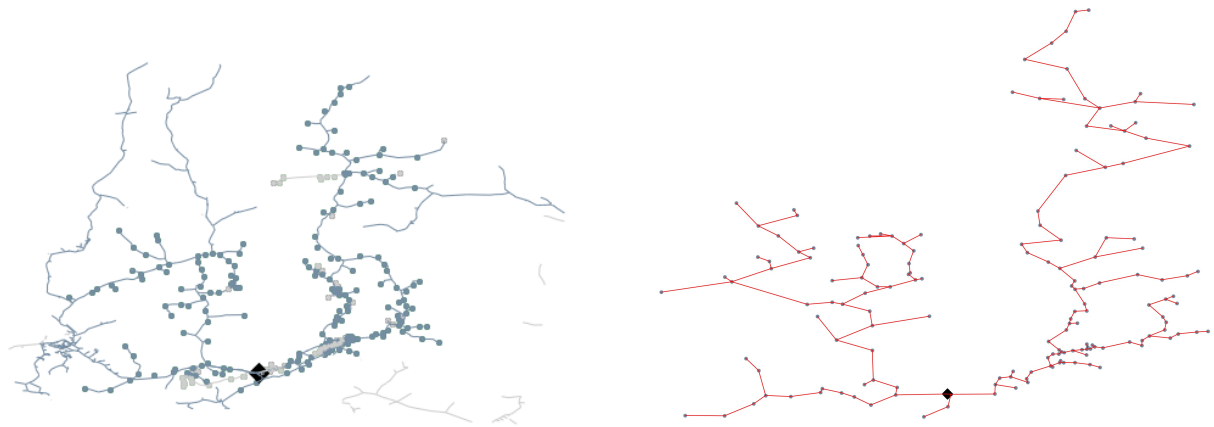


Figure 5.2.: *Comparison of power flow based power distance and artificial grid based power distance, for* $\alpha = 0.4$

The artificial grid only uses the coordinates and demand per node and does not rely on the grid data. The lower data requirements made it possible to calculate the artificial grid based power distance for all test cases. For each case, an artificial radial grid is built based on a representative hour from which the hourly power distance can be computed by using a power flow calculation. For test cases with more than one transformer station, individual grid systems are created around each transformer station. The nodes in the system are assigned to the closest transformer station and the total power distance is a sum of the power distances in each sub-system. Lines that are included in the real grid but do not connect to a substation are not constructed in the artifical grid.

Figure 5.3 gives a comparison of the real grid and the artificial grid for the Test Case A. It is clearly visible, that the artificial grid is a radial system with no meshes. Generally the artificial grid resembles the real grid well, with the exception that meshes are not accounted for, which underlines the suitability of the computation method for the purpose of constructing an idealised grid.

The hourly results for all test cases are shown



(a) *real grid for Test Case A* (b) *Artifical grid for Test Case A*

Figure 5.3.: *Comparison of real and artificial grid for Test Case A, for* $\alpha = 0.4$

in Appendix Figure 7.2. As for the power-flow based power distance, a flattening compared to the demand data can be observed. The intensity of the resulting power distance, however, varies with the test cases. A comparison of the power-flow based power distance and the artificial grid based power distance is given in Figure 5.2. For all investigated test cases the artificial grid based power distance and the power-flow based power distance show a linear relation. This indicates that the artificial grid based power distance is a good proxy for the power-flow based power distance.

Differences between the test cases can be observed in the ratio between the two parameters. Generally the artificial parameter is smaller than the power-flow based parameter which can be attributed to geographical and topographical constraints in the real grid. The artificial grid will always create direct connections between two nodes and the resulting grid will be radial without meshes. As a result, the total length of lines in the grid will be shorter than in the real grid used for the power flow calculation. The ratio is closest to 1 in rural areas with long lines and few meshes, such as Test Case D for which lines are located on the coast of a fjord. In urban areas, where the grid is typically more meshed and existing infrastructure constrains the building of direct lines between two nodes, the power-flow based parameter can be more than twice as high as the artificial grid based power distance.

We further investigated the difference between the power-flow based power distance and the artificial parameter by comparing the total length of lines in the real and artificial grid (see Table 5.1). For all cases the artificial grid has a shorter total line length than the real grid. This is a result of meshes in the real grid, lines that are built around obstacles and lines that do not end at a substation. The difference in total length can give a first indication of the differences in power distances but does not fully explain the relation between the power distance in the real and artifical grid.

Therefore we additionally investigate the av-

Table 5.1.: *Comparison of total length in real and artificial grid*

Test case	L^{real} [km]	L^{art} [km]
Test Case A	426	183
Test Case B	202	118
Test Case C	109	28
Test Case D	149	106
Test Case E	86	33
Test Case F	540	102

erage relative line length. This means the average ratio of the actual line length in the real grid compared to the euclidean distance between two substations. The ratio between the two lengths can be a indicator of topographical differences and could be used to adjust the artificial grid based power distance. Table 5.2 compares the ratio of the power-flow based power distance and the artificial grid based power distance (P_d^{flow}/P_d^{art}) to the ratio of total line length in the real and artificial grid (L^{real}/L^{art}) and the average relative line length ($\sum_{ij} l_{ij}^{real}/l_{ij}$). For all test cases, except for Test Case A, the ratio of power distances is inversely proportional to the ratio of real and artificial line length. This supports the hypothesis, that grids with generally shorter lines in the real grid are less well represented by the artificial grid and the artificial power distance is underestimated compared to the power-flow based power distance.

A similar trend can be deduced from the relation between the power distance ratio and the average relative line length – the closer the line length ratio is to 1 the closer the results for the power distance are. This can be interpreted that the artificial grid, with its connections along the euclidean distance, resembles the real grid closely and thus results in a similar power distance output for the power flow and the artificial grid based method. Test Case A is an outlier in this series which may be due to the fact that the grid has large meshes despite substations being very spread out. As a result power can be distributed to the most distant substation along several paths with lower flow per path compared to the artificial grid, where only one connection is established. The difference to the results of the remaining test cases highlights that other factors than the difference in line lengths may impact the result of the artificial grid computation. A clearer investigation of other potential impacts can be performed in future work when more and larger test cases are made available.

The differences in ratio between the two parameters could be corrected by a geographical factor. The investigated test cases are only a limited sample of grid areas in Norway and are thus not representative for determining geographical or topographical correction factors. From a larger data sample and detailed geographical input it could be possible to investigate how the ratio changes with the conditions in each area. As a result a geographical adjustment, similar to that used in the current DEA benchmarking, can be defined.

Table 5.2.: *Ratios of lengths in the real and artificial grid*

Test case	ratio P_d^{flow}/P_d^{art}	ratio L^{real}/L^{art}	ratio $\sum_{ij} l_{ij}^{real}/l_{ij}$
Test Case A	0.95	2.32	1.07
Test Case B	1.48	1.71	1.17
Test Case C	2.34	3.89	1.27
Test Case D	1.17	1.41	1.07
Test Case E	1.51	2.60	1.35

5.4. Demand distribution based power distance

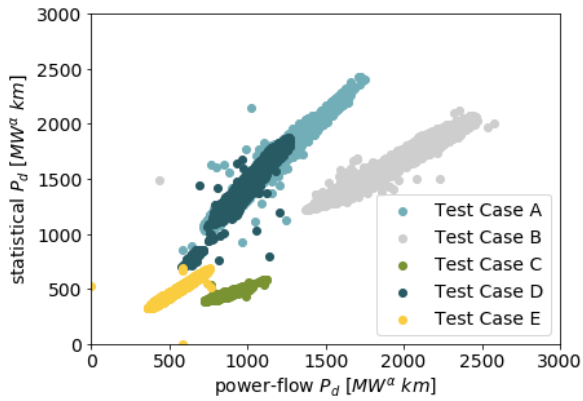


Figure 5.4.: *Comparison of power flow based power distance and demand distribution based power distance, for $\alpha = 0.4$*

The demand distribution based power distance does not rely on a real or artificial grid. Instead it uses the coordinates and demand per node to calculate a distribution of demand around each

transformer station. This method could be applied to all test cases without limitations due to data quality or availability. The hourly results for all test cases is shown in Appendix Figure 7.4.

The demand distribution based power distance and the power-flow based power distance show a linear relation for all test cases. As in the previous comparison for the artificial parameter, the ratio between the two power distance parameters varies between test cases (see Table 5.3).

Table 5.3.: *Ratios of statistical to power-flow based power distance*

Test case	ratio P_d^{flow}/P_d^{art}
Test Case A	0.70
Test Case B	1.17
Test Case C	1.89
Test Case D	0.69
Test Case E	1.12

In urban areas with shorter distances between nodes and more meshes in the system, the statistical power distance is lower than the power-flow based power distance. Similarly to the artificial grid, the statistical parameter does not account for meshes and indirect connections. As a result, the overall parameter is lower than the power-flow based power distance. For less dense demand distributions the statistical power distance is larger than the power flow based power distance. This can be attributed to the fact that nodes may not be in angular proximity according to calculations but are most efficiently connected along one line e.g. in Test Case D along a fjord. As a result, the demand distribution based power distance counts nodes with individual connections over long distances whereas the real grid connects these nodes over one line for which the total length is shorter.

5.5. Comparison of yearly output parameters - energy distance

The power distance parameters described in the previous sections are an hourly output parameter for each grid system. To obtain a fair benchmarking parameter the hourly values need to be aggregated into one or more annual values. As discussed in Section 3 different cost drivers in a distribution grid can be represented by different aggregation parameters. Specifically the 1-norm, that accounts for the average demand or power distance and the infinity-norm that accounts for

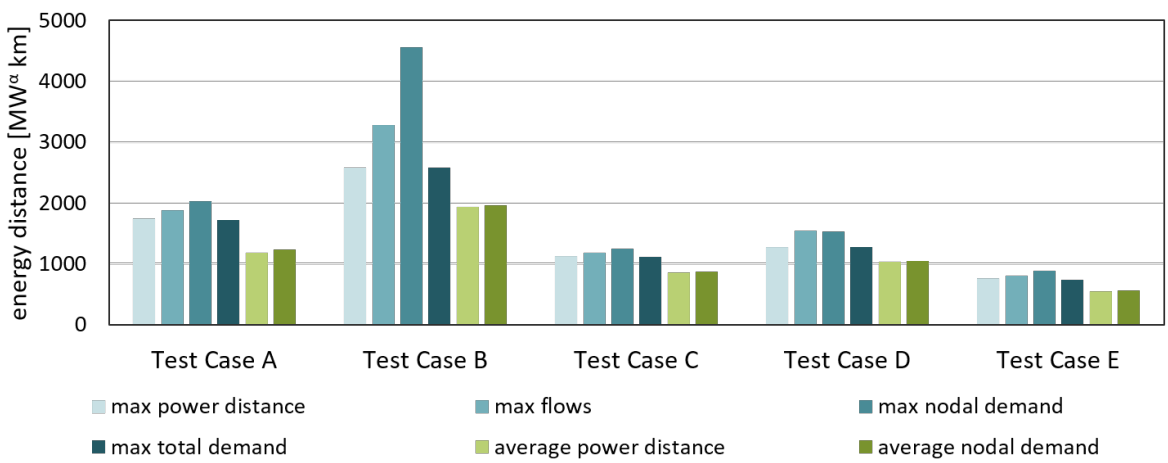


Figure 5.5.: *Comparison of infinity-norm and 1-norm energy distance parameters for different DSOs, from the power-flow based power distance for* $\alpha = 0.4$

the maximum. In this section we will investigate different aggregation methods for the hourly power distance parameters to an annual energy distance based on the test cases. General trends can be deduced from the results but absolute comparisons need to be made with caution, as the time periods of provided data vary. All aggregations will be performed based on the power flow based power distance.

The following aggregation methods are investigated:

1-norm:

average P_d: The average power distance in the year. Determined after calculating the power distance for each hour.

average nodal demand: The power distance calculated for the average demand at eaxch substation in the system. The power distance is only calculated once.

∞-norm:

max P_d: The maximum power distance in the year. Determined after calculating the power distance for each hour.

max total demand: The power distance calculated for one hour with maximum demand in the system. The power distance is only calculated once.

max nodal demand: The power distance calculated for a system where each node is assigned its maximum yearly demand. The power distance is only calculated once.

max flows: The power distance calculated for a system where each line element is assigned its maximum yearly flow. The power distance is only calculated once.

Figure 5.6 shows a comparison of the six mentioned methods for all test cases where the power-flow based power distance could be calculated. The two 1-norm parameters are fairly similar for all DSOs with the nodal average being slightly higher than the average power distance. Since the demand is scaled with the alpha parameter its impact in the power distance is lower than in the original demand data. Taking an average prior to applying the power distance calculation will thus lead to a higher energy distance than performing an averaging operation after calculating the power distance.

For the infinity-norm parameters the max nodal demand is consistently the highest followed by the max flows. The nodal and flow maximum tend

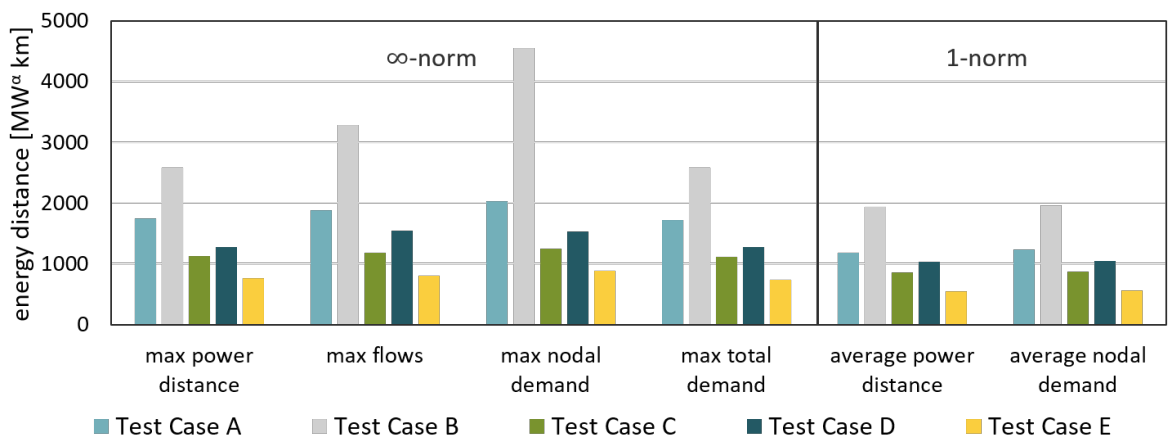


Figure 5.6.: *Comparison of energy distance parameters for different methods of defining the maximum and average value, from the power-flow based power distance for $\alpha = 0.4$*

to overestimate the real situation in the grid as the demand peaks per node or flow peaks per line do not occur in the same hour of the input data. Note that the maximum flow aggregation can only be performed for power distance calculations that are based on a grid, real or synthetic. For larger systems, as for example in Test Case B, the difference between the max power distance and the max nodal demand is higher than in smaller systems.

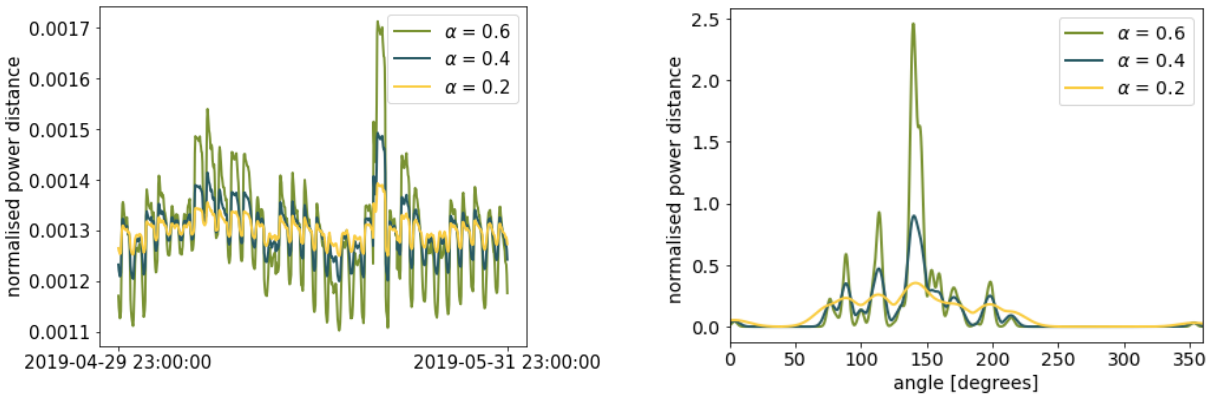
This can be attributed to the fact that the probability of the maximum demand per node to occur at the same time is lower for larger systems. Therefore the difference between the hour of maximum demand and the maximum nodal demand will be larger. This is more clearly visible in Figure 5.6 where the largest system (Test Case B) sees the largest difference between the maximum aggregation methods. The max power distance will be at least the same as the max demand – for the max power distance a power distance is calculated for each hour of demand, including the hour of maximum demand. The hour of highest demand does not have to correspond to the hour of the highest power distance. It can occur that in the hour of highest demand the largest demand peaks occur close to the transformer station, resulting in a comparably low power distance, whereas the hour of the highest power distance may have more spread out demand.

5.6. Comparison of different alpha parameters

In the previous sections, results were presented for an alpha parameter of 0.4. The chosen value is based on the different investigated methods to determine the α-parameter which give a range between 0.3 and 0.5 (see Section 2.7). In different computation methods, the alpha parameter can have a different impact, which will be discussed below.

Power-flow based power distance

For the power flow based power distance computation the alpha parameter has a flattening effect compared to the demand input. For larger alpha the resulting power distance will more closely resemble the demand profile through its peaks and valleys, whereas for lower values the impact of the demand is lower in the overall result. This is illustrated in Figure 5.7 (a) for the Test Case C. An alpha of 0.6 sees large daily peaks while the



(a) *Hourly power flow based power distance of Test Case C*

(b) *Demand distribution for different alpha parameters in one hour for Test Case D*

Figure 5.7.: *Comparison of results for different alpha parameters*

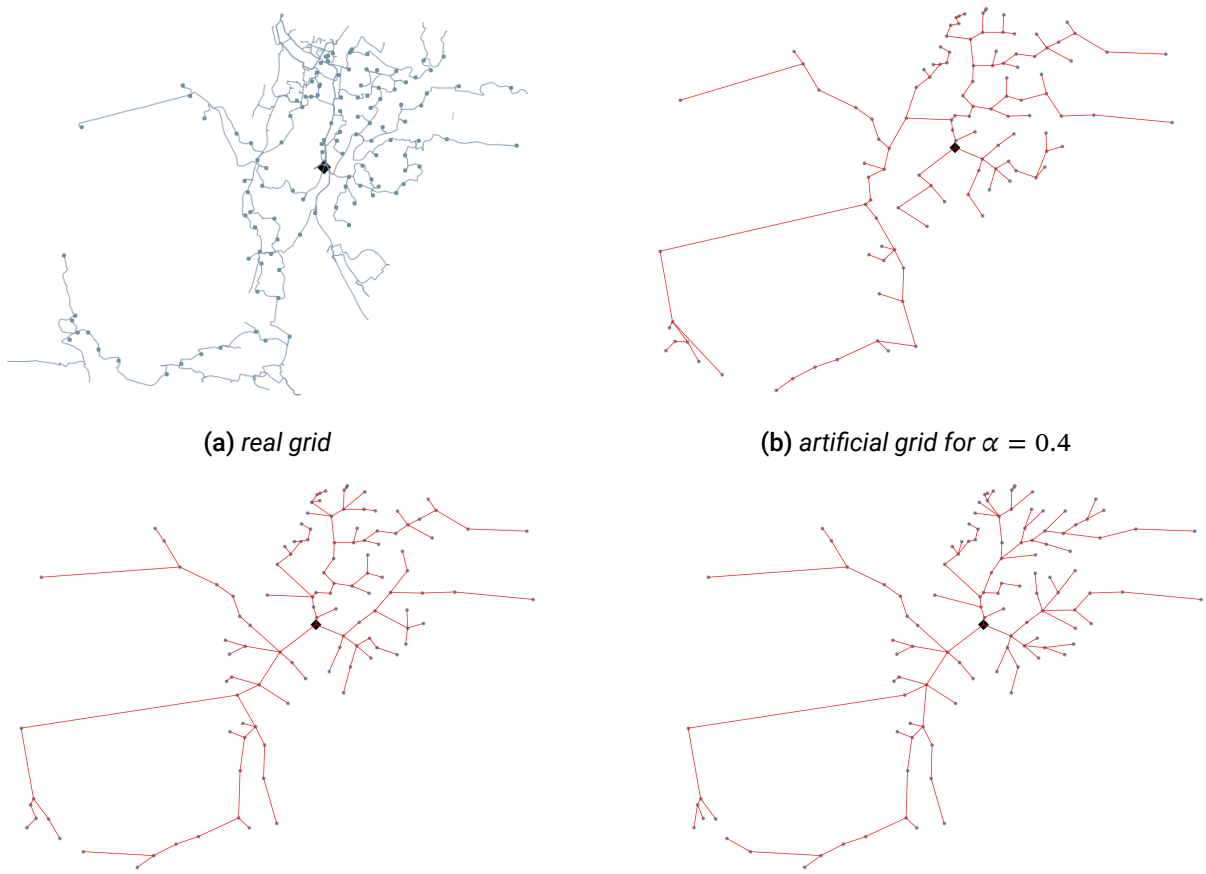
results for alpha = 0.2 are generally flatter. In the most extreme cases an alpha of 0 would lead to a flat curve independent of demand and an alpha parameter of 1 would see the same variations as the demand input.

Demand distribution based power distance

The demand distribution that is calculated to obtain the statistical power distance depends strongly on the alpha parameter. Most importantly the alpha parameter is used to determine the standard deviation of the normal distribution that is created around each demand node. For lower alpha parameters the individual distributions per node will be broader, thereby creating more overlaps with other nodes in angular proximity. This signifies that the nodes can be more efficiently connected over one line with high capacity. As alpha is increased the peaks become sharper and less overlaps are visible, which indicates that individual connections to each of the nodes are preferred. Figure 5.7 (b) shows a normalised distribution for different alpha parameters. The sharpest peaks are visible for an alpha of 0.6 while the distribution for α = 0.2 is smoother and almost shows no distinct peaks.

Artificial grid based power distance

When the artificial grid based power distance is calculated the alpha parameter impacts the output in two ways: (1) through the scaling of demand in



(a) *real grid*

(b) *artificial grid for* $\alpha = 0.4$

(c) *artificial grid for* $\alpha = 0.6$

(d) *artificial grid for* $\alpha = 0.8$

Figure 5.8.: *Comparison of real and artificial grid for different alpha parameters for Test Case E*

the power distance calculation and, more importantly, (2) through the creation of the grid based on the minimal increase in power distance. Figure 5.8 shows a comparison of the real grid of Test Case E (top left) and the idealised grid that is established by the algorithm for different alpha values of 0.4, 0.6 and 0.8. It is immediately visible that the artificial grid has a less complex topology than the real grid. At first sight differences in the areas with a large density of substations are not as visible. One can, however, quickly identify that the connections in the lower left part of the grid differ greatly between the real and the artificial grid for all alpha parameters. While the real grid connects all nodes via a long U-shaped sequence of lines, the artificial grid builds two individual lines. Which nodes to the far left are connected to each of the two branches also differs for different alpha parameters. Depending on the value of α the connection that increases the total power distance by the least changes.

The largest differences between the artificial grids created based on different alpha parameters are visible in the top right part of the grid. In this area the real grid is characterised by a large number of meshes. Generally, lower alpha parameters incentivise building more lines with higher capacity. As a result, there are more common connections in the artificial grid for α = 0.4. For a higher alpha parameter building more lines with lower capacity is more efficient. This dynamic is clearly visible in the artificial grid for α = 0.8 where a large number of lines branch out and supply individual nodes. Almost every substation is connected by its own line in the α = 0.8 grid, while the α = 0.4 grid has more lines that serve multiple nodes. Overall, a grid that is established with an alpha parameter of around 0.4 resembles the real grid the most, despite being radial.

The calculation of the power distance will strongly depend on the grid topology which determines the length of each line and thus the path along which power can be transferred to each node. As a power flow calculation is used to determine the flow on each line and then compute the total power distance, the demand data is influenced by the α parameter in a similar way as described for the grid-based computation. A lower alpha parameter reduces the total output and has a more flattening effect.

Energy distance

For the yearly output parameter, α influences the result through the hourly power distance calcula-

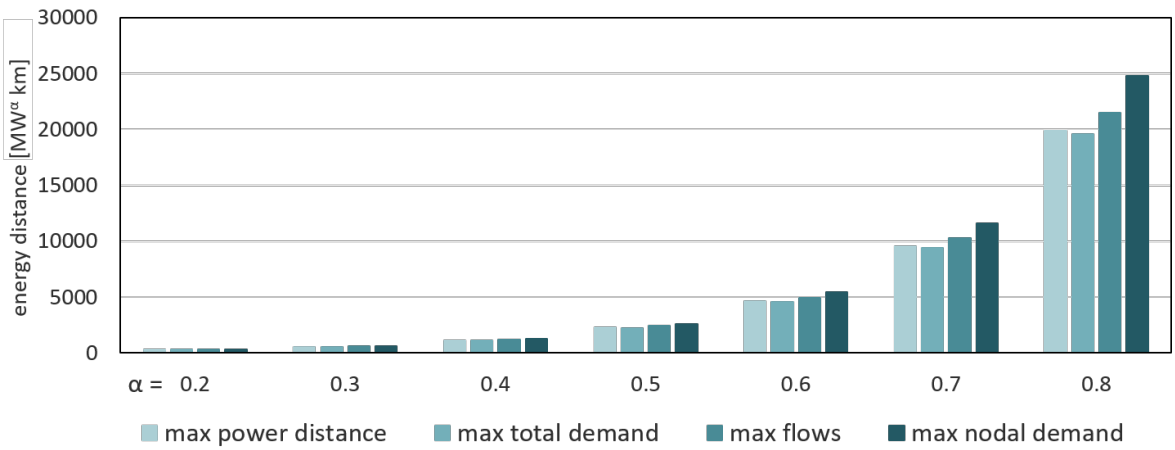


Figure 5.9.: *Comparison of infinity-norm energy distance parameters for different alpha parameters for Test Case C*

tions. Figure 5.9 illustrates how the output for the four infinity-norm aggregation methods changes with the alpha parameter in the Test Case C. For all aggregation methods, an increasing trend with alpha can be observed. A higher alpha will generally increase the power distance per hour because the demand is flattened less compared to an alpha parameter close to zero. As the impact of demand becomes smaller with a smaller alpha parameter it becomes less likely that the hour of maximum demand and the hour of maximum power distance overlap. On the other hand, a larger contribution from demand in the multiplication with the line length potentially leads to a larger spread between the hour of maximum demand and the hour of maximum power distance. Whether the hour of maximum demand will also result in the highest power distance depends most on the grid topology and therefore differs from case to case. The maximum nodal energy distance is consistently the highest output and increases more strongly with a higher alpha parameter. The exact increase is again determined by the grid topology where alpha has the largest impact on nodes that are located far from the transformer station. The long line length will be multiplied with a larger value P^{α} as alpha increases towards one. The maximum flow energy distance is always higher than the maximum hour and lower than the maximum demand. The increase with alpha occurs in the same scale as for the maximum nodal demand, related to the impact of the P^{α} term with higher alpha.

6. Evaluation

In Section 1.3 we described the most important factors for a suitable DEA benchmarking output. Optimally, an objective parameter should be highly exogenous in representing the task of the grid company with low requirements for data and computational effort. Additionally, the parameter should be intuitive to understand and comparable between grid companies with different external conditions. After computing the power distance using four methods for several test cases we can analyse the suitability of each parameter based on the mentioned key factors. In this section we compare the main results for the different parameters and discuss feedback from the reference group on the results. Table 6.1 summarises the suitability of each power distance parameter with respect to these factors.

6.1. Discussion of key factors

6.1.1. Task representation

A benchmarking parameter should ideally reflect the task of the DSO rather than its effort. The task of each grid company is to supply power to cover their customers demand at any time and location. An output that reflects this task should consider the demand per customer and its geographic distribution to account for the incurred cost for transporting power over long distances. The existing grid infrastructure takes the effort of the DSO, i.e. the investments in infrastructure that the grid company takes to supply power to all customers, into account. Of the investigated power distance parameters, the minimal and power-flow based power distance use the existing grid to investigate how power can be ideally distributed to cover the demand at each node. These approaches use the existing infrastructure, which represents the DSOs effort, as well as the distribution of demand, to reflect the task of the DSO. The artificial grid based and demand distribution based power distance are independent of the existing grid and thus do not account for the effort of each grid company. Instead these methods only consider the task of the DSO to supply demand to each node in the system, either based on an idealised grid or a statistical distribution. The current DEA benchmarking uses the length of line, the number of substations and the number of customers as output. The former two reflect the effort of the DSO while the latter partly represents the task of the DSO. However, the main component of a grid company's task lies in supplying power to cover demand for their customers. The number of customers does not reflect the total demand or its distribution.

The annual energy distance parameter should reflect the task of the DSO and the different drivers for the cost that a grid company is exposed to as a result of fulfilling its task. As discussed in Section 3 different aggregation methods for hourly data can reflect several of the most important cost drivers. We suggest using several yearly parameters to account for the investment costs, the cost of delivering energy and administrative costs. The cost of delivering energy should be reflected by an average value, for which the average yearly demand per node or the average power distance in a year are similarly suitable. For the administrative costs we propose to keep the number of customers in the DEA model. The representation of the investment costs is not as straight-forward. The DSO should equip its grid so that it can handle the maximum load in the system. The maximum load can be simply interpreted as the hour of maximum demand in a year or the maximum power distance.

Table 6.1.: *Evaluation of key considerations for benchmarking parameters*

	$current\ DEA$	P_d^{min}	P_d^{flow}	P_d^{art}	P_d^{stat}
Task representation	medium	very good	very good	good	good
Exogeneity	low	medium	medium	high	high
Data requirements	very low	very high	very high	medium	medium
Computational effort	very low	too high	high	medium	medium
Intuitiveness	very high	high	high	high	low
Comparability	medium[a]	very high	very high	high[a]	medium[a]
Overall suitability	medium	low	high [b]	high	medium

[a]Before geographical correction
[b]Constrained by data availability

However, the maximum demand in one year does not necessarily reflect the total possible maximum that could occur in a system. We therefore consider it more suitable to account for the maximum demand per substations or alternatively the maximum flow per line segment, for methodologies that use a grid. If the low-voltage distribution grid is also accounted for, the maximum demand per substation should be the sum of maximum demands per customer.

6.1.2. Exogeneity

The concept of exogeneity in the context of a power distance parameter refers to an output that is not under control of the benchmarked entities, i.e. independent of actions taken by the DSO. The task of the DSO should be to build and maintain a grid that optimally supplies power to their customers, not to adapt the grid to the optimal output in the benchmarking and accordingly receive higher revenues. Consequently, an ideal benchmarking parameter should be entirely independent of the grid company's investment or operational decisions. In the current DEA benchmarking the length of lines is a parameter that is under control of the DSO. This could, in theory, lead to grid companies building longer lines than necessary to improve their benchmarking output. Similarly building more substations could have a positive impact on the output for a DSO. The fact that the presently used parameters are under control of the DSO was one of the main concerns that motivated the definition of new output variables.

The grid-based power distance parameters are still partly under control of the DSO, as the existing grid infrastructure is used as input. Through the choice of the alpha parameter an incentive to investment in more substations than needed can be avoided. However, building longer lines than required may be incentivised by these grid-based variables as it can lead to an improved benchmarking output for the grid companies. In contrast, the grid-free variables are highly exogenous and cannot be influenced by the grid companies. For the parameters where only demand per substation and substation location are used as input, the grid companies should not be able to impact customer demand and their location. The location of substations can, in theory, be adapted by the grid companies to improve the benchmarking output. However, the power distance computation for the high voltage distribution grid should be complemented with an appropriate aggregation of demand per customer in the low voltage distribution grid around each substation. Through the choice of demand aggregation method in the low-voltage distribution grid, the incentive for selecting a substation location based on the benchmarking output rather than system efficiency can be removed. Generally, the position of a substation is constrained by multiple factors related to geography and external conditions, thus

an incentive to place a substation differently does not mean the different placement will be technically possible. The annual energy distance parameters will be as exogenous as the hourly power distance computation that is used.

6.1.3. Data requirements

In the years since the introduction of the current DEA benchmarking variables, large improvements in available data have been made. Hourly demand and generation measurement data from smart-meters can be obtained from Elhub and RME has built a comprehensive database of the grid infrastructure in Norway. These data inputs open new possibilities for including more detailed data in the DEA model but also pose challenges in terms of data quality and data handling. Generally, the benchmarking output should not pose unrealistic data requirements on the grid companies or the regulatory authority. The data used should be of high quality to avoid any adjustments that may distort the final output. For a power distance parameter the requirements on data quality should be high enough to correctly account for the (hourly) demand distribution but low enough to be fulfilled from existing data without extensive (manual) adjustments. For any power distance parameter, demand data per substation and substation location are crucial input. We found that geographical locations of substations could be easily extracted from the RME grid data and hourly demand data from Elhub could be used directly. Despite some challenges in matching the station data with the corresponding demand, we consider the quality of available data sufficient for the purpose. Hence, the grid-free power distance parameters, i.e. artificial grid and statistical output, could be calculated without limitations from the available data.

For power distance parameters that use the existing grid large amounts of data needed to be provided on the grid infrastructure. This data was made available as individual line segments and substations which had to be compiled into a complete grid system per test case. Despite the high quality of available data, we encountered several issues when compiling the grid data (see Section 4.2). At present, we do not consider the available data to be sufficient for an objective benchmarking through a grid-based power distance parameter.

The computation of an annual energy distance output does not pose any additional data requirements compared to the power distance computations as long as consistent data in hourly resolution is made available for each substation.

6.1.4. Computational effort

The benchmarking output needs to be computed for each grid company in Norway on annual basis. Thus, it should be possible to compute any output that is used in the DEA benchmark with reasonable computational effort. In the proposed methodologies in this report, it was shown that the minimal power distance could not be solved in a reasonable time for any of the test cases. The fact that the minimal power distance relies on an iterative optimisation process, makes it increasingly complex with larger and more meshed grid systems. Logically, a parameter that cannot be computed for every topology in the high voltage distribution grid does not qualify as a benchmarking output. For all other methodologies the computational effort was no limitation in calculating a power distance output. There were differences in the number of required calculation steps and the computation time but all parameters could be obtained in a reasonable time. Computing hourly values for any power distance parameter can be a time-consuming task but if the result for one hour can be calculated, repeating the task for each hour in the year and performing annual aggregations is not constrained by the computational complexity. We generally believe that a lower computation time should not come at the expense of the output quality. Since the computation is typically only performed once a year, we consider any parameter that can be computed within a reasonable time to be suitable for the purpose.

When considering the complexity of a compu-

tation all calculation steps need to be taken into account. For the power distance computations this includes any data handling and adjustments needed to compile the grid data.

6.1.5. Intuitiveness

A thorough understanding of the methodology behind the revenue regulation is crucial for the regulatory authority that applies the benchmarking model, the grid companies, whose revenues depend directly on the output and potentially large customers, who are affected by the resulting grid tariffs. While an ideal benchmarking parameter should capture the factors that affect the task of the DSO in detail, it should not be too complex for affected entities to understand. In the case of the Swedish NPAM a lack of documentation and general understanding of the methodology was one of the reasons that the regulatory change was not well received by the grid companies. A similar outcome should be avoided if a new benchmarking parameter is introduced in the DEA model. In our experience and from the feedback received from the reference group, grid-based parameters are easier to understand than those that do not rely on the real grid. In the context of grid regulation, it is intuitive to refer to the existing topology and consider how power can be transported in a given grid system. Though the optimisation process to compute the minimal power distance is highly complex for a meshed system, the general concept of removing line segments until a radial grid remains can be easily understood. For a radial system the computation of flows per line segment is simple. For the power flow based method the general methodology is similarly simple to grasp as it is based on the real grid topology and physical flows – two concepts that anyone working in or with a grid company will likely be familiar with. The calculation of individual flows per line segments can be less intuitive, as there are several paths along which power can flow in a meshed grid. Overall we consider the grid based parameters to be most easily understood.

The concept of an artificial radial grid and the computation of power flow in the resulting system is slightly more complex. The power flow computation in the artificial grid itself is relatively simple, as there are no meshes in the grid. But the creation of the idealised grid based on the minimal increase in power distance is more challenging to understand. The methodology operates in a way that the connections between nodes are not established merely based on minimal distance but also account for the cost-scaling of lines with capacity and the remaining demand in unconnected nodes. While this approach creates a more cost effective grid than a grid based on minimal distances, it is also less intuitive to understand. Nevertheless, the increased complexity did not pose problems to the participants in the reference group, who developed an understanding of the concept of the artificial grid based power distance.

The demand distribution based power distance is the most abstract parameter analysed in this report. Since no real or artificial grid is used, the methodology is less intuitive to understand. The output is of a similar format as for the (real or artificial) grid-based parameters but calculation steps to obtain a power distance output are founded on statistical concepts rather than definite paths of power flow. Generally, the concept of angular proximity and its application in the demand distribution based power distance was considered complex to understand by some members of the reference group.

6.1.6. Comparability

In order to objectively becnhmark the performance of different grid companies the benchmarking output should be comparable between areas with different external conditions. As opposed to the Spanish RNM, where the efficiency of each DSO is determined through benchmarking a real grid against an artificial grid, the Norwegian DEA model evaluates relative efficiency between all grid companies. Hence, a benchmarking parameter should be as objective as possible for each considered grid company.

Parameters that are based on the real grid will be highly comparable between grid companies because the external conditions will be reflected both by the total incurred cost and the resulting grid topology. Differences in demand distribution, topographical conditions and area constraints e.g. through water bodies or natural parks, are indirectly accounted for through the layout of the real grid. In contrast, the artificial grid based power distance and the demand distribution based power distance do not consider differences in topography between grid companies. The impact of different topographies for rural and urban areas could clearly be identified in Figures 5.2 and 5.4 for the artificial and statistical power distance respectively. We suggest using a geographical correction factor to these parameters ex-post. Such a correction factor could be determined in a similar manner to the current DEA benchmarking for the artificial grid and based on a ratio of area types for the demand distribution based power distance.

For the annual aggregation of the power distance into an energy distance parameter, comparability is also a crucial factor. The number of customers, to account for administrative costs and the average demand per grid area, to reflect the cost of energy delivered, are comparable between DSO for all methods. For the infinity norm, i.e. the situation of maximum load, the selection of the maximum demand impacts the comparability of the output. For a larger grid area the probability that all nodes experience their maximum load at the same time will be lower than for a small region with few substations. Despite the low probability, such a situation may occur and the respective grid company will need to equip the grid accordingly. However, when using the hour of maximum demand or the maximum power distance in a year DSOs with small areas and few substations may have an advantage compared to those who need to serve larger areas. We therefore propose to use maximum demand per node or the maximum flow per line for the calculation of the infinity norm. By using these values a ”worst-case” demand scenario is simulated which should reflect the investments necessary to operate the grid and is therefore comparable for all DSOs.

6.1.7. Incentives

Each proposed method to calculate a power distance and subsequently an energy distance may create different incentives for the grid companies. Ideally a benchmarking parameter should not disincentivise any of the following; investments that improve security of supply, providing of accurate and updated data, building a line in the shortest/most efficient path, among others.

Some possible implications were already touched upon in the previous paragraphs and will be described in more detail while others are yet to be discussed. The current DEA benchmarking uses the length of lines and number of substations as input. Theoretically DSOs can improve their benchmarking output by building more substations and longer lines than what would be necessary. By taking the demand of customers and their location into account power flow parameters aim at removing these incentives. The incentive of building longer lines may still be given for the power flow based power distance and the minimal power distance, especially for substations that lie at the end of a grid line. All considered methods may incentivise placing substations in a way that maximises the output. However, this can be avoided when the regulation in the high voltage distribution grid is complemented with an appropriate aggregation of demand per customer in the low voltage distribution grid. Generally, we deem it unlikely that grid companies will change the position of substations to improve their benchmarking output.

For methodologies that use a radial grid, i.e. the minimal power distance and the artificial grid based power distance, investments that improve security of supply by building a meshed system will not be beneficial for the benchmarking output. The methodologies will always identify a radial system as the most efficient and would not promote a system that aims to improve security of supply. This

was also demonstrated in [5]. For the artificial grid based power distance and the statistical power distance any investment that does not connect a new substation will not affect the output, as the artificial grid is created independently of actual grid topology and the statistical calculation does not account for any grid. As a result, lines that do not affect the number, location or demand of substations are not accounted for.

When grid data is used for a computation it has to be ensured that grid companies are incentivised to provide the most accurate and recent data. Some examples for when this may not be the case are when manual adjustments of missing lines improves the output compared to the real grid situation, when missing demand data in hours of low load lead to a higher annual output, when the addition of a new substation between two transformer stations leads to different possible flows in the grid and not reporting the station leads to a better relative performance.

6.2. Feedback from the reference group

All investigated methods and the results of the test cases were discussed with a reference group of seven grid companies in cooperation with RME. The exchanges with Norwegian DSOs offered insights into common concerns and potential challenges. The following is a non-exhaustive list of feedback that we received from the reference group:

Historic grid It was mentioned that using a grid-free parameter would not account for historic grid developments. In many cases historic assets may not be the most efficient solution in today's grid but were necessary investments at the time of construction. In this context it was discussed that this may apply to all grid companies to a similar extent.

Alpha parameter A participant mentioned that the cost of lines will not scale with the line voltage, as this is often restricted by the existing grid. Instead the most relevant variable is the cable diameter. This motivated our approach to calculate α from the cost catalogue.

Exogeneity The involved grid companies underlined that it was not common to plan assets based on the benchmarking output as this may change over time while infrastructure is usually in place for a long period. Consequently, exogeneity in an output parameter is desired but should not be the main focus.

Investment cost The investment cost in a grid will depend on the maximum load that needs to be covered from each customer even if these do not occur at the same time. This feedback indicates that using a nodal maximum demand for the infinity norm could be the most suitable approach.

Cost of losses The cost of losses does not depend on the peak demand but rather on the fluctuations in demand. At the same time the length of lines is already a proxy for the losses in a system. This topic should be investigated further to determine whether it is sufficient to account for a maximum and average situation or whether and additional measure of demand fluctuations should be included.

Intuitiveness Several participants in the reference group agreed that the power flow based power distance is the most intuitive and that the demand distribution based power distance is complicated to understand.

Individual challenges Several grid companies outlined challenges that they were facing in a specific situation. An example was the charging of electric ferries which leads to high demand peaks for a shorter duration than one hour, which would not be captured by calculating hourly power distance values. Another case was the investment in a connection to a large customer that may disconnect from the grid in the future. The costs would need to be covered by the DSO but if no demand is

recorded at this customer, the power distance would not capture this investment. These individual concerns are addressed in more detail in Section 3.

7. Recommendations and conclusions

From the investigations performed in this project we conclude that a relevant measure for the task of a DSO is the electric power distance within its grid area, given an obligation to cover all loads and to handle power fed into the grid from distributed generation. The power distance can be computed through different methodologies to obtain the relevant distances over which the power must be distributed to satisfy the exogenous demand. For any computation method, the magnitude of the power distance strongly depends on the underlying distribution of demand, and therefore captures that the task of supplying power to clients depends on the demand distribution.

The following section analyses each power distance parameter as well as the annually aggregated energy distance with respect to applicability and regulatory implications. We then present our recommendations and outlook for further work.

7.1. Discussion of applicability and incentives per parameter

7.1.1. Minimal power distance

The minimal power distance is calculated from the flows on all lines in the minimal grid that is required to supply demand within a given topology. It is based on the assumption that an optimal grid that minimises the electric power distance will always be a radial system, or a tree in a graph theoretic sense. An iteration over all mesh elements in a grid allows to find a globally optimal solution for any grid system. This approach was demonstrated on test grids in [5].

In this project the minimal power distance could not be computed for any of the test cases due to high computational complexity. For the optimisation the complexity scales exponentially with the number of meshes in a system. In the investigated test cases, and likely any real grid test case, the optimisation could not find a solution.

Besides computational challenges, the data requirements and need for data handling are very high for the minimal power distance. Data on all lines, transformers and substations, including hourly demand and generation need to be available in high spatial resolution. Despite the high quality of available data from Elhub and RME, manual adjustments were needed to compile complete grid systems and match stations with demand. For a benchmarking parameter on national level, we deem it impossible to perform any manual adjustments without affecting either the consistency of the output or the required effort.

The minimal power distance reflects the distribution of demand in a given grid topology and is a more exogenous parameter than the current output in the DEA benchmarking (length of lines, number of customers). Nevertheless, the parameter is not fully exogenous as the grid topology is partly under control of the DSO through investment decisions.

It was also shown in [5] that the minimal power distance as a benchmarking parameter may give disincentives to invest in grid reinforcements that improve security of supply. Since the resulting minimal grid will always have a radial topology, any investment that creates a mesh in the grid will be considered inefficient.

Despite its benefits compared to the current benchmarking output, the computational complexity of this method makes it impossible to calculate a minimal power distance for all high voltage distribution grids in Norway, thus disqualifying it as a

benchmarking parameter.

7.1.2. Power-flow based power distance

The power-flow based power distance uses the full meshed grid to compute a physically optimal power flow along each line. Power can flow along any line segment in the grid, for which line specifications are kept constant, to supply the demand at every substations.

By using the real grid topology and considering physical flows, the power-flow based power distance gives a good representation of the task of a DSO to supply power to each customer through the existing grid infrastructure. The distribution of demand around each transformer station is accounted for by the multiplication of flow per line with the line length. In contrast to the minimal power distance, the power-flow based power distance does not remove meshes from the grid and instead uses the full meshed system for determining the optimal power flow. As a result, investments in lines that enforce the grid and thereby create meshes will be considered in the power-flow based power distance.

Similarly to the minimal power distance, data requirements for the power-flow based power distance are challenging. The methodology requires a full dataset of lines and demand stations in a grid with details on geographical location, line length and hourly demand/generation. At present, the grid data still requires some additional adjustments to connect lines, substations and transformers and to remove parallel lines. These adjustments can be automated based on typical connection patterns and require limited computational effort. However, any adjustment will lead to a less realistic representation. Applying any kind of adjustment to the input data can also lead to disincentives to provide the most accurate and updated data for the grid companies. In this project, some data issues could not be handled through automated processes. Examples are the manual addition of missing lines to connect free substations, assigning substations to transformer stations and matching station coordinates to the demand data.

For the case of Test Case F, the provided grid data could not be compiled to a complete grid. Consequently the power flow based power distance could not be computed. A situation were the benchmarking parameter cannot be computed for specific grid companies or grid areas disqualifies the parameter for the application on national level.

Furthermore, the use of real grid data influences the exogeneity of the parameter. The topology of the real grid will always be under control of the DSO to some extent.

Technically line specifications such as line voltage and resistance could be used as an input to the power flow computation to give an even more realistic representation of the existing grid. However, we consider it difficult to collect consistent data on line specification from all grid companies. This may lead to a distortion in the benchmarking output or an incentive to provide less accurate data. Instead we would argue, that all DSOs face similar constraints from the grid properties and not accounting for line parameters will not affect the relative benchmarking strongly.

Overall, the power-flow based power distance is well suited to represent the task of the DSO and provides improved exogeneity compared to the current benchmarking. For the time being, the requirements on data quality cannot be sufficiently fulfilled to ensure an objective benchmark. We propose to aim at improving data quality to an extent where it becomes possible to use the power-flow based power distance for regulatory purposes. This can be a focus of further work from the side of the regulatory authority.

7.1.3. Artificial grid based power distance

The artificial grid based power distance computes the flows on each line of an idealised grid to supply the demand at all nodes. The motivation behind

this method is to have an exogenous parameter that can be computed with low data requirements. The artificial grid based power distance only uses the location and demand per substation as an input. These data are presently available from Elhub and RME making it possible to compute the output for any given grid area.

The artificial grid is created as a radial system, as the minimal required grid to supply all nodes is always a tree-like topology. Unless new demand nodes are added to the system, an investment in new or reinforced lines will not affect the output. This may result in disincentives to perform investments that improve security of supply.

The resulting parameter is highly exogenous, as neither the location of customers nor the demand is under control of the DSO.

The artificial grid based power distance does not account for geographical or topographical constraints, as the real grid would. This is made visible by the different ratios between power-flow based power distance and artificial grid based power distance for urban and rural test cases. We want to refrain from creating an overly complex idealised grid, comparable to the Swedish model, and instead choose to compute a simple artificial grid that captures the challenges of supplying demand to distributed nodes. Rather than including external constraints through topography, weather or natural reserves in the initial computation, we suggest performing an ex-post adjustment of such factors. Similar to the topography adjustments in the current benchmarking, the output from the artificial grid based power distance can be adjusted in an additional step. In Section 5.3 we described a first approach to compare the line lengths in the real and artificial grid as an indicator of topographical differences.

If such a correction factor can be identified, and the required data is more easily accessible than the grid-data needed for the power-flow based power distance, we suggest to use the artificial grid power distance in DSO income regulation.

7.1.4. Demand distribution based power distance

The demand distribution based power distance calculates an output parameter that is merely based on the demand and location of each node. By not using a grid altogether, the statistical parameter is highly exogenous and has low data requirements.

The only required input data is the location of each substation and the demand per node which are readily available from the grid database and Elhub, respectively. The computational effort is comparably low, as no grid needs to be created and no power flow calculation is used. A major limitation of the demand distribution based power distance is that it is not as intuitive to understand as grid based parameters. Using a parameter that is detached from a grid makes it harder to compare to the real grid.

The fact that only the nodes in a system are included in the calculation also means, that investments that do not create connections to new substations will not increase the output and are therefore not considered efficient by the benchmarking algorithm even if they improve security of supply or enforce the grid to supply higher local demand.

Similar to the artificial grid based parameter, the statistical parameter does not account for geographical differences or other limitations between grid areas. Potentially, a correction factor could be applied that investigates the ratio of area types, e.g. buildings, forest, water, agricultural land, in the grid area.

Overall, we consider the demand distribution based power distance to be a useful statistical parameter. For the application in a regulatory setting, however, it may be too complex to intuitively understand for the relevant entities. Further investigations on the robustness of the method and the exact implications of the alpha parameter are recommended.

We propose testing a similar approach to aggregate demand data from individual metering points to substation demand. In this context the statistical parameter could be a useful aggregation

tool that does not disincentivise 400 V lines in the low voltage distribution grid.

7.1.5. Yearly benchmarking parameter - energy distance

The output parameter(s) in the DEA benchmarking should reflect different cost drivers in building and maintaining the grid. To reflect administrative costs we propose to keep the number of customers as an output parameter. The costs for administrative tasks will not scale with transferred power or customer demand but is directly proportional to the number of customers. The number of customers is also an exogenous parameter that is not under control of the DSO.

To account for the overall cost of energy delivered we propose to use a 1-norm energy distance, which reflects the average demand supplied to each node. Two aggregation methods were investigated to calculate the 1-norm: either using the average demand per node and computing one power distance from it or calculating the power distance for each hour and taking the annual average. Taking the average of the demand prior to calculating the power distance gives a consistently higher result than the average power distance due to the scaling of demand in the power distance formula. The effect of the averaging operation is the same for every DSO and calculating the average before computing the power distance reduces the computational effort. We therefore suggest to use the power distance from the average demand as 1-norm energy distance.

To reflect the investment cost we have looked at maximum load situations in the grid. We argue, that a grid investment will be taken in a way that the maximum load in the grid can be covered at any time. In Section 5.5 we discussed that the maximum load situation in a grid can be interpreted differently. Either by considering the hour of maximum demand in the analysed year, the maximum power distance from the hourly output or the maximum load per line or node irrespective of hour. The use of the maximum flow per line or the maximum demand per node may overestimate the output parameter compared to calculation based on a real situation, i.e. max demand or max power distance. However it needs to be considered that just because the combination of high demands in several nodes did not occur at the same time in the considered year does not mean it cannot occur at other times. The grid should in any case be equipped to supply the requested demand to its customers. Thus an infinity-norm based on the maximum nodal demand or maximum flow per line is most applicable to reflect investment costs.

7.1.6. Alpha parameter

The alpha parameter describes the scaling of cost with line capacity. An alpha parameter between 0 and 1 reflects that it is more cost efficient to build one line with high capacity rather than two lines of half the capacity to supply demand.

The results shown in Chapter 5 are based on calculation with an alpha parameter of 0.4. Additionally we performed calculations for different alpha parameters in all methods. Generally a low alpha parameter leads to a flattening of the resulting power distance compared to the demand data. For the artificial grid, the alpha parameter is used as an input to build the grid with the smallest incremental increases in power distance. We have seen that an artificial grid created with an alpha of 0.3-0.5 resembles the real grid the most, which indicates that these values reflect the real investment conditions in a grid most accurately. For the demand distribution based power distance the alpha parameter influences which nodes are considered to be in angular proximity, i.e. for which nodes it is more efficient to be connected to the transformer through one strong line compared to two individual lines.

In this project, we investigated three methods to calculate the alpha parameter from available data. Our conclusion is that there is no definite answer as to which alpha parameter should be used. Instead we have identified a range of $\alpha =$

$0.3 - 0.5$ in which the alpha-parameter is likely to lie. In future work, the alpha parameter should be investigated in more detail with respect to the individual methods. It is possible that for different methodologies slightly different alpha parameters are most suitable.

7.2. Recommendations

In this project we have investigated four methods to compute a parameter that reflects the distribution of demand in the high-voltage distribution grid. The methodologies were tested on six selected test cases for representative grid companies in Norway to obtain first conclusions on their applicability. All recommendations in this section are based on a limited amount of data for a small number of test cases and are thus not necessarily representative for the distribution grid on national level. For more general recommendations these methods would need to be applied to full grid areas per DSO or ideally the entire Norwegian high voltage distribution grid. This should be the focus of future work on this topic, which will be further discussed in Section 7.3.

We consider the power-flow based power distance to be the most suitable parameter to represent the task of the DSO. The output parameter is also highly comparable between DSOs in regions with different external conditions and is intuitive to understand. Despite lower exogeneity than the artificial grid and demand distribution based power distance, the power-flow based power distance would be an improvement compared to the current benchmarking method. From the feedback of the reference group, we also conclude that exogeneity is not the most crucial factor in this context as grid companies are unlikely to adapt their investment decisions to influence the DEA output. Calculating the power-flow based power distance is associated with high computational complexity and high data requirements. While the former does not pose limitations to its applicability in the DEA benchmarking the data requirements disqualify the power flow based power distance from being used in a regulatory setting. Technically, all required data is available from Elhub and the RME grid database but compiling the data on individual grid assets into a complete system calls for complex handling and manual adjustments. We conclude that for the time being the available data is insufficient to reliably compute the power-flow based power distance for all grid areas in the Norwegian distribution grid. We recommend that additional effort is put into ensuring a high quality and consistency of data for the entire Norwegian distribution grid.

Should it not be possible to achieve a level of data quality that is sufficient to compute an objective power-flow based power distance, we propose to use the artificial grid based power distance with an ex-post geographical adjustments. The artificial grid based power distance has lower data requirements and computational complexity than the power flow based method. Additionally, it is highly exogenous and comparability between grid companies can be achieved through adjustment factors. For this methodology, special focus should be put on defining the representative hour and alpha parameter for which the artificial grid is created. Additional work, beyond the scope of this project, will be required to identify geographical adjustment factors based on larger test cases. Optionally it can be investigated to use an artificial extension grid, similar to the Spanish RNM brownfield model to account for historic grid infrastructure.

If only the high-voltage distribution grid is used to compute the power distance parameters, the distribution of demand in the low-voltage grid needs to be accounted for in a suitable way. The used benchmarking method should reflect the difference in required investment depending on the needed substations. This could be achieved by applying an approach similar to the demand distribution based power distance or more simply by keeping the number of substations as an output parameter, or a ratio of line length in different voltage levels. The choice of aggregation and output parameter in the low-voltage distribution grid will depend on data availability and computational complexity, which

should be addressed in further work.

In any case, we recommend testing all proposed methodologies on larger test cases to compare the respective output to the current DEA benchmarking. This will allow analysis of how each parameter reflects the actual investment costs that a grid company faces and how the relative efficiency of grid companies changes for different parameters, compared to the current benchmarking.

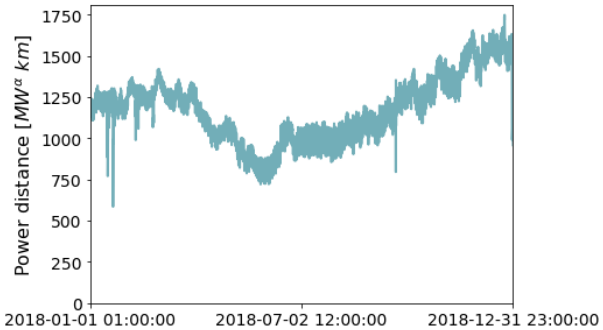
For the annual aggregation of hourly power distance parameters to a benchmarking output we propose to use a 1-norm, that reflects the average load situation, and an infinity-norm, that represents the maximum load situation. Additionally, we suggest to keep the number of customers as a DEA output to reflect the administration costs. For the 1-norm, we suggest using the average demand per substation and calculating one power distance from as it is the least computationally intensive approach. For the infinity norm, we believe that the investment cost is best reflected by a ”worst-case” situation of maximum possible load in a system. This is reflected by either the maximum demand per node or the maximum flow per line, irrespective of hour. The infinity-norm should be further investigated once larger test cases are made available to compare the output to the incurred investment costs.

7.3. Outlook

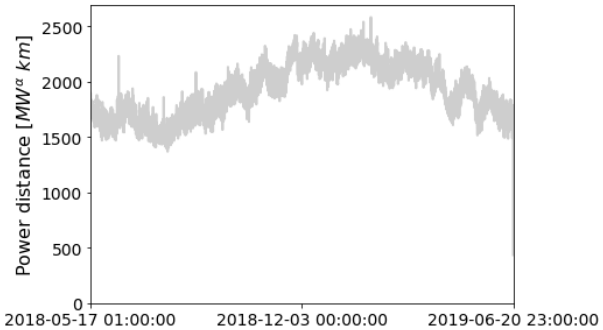
The focus of further work should lie in identifying advantages and disadvantages of each parameter and the interpretation of the obtained results in the context of regulatory implications. Though certain aspects were covered in this report, the proposed methodologies need to be tested on larger test cases and ideally all high voltage distribution grids in Norway. Calculating the power distance from large test cases will make it possible to compare the different methods to each other and to the existing DEA benchmarking model. From the obtained results, effects such as unwanted bias (e.g. urban grids are depicted less effectively than rural grids), investment disincentives (e.g. building additional lines to increase security of supply to an already connected customer can reduce the output) or limited exogeneity (e.g. the result of the power distance can be influenced by the DSO) can be identified. The target of this analysis should be to conclude which method is best suited for the purpose of DSO revenue regulation and which additional adjustments to the final benchmarking, for example a topographical correction factor, may be needed.

Appendix

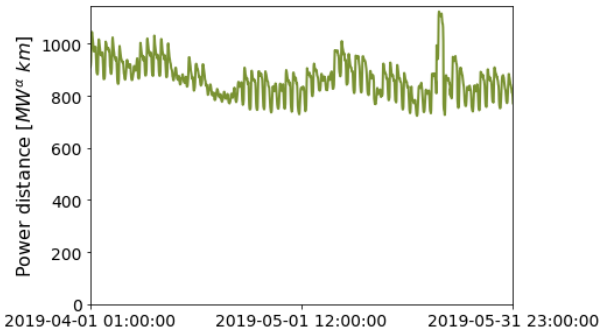
Power-flow based power distance results



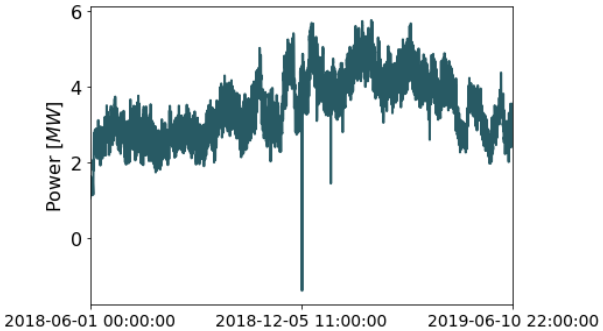
(a) *Test Case A*



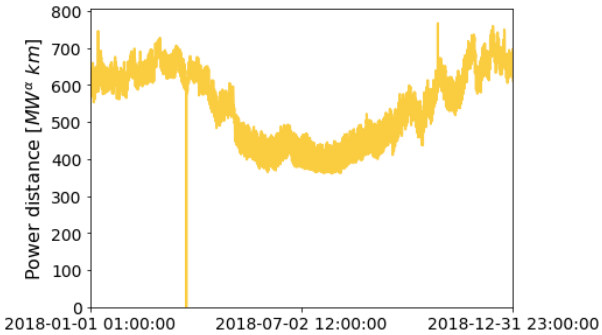
(b) *Test Case B*



(c) *Test Case C*



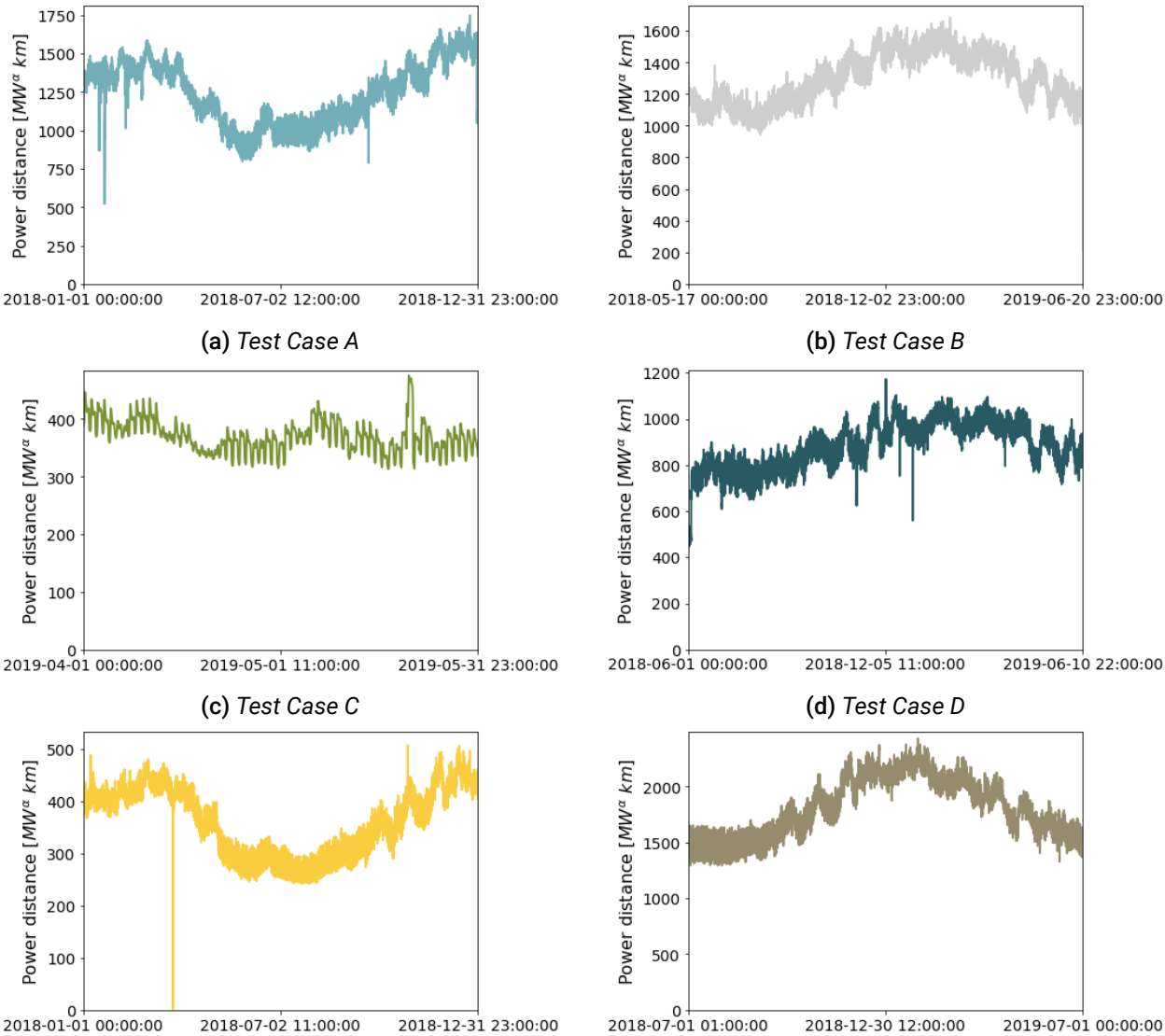
(d) *Test Case D*



(e) *Test Case E*

Figure 7.1.: *Hourly results for the power-flow based power distance per test case, for* $\alpha = 0.4$

Artificial grid based power distance results



(a) *Test Case A*

(b) *Test Case B*

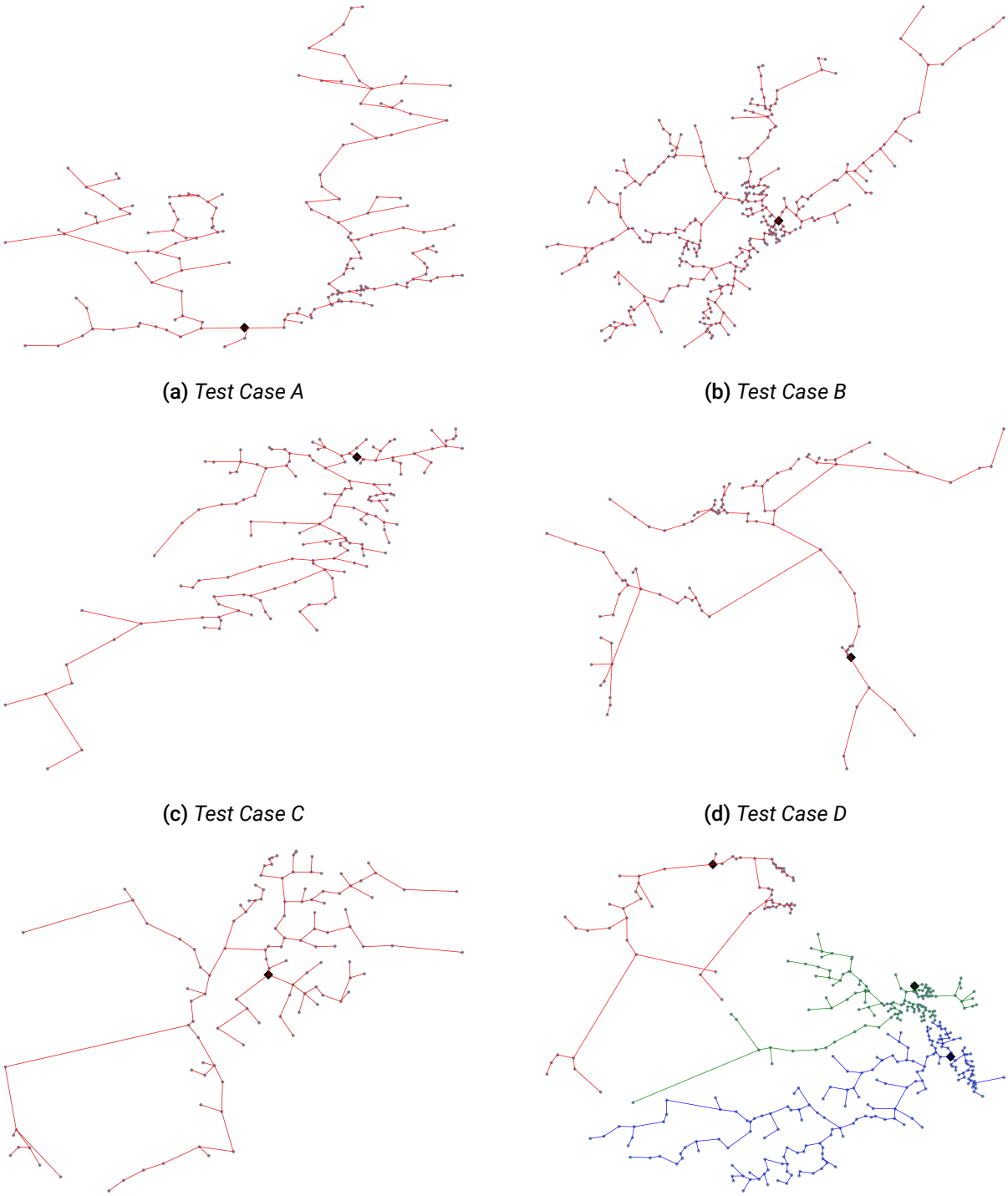
(c) *Test Case C*

(d) *Test Case D*

(e) *Test Case E*

(f) *Test Case F*

Figure 7.2.: *Hourly results for the artificial grid based power distance per test case, for* $\alpha = 0.4$



(a) *Test Case A*

(b) *Test Case B*

(c) *Test Case C*

(d) *Test Case D*

(e) *Test Case E*

(f) *Test Case F*

Figure 7.3.: *Artificial grid for each test case created for $\alpha = 0.4$, for test cases with more than one transformer station each individual tree is shown in a different colour*

Demand distribution based power distance results

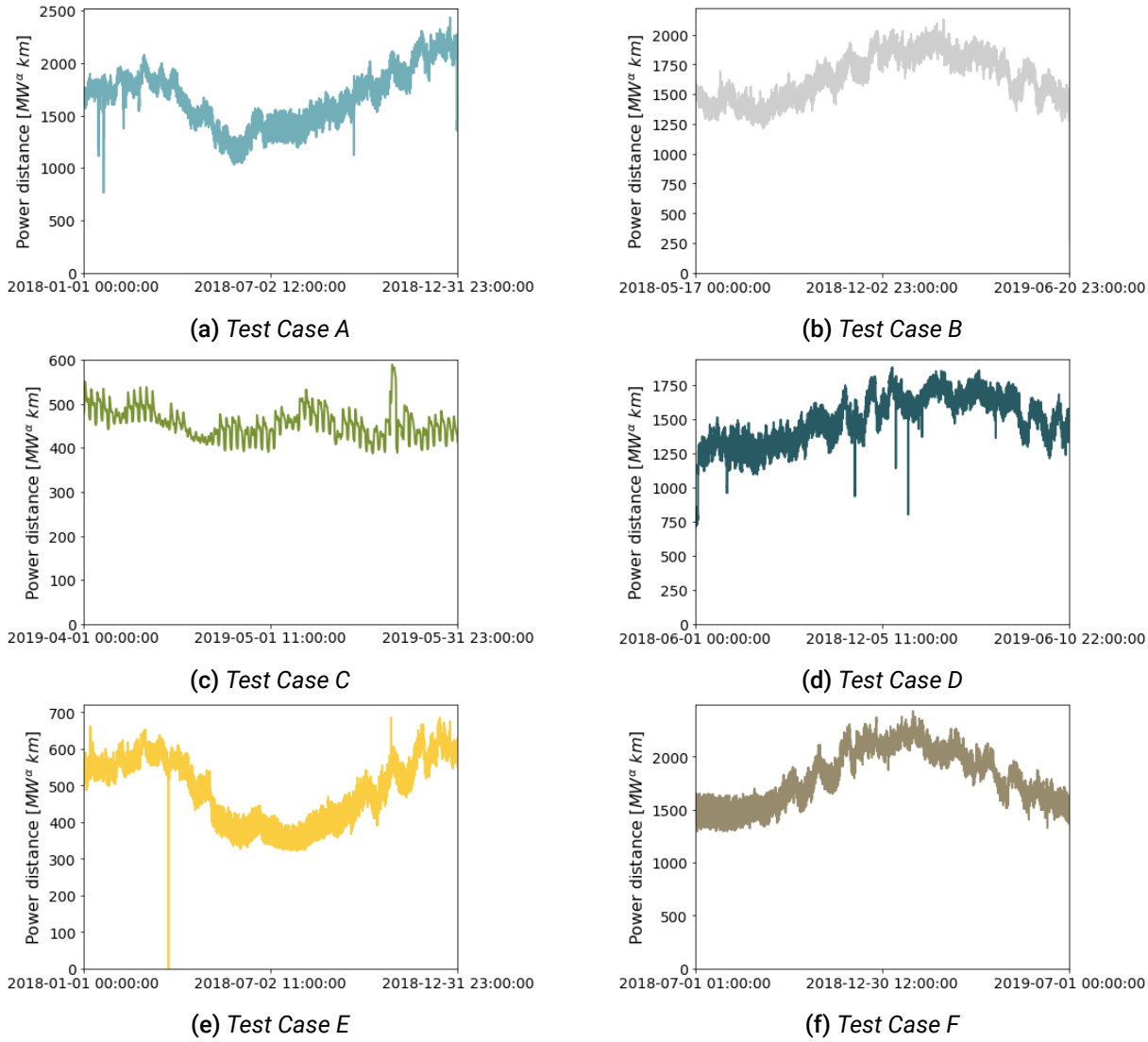


Figure 7.4.: *Hourly results for the demand distribution based power distance per test case, for* $\alpha = 0.4$

Energy distance

Infinity norm algorithm

Algorithm 1 ∞-norm based on max PF

```
procedure IDENTIFY_MAX_PF(D_ih, L_ij)
    P^max ← 0
    P_D^∞ ← 0
    P^E
    for h ← 1..H do
        Calculate OPF (D^h, L)
        P^E ← solution
        for e ← 1..E do
            if P_e^E ≥ P_e^max then
                P_e^max ← P_e^E
            end if
        end for
    end for
end procedure
P_D^∞ ← Σ_{e∈E} L_e * |P_e^max|^α
Return P_D^∞
```

A. Acronyms

AMS Advanced Metering System

CENS Cost of Energy Not Supplied

DEA Data Envelopment Analysis

DSO Distribution System Operator

NPAM Network Performance Assessment Model

RME Reguleringsmyndigheten for Energi

RNM Reference Network Model

SM Smart Meter

B. References

[1] C.M. Domingo Comillas Universidad Pontificia. *RNM: Reference Network Model*. URL: http://www.iit.comillas.edu/technology-offer/rnm (visited on 10/30/2019).

[2] A.Sanchez T. Gomez C.Mateo. "La retribucion de la distribucion de electricidad en Espana y el Modelo de Red de Referencia". In: *Estudios de economica aplicada* (2011).

[3] J. Wallnerström L. Bertling M. Larsson. "Review of the Swedish Network Performance Assessment Model". In: *Electric Power Systems Research* (2008).

[4] M. Larsson. *The Network Performance Assessment Model*. 2005.

[5] THEMA Consulting Group. *Computing the power distance parameter*. 2018.

[6] SINTEF Energi AS. *Planleggingsbok for kraftnett - Kostnadskatalog distribusjonsnett*. 2019.

Disclaimer

About THEMA:
THEMA Consulting Group is a consulting firm focused on electricity and energy issues, and specializing in market analysis, market design and business strategy.

THEMA Consulting Group

Øvre Vollgate 6
0158 Oslo, Norway

support@thema.no
https://www.thema.no/

Berlin office

Friedrichstrasse 68
10117 Berlin, Germany

NVE
Reguleringsmyndigheten
for energi – RME

Reguleringsmyndigheten for energi

MIDDELTHUNSGATE 29
POSTBOKS 5091 MAJORSTUEN
0301 OSLO
TELEFON: (+47) 22 95 95 95

www.reguleringsmyndigheten.no