# Monte Carlo Sampling Techniques for the Efficient Estimation of Risk Metrics of a Stochastic Distribution Grid Power Demand Model

by

**Julian Betge**

to obtain the degree of Master of Science
at the faculty of Electrical Engineering, Mathematics and Computer Science at Delft University of Technology,
to be defended publicly on Monday August 31, 2020 at 13.00.

Master of Science Sustainable Energy Technology
*(M. Sc.)*

| | | |
|---|---|---|
| **Student number:** | 4944194 | |
| **Project Duration:** | 08.12.2019 - 31.08.2020 | |
| **Thesis committee:** | Dr. Simon Tindemans | Supervisor |
| | Barbera Droste (Alliander) | Supervisor |
| | Jacco Heres (Alliander) | Co-Supervisor |
| | Dr. Jose Rueda Torres | Chair |
| | Dr. Hesan Ziar | Committee Member |

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


# Abstract

The Distribution System Operator (DSO) Alliander has the ambition to explore possibilities of improving its predictive demand modelling applications. This research aims to contribute to one of these, the Advanced Net DEcision Support (ANDES) model, which provides detailed predictions for long-term capacity planning. This thesis pursues two main objectives: Firstly, the creation of a probabilistic power demand model which reflects the volatile nature of real customer demand in an adequate manner. Secondly, the development and evaluation of methods to estimate certain model output quantities of interest in a computationally efficient manner.

To this end, variance reduction techniques have been investigated. Self-tuning Importance Sampling (IS) methods giving different weight to time steps and/or customer load profiles have been developed. In order to make the optimisation stage superfluous, additionally an approach to find a generalised asset distribution has been investigated.

The essential finding from evaluating all considered methods was that their performance in terms of efficiency and accuracy depends mainly on two variables – the order of magnitude of the estimated quantity and the number of customer connected to an asset in question. For small assets and an estimated overload probability of the order $10^{-5}$ or smaller, all profile IS methods and especially the generalised bin probability IS showed the strongest performance with average speed-ups of 5-30 times with respect to the reference method of sampling full annual traces. For assets with more than 80 customers and small overload probabilities, the profile IS methods were found to frequently produce estimates of a significantly too small order of magnitude. Conventional Monte Carlo (MC) sampling and time step IS, in turn, produced reliable estimates regardless of the number of customers. For assets of all sizes with an estimated overload probability of the order $10^{-4}$ and larger, conventional MC sampling showed the best performance with speed-ups above 5 times. Overall, conventional MC sampling performed robustly in all circumstances, while IS demonstrated its potential to significantly increase the estimation efficiency of rare event probabilities in certain cases.

To determine which magnitudes demand maxima and minima can potentially reach, Extreme Value Theory (EVT) has been applied. The computationally more efficient methods and extreme value inference were considered not compatible, sampling full annual traces appears to be required for a reliable estimation of maximum and minimum demand return levels.

Based on the findings of this thesis, a flexible algorithm could be investigated in future research which employs IS for rare event probabilities and conventional MC sampling otherwise. For an integrated evaluation of all risk metrics, the algorithm could initially sample 200 entire annual traces to be used for extreme value inference.

# Acknowledgments

First of all, I would like to thank my supervisors **Dr. Simon Tindemans, Barbera Droste and Jacco Heres** for supporting me on every step of the way with patience, great ideas and helpful insights. Whenever problems arose or questions came up, I was sure I could receive good advice. Such great support cannot be taken for granted which is why I am extremely grateful and indebted to all of them.

I would also like to thank the other committee members **Dr. Jose Rueda Torres and Dr. Hesan Ziar** for taking interest in my research and for their willingness to be part of the committee.

I would like to thank all colleagues at Alliander for making me feel welcome from the first day on, especially Lloyd, Casper and Stan. It has been a pleasure to be a part of the team.

Last but not least, I am very thankful towards my friends from all the different periods of my life so far. Without my friends from Delft and other places I might not be at this point right now and surely the two years of study would have been much harder and less joyful without them. In no particular order, thank you Damiano, Fanis, Patrycja, Monica, Chatu, Jop, Megan, Anurag, Gopan, Oskar, Srishti and many more for being part of my life. A special thanks also goes to my house mates Sarah and Jan, without whom I would surely not have survived the last weeks. Finally, I would like to express my thankfulness towards my parents and my sister for their patience, support and love.

# Contents

# List of Acronyms

**ADMD** After Diversity Maximum Demand.

**ANDES** Advanced Net DEcision Support.

**BM** Block Maxima.

**CBS** *Centraal Bureau voor de Statistiek*.

**CE** Cross-Entropy.

**CMC** Crude Monte Carlo.

**DSO** Distribution System Operator.

**ECN** *Energieonderzoek Centrum Nederland*.

**EV** Electric Vehicle.

**EVT** Extreme Value Theory.

**FACE** Fully Automated Cross-Entropy.

**GEV** Generalised Extreme Value.

**GPD** Generalised Pareto Distribution.

**HLI** Hierarchical Level I.

**HLII** Hierarchical Level II.

**HLIII** Hierarchical Level III.

**HP** Heat Pump.

**HV** High Voltage.

**i.i.d.** independent and identically distributed.

**IRENA** International Renewable Energy Agency.

**IS** Importance Sampling.

**KvK** Netherlands Chamber of Commerce, *Kamer van Koophandel*.

**LS-HLD** *Laagspannings-Hoofdleiding*.

**LV** Low Voltage.

**MC** Monte Carlo.

**MS-HLD** *Middenspannings-Hoofdleiding*.

**MSR** *Middenspanningsruimte*.

**MV** Medium Voltage.

**NEDU** *Nederlandse Energiedatauitwisseling*.

**OS** *Onderstation*.

**PDF** Probability Density Function.

**PMF** Probability Mass Function.

**POT** Peak Over Threshold.

**PV** Solar Photovoltaic.

**SJV** *Standaard Jaarverbruik*.

# List of Symbols

$\boldsymbol{RE}$ The relative error or coefficient of variation of an estimator or an estimate.

$\boldsymbol{SE}$ The standard error of an estimator or an estimate.

$\boldsymbol{\Omega}$ A sample space. In most cases, the sample space of the demand model referred to.

$\boldsymbol{X}$ A random state or realisation from a sample space.

$\boldsymbol{x}$ A specific state or realisation from a sample space.

$\hat{\boldsymbol{RL}}^{\boldsymbol{k}}$ The estimated return level for a period with a length of $k$ blocks.

$\mathbb{1}_{\boldsymbol{A}}$ The indicator function which returns a value of 1 whenever event $A$ occurs and 0 otherwise. The indicator function has the useful property $\mathbb{E}[\mathbb{1}_A] = Pr(A)$, therefore its expectation is equal to the probability of event $A$ occurring..

$\xrightarrow{\boldsymbol{d}}$ Convergence in distribution.

$\boldsymbol{d_{cap}}$ The power capacity rating of an asset.

$\boldsymbol{d_{crit}}$ A critical power value above the power capacity rating of an asset.

$\boldsymbol{r}$ A generic risk metric of the form $r = \mathbb{E}[M(\boldsymbol{X})]$.

$\boldsymbol{G_{\xi,\sigma,\mu}}$ The Generalised Extreme Value (GEV) with shape parameter $\xi$, scale parameter $\sigma$ and location parameter $\mu$.

$\hat{\boldsymbol{R}}$ An estimator of $r$, describing the procedure to obtain an estimate.

$\hat{\boldsymbol{r}}$ An estimate of $r$.

$\boldsymbol{\mathcal{N}(\mu, \sigma^2)}$ The normal distribution with expectation $\mu$ and standard deviation $\sigma$.

# 1 Introduction

With the increasing uptake of distributed energy resources in the course of the energy transition, the structure of the energy system is becoming increasingly decentralised. Departing from the traditional centralised structure entails that the roles and activities of the principle actors in the energy system change. Within this context, the International Renewable Energy Agency (IRENA) sees growing responsibilities for Distribution System Operators (DSOs): Distributed generation leads already presently to bi-directional and less predictable power flows on the lowest grid levels. This requires DSOs to step in and actively operate their grid in order to prevent congestion. Furthermore, an expanded DSO role could involve the direct operation of distributed flexibility assets or at least facilitation of a market where flexibility can be traded [49]. Next to these rather short-term operational and economic aspects, also the long-term grid planning approaches of DSOs are strongly affected by the characteristics of distributed energy resources. The IRENA suggests here that DSOs may use customer consumption, production and other types of data to create better demand forecasts [49].

The Dutch DSO Alliander is already putting the latter idea into practice: The Advanced Net DEcision Support (ANDES) model which Alliander develops is a novel long-term capacity planning tool which incorporates large volumes of various types of customer data. This thesis aims to contribute to the further development of the model by investigating the use of Monte Carlo sampling techniques tailored for the context of long-term capacity planning. Why this focus is relevant and which aspects of the ANDES model it concerns, is motivated more closely in the following section 1.1. On the basis of this initial assessment of the challenge, the principal objectives of this thesis are described in section 1.2. A set of research questions is subsequently derived from the thesis objectives in section 1.3. The notation and the overall structure of the present thesis report is explained in section 1.4 and section 1.5, respectively.

## 1.1 Motivation

An important objective of capacity planning from a DSO perspective is to ensure sufficient power distribution capacity is available in the short, medium and long-term, in order to minimise interruptions in the power supply to grid customers as far as possible. However, since capacity expansion is very expensive, it is also an important concern of capacity planning to identify economically viable grid investment paths and to prevent unnecessary overinvestment. This classic dilemma of capacity planning between the security of supply and the costs has increasingly become a trilemma in recent decades, due to the urgency of reducing greenhouse gas emissions of the energy sector – thus, ideally the energy supply should be secure, cheap and environmentally sustainable. Even though the latter dimension is not under the direct control of DSOs, the energy transition requires fundamental changes in the way capacity planning is approached.

To respond to the need for new capacity planning approaches, the ANDES model [81] is

being developed within Alliander. Traditional methods often use rather simple linear models combined with empirical observation to forecast peak loads in a top-down manner. The ANDES model, in contrast, relies on a bottom-up approach to obtain time-resolved asset load profiles by aggregating load profiles of individual customers. For different diffusion scenarios of key low-carbon technologies, the approach allows to predict which assets might be overloaded in the future. The model covers the entire Alliander grid, has a time horizon of up to 40 years with quarter-hourly resolution and currently comprises five future scenarios which entails a considerable computational effort and large volumes of output data.

Even though the ANDES model represents a major improvement with respect to traditional capacity planning methods, it is still under active development and not all of the underlying assumptions and modelling choices are ideal currently. Among the most problematic aspects is the usage of averaged category load profiles to approximate the electricity demand of several categories of unmonitored customer. Average profiles are typically smoother and show less stochastic features which are characteristic of real customer demand. On aggregate, this may lead to the underestimation of peaks and troughs. Furthermore, currently the predictions of ANDES are deterministic and their uncertainty is not quantified. In previous thesis work for Alliander by Valckx et al., a Monte Carlo simulation approach based on the random assignment of measured smart meter profiles to unmonitored customer has been investigated [92]. The approach is shown to improve the prediction of peak loads, however this comes at a significantly increased computational cost. Therefore, it would be desirable to develop a method which evaluates a probabilistic electricity demand model with improved prediction accuracy within a reasonable time.

## 1.2 Thesis Objectives

The general motivation to conduct this research, described in the previous section, can now be translated to specific thesis objectives. The two overarching objectives are to, firstly, set up a probabilistic electricity demand model and, secondly, develop a method to efficiently estimate certain model output quantities of interest. The demand model should be similar to the ANDES model in many respects, except for the usage of average category profiles, which are undesirably smooth as compared to real demand profiles. In other words, the demand model should ideally reflect the stochastic and volatile nature of actual customer demand as much as possible. The model output quantities of interest will be called risk metrics in the following, as they capture the risk for grid assets to exceed operational constraints which are derived from actual physical limits.

The risk metrics are chosen based on the input information requirements of capacity planning. Among these, overload forecasts are an indispensable type of analysis feeding into the capacity planning process which has to find a reasonable compromise in light of the trilemma characterised above. Specifically, it is desirable for DSOs to predict in detail which grid assets are likely to be overloaded in the future and what the likely magnitude of the overloads is. In accordance with standard criteria used by Alliander, this can be broken down into three questions, each of which is addressed by a risk metric:

- **Question addressed by risk metric 1:** *For a given grid asset, what is the probability per time step to exceed a critical power level above the rated capacity?* This critical power level

lies above the rated power capacity, it is for example 110% for substations which transform High Voltage (HV) to Medium Voltage (MV).

- **Question addressed by risk metric 2:** *For a given grid asset, what is the expected total duration of exceeding the rated power capacity per year?* For HV/MV substations, for example, the threshold which should not be exceeded are 87.5 hours.
- **Question addressed by risk metric 3:** *How high can the maximum demand (and how low can the minimum demand) on a given asset become?* There are no standardised asset level criteria for this question as for the previous two questions, because this question serves to find out how much additional capacity is necessary to prevent or limit overloads to a desired frequency.

There are, of course, other operational constraints relating to voltage stability, current ratings and power quality in terms of frequency and harmonics content. The focus on power ratings and load extremes in terms of power here is due to the fact that the ANDES model has the same focus. This focus is sensible, since power is the most relevant quantity for long-term capacity planning. To sum up, the principle objectives of this thesis are to, firstly, create a probabilistic demand model and to, secondly, investigate methods which allow to efficiently estimate the three detailed model output quantities or risk metrics of interest.

## 1.3 Research Questions

The following research questions are addressed in this thesis:

1. Which approaches exist to model distribution network power demand on the basis of limited customer monitoring data?
2. Which methods are able to improve the computational efficiency of estimating model output quantities of interest?
3. What assumptions underlie the current ANDES model? Which of them are reasonable in light of research question (1) and which requirements for the implementation of the demand model can be derived from the discussion?
4. How do the most promising methods from research question (2) perform in the given context?
5. How compatible are the Monte Carlo (MC) sampling-based methods developed for the first two model output quantities of interest and extreme value inference for the third quantity in the given context?

## 1.4 Notation

As this thesis is tailored for specific applications within Alliander, for greater ease of understanding the Dutch abbreviations of some specific terms are used throughout this report. This

primarily concerns the names of assets of different levels in the grid hierarchy. Furthermore, all non-abbreviated Dutch terms are set in italics. Table 1.1 gives an overview of the terms with English definitions, additionally the terms can be found in the list of acronyms.

**Table 1.1:** Overview of the Dutch terms and abbreviations used.

| Dutch term and abbreviation | English definition |
|---|---|
| *Onderstation* (OS) | High-voltage/medium-voltage substation |
| *Middenspanningsruimte* (MSR) | Medium-voltage/low-voltage substation |
| *Middenspannings-Hoofdleiding* (MS-HLD) | Medium-voltage cable |
| *Laagspannings-Hoofdleiding* (LS-HLD) | Low-voltage cable |
| *Standaard Jaarverbruik* (SJV) | Standard yearly consumption |

Alliander is an umbrella company with several divisions. Liander is the division of Alliander which carries out the actual network operation and management - and is, thus, responsible for the core DSO activities. The data-science division in turn, within which the ANDES model is being developed, is directly part of Alliander. Throughout this thesis, for the sake of better readability, the internal organisational structure is not further differentiated and reference is always made to the umbrella company Alliander.

In terms of mathematical notation, the convention is followed to denote random variables in capital letters, e.g. $X$, and specific realisation of these random variables in small letters, e.g. $x$. Estimators and estimated quantities are denoted with a hat '$\hat{}$'. Furthermore, vectors are set in bold, for example $\boldsymbol{v}$ for a specific vector and $\boldsymbol{V}$ for a random vector.

## 1.5 Thesis Structure

This thesis is structured such that reviewing and reflecting on the work of others (chapters 2-4) is separated as much as possible from contributions of this work (chapters 5-8).

The review of previous research and the presentation of methodological foundations in chapters 2-4 establish the general context in which the specific research of this thesis is embedded. The literature review in chapter 2 serves hereby as the basis for the following two chapters. Against the backdrop of the demand modelling approaches reviewed in section 2.2, the assumptions underlying the ANDES model are examined and the demand modelling approach adopted for this thesis is motivated in chapter 3. In a similar way, the literature review in section 2.3 sets the stage for chapter 4 by identifying suitable methods to obtain the risk metrics of interest introduced above. In chapter 4, the general methodological foundations of the selected methods are then presented.

The chapters 5-8 describe and discuss the specific contributions of this thesis. In chapter 5, the demand model used in this thesis is specified and the implementation process is detailed. Chapter 6 focuses on the methods developed to evaluate the demand model such that the risk metrics can be obtained in a computationally efficient way. In chapter 7, the results of tuning certain parameters of the methods and regarding their computational efficiency are presented. Furthermore, also the risk metric values obtained for various sets of test assets are shown and

analysed. Finally, in chapter 8, conclusions are drawn and possibilities for further research are discussed.

# 2 Literature Review

This chapter lays the foundation for the following two chapters by reviewing relevant previous research. In section 2.1, an overview of power system capacity planning and reliability evaluation is given, as this is the general context in which the Advanced Net DEcision Support (ANDES) model and this research fall. Subsequently, some of the literature of more specific relevance is reviewed in order to address the first two research questions: Concerning research question (1), the distribution network demand modelling approaches of other studies are considered in section 2.2. Concerning research question (2), suitable methods to obtain the risk metrics of interest in a computationally efficient way and power system studies in which these methods have found application are reviewed in section 2.3. As several areas of research with a large body of literature are studied in this chapter, the following review cannot be comprehensive. It is rather of exemplary nature and geared towards the most relevant aspects for this thesis.

## 2.1 General Overview of Power System Capacity Planning and Reliability Evaluation

Power system planning is characterised by the frequently conflicting objectives of reliability, affordability and sustainability of the power supply – giving rise to the fundamental trilemma touched upon already in section 1.1. The trilemma is not only faced by the power sector, but rather by the overall energy sector. Its importance shows, for example, in the fact that the World Energy Council publishes an index which ranks countries based on how well they are able to balance the conflicting dimensions [23].

An intuitive approach to the trilemma from a power system planning perspective is to formulate it as an optimisation problem whose solution indicates where the optimal trade-off between the conflicting objectives might lie. To that end, Aghaei et al. propose a multi-period multi-objective optimisation method of a generation expansion planning model which minimises cost and environmental impact while maximising reliability. As a result, the optimal size, installation time and technology type of new generating units are obtained [2]. Tekiner et al. formulate a similar generation expansion planning problem also as a multi-period multi-objective optimisation. In their study Monte Carlo simulation is used to generate component availability and demand scenarios over which the optimisation using mixed integer linear programming is carried out. The authors highlight that expansion decisions differ depending on the weighting of cost relative to greenhouse gas emission minimisation [89]. The multi-objective optimisation in [4] also relies on mixed integer linear programming and explicitly takes demand side management as an option in the planning process into account.

When it comes to power distribution grid planning in the traditional sense, the trade-off between reliability and affordability is the relevant one, since the environmental impact of generation is outside the scope of distribution grid planners. Finding a trade-off in this context

is the goal of various optimisation approaches concerned with optimal distribution system capacity and topology reviewed by Khator and Leung in [55]. Next to optimisation approaches, Khator and Leung also identify heuristic based and artificial intelligence approaches which have a reduced computational burden and attempt to find at least near-optimal solutions. In [56] such a heuristic is used, while the paper also goes beyond the traditional grid planning scope of the review in [55]. The latter is due to a focus on optimal investment in distributed generation technologies by distribution companies, adding the generation aspect to the picture (however, the environmental impact of the distributed generation is not part of the optimisation heuristic here).

Optimisation studies and other approaches which are concerned with balancing the conflicting objectives of overall power system planning rely on assessments of the reliability, affordability and sustainability dimensions. Out of these, the evaluation of power system reliability is especially challenging and forms a field of study on which a large body of literature has emerged. This literature will be given a closer look in the remainder of this section, as this thesis aims to contribute to distribution grid reliability assessment methods. It is useful to begin with some general considerations which provide a structure to the subsequent review. The classic work of Billinton and Allen [9] serves as the basis for these considerations and other parts of the generic material in this section.

Power system reliability can be evaluated employing deterministic or probabilistic methods and criteria. Billinton and Allen draw attention to the fact that while deterministic criteria may be easier to understand, they fail to reflect the stochastic nature of component and system behaviour as well as customer demand. Deterministic approaches can be used to analyse the impact of certain worst case scenarios, however their potentially very low probability of occurrence is not accounted for which could lead to an overly conservative system design. Another fundamental distinction to recognise, highlighted in [9], is that between system security and system adequacy which are the major subdivisions of system reliability. System security refers to the system's ability to cope with disturbances which endanger its functioning. System adequacy in turn means the system's ability to meet customer demand within its operational constraints. The term, therefore, encompasses generation, transmission and distribution capabilities. System security is concerned with dynamic and transient phenomena, while system adequacy is centred around static properties [9]. The main focus here lies on adequacy considerations.

For a quantitative assessment of adequacy, it is common to use various indices (the specific indices used in later chapters of this thesis are called risk metrics) which condense model outputs or empirical observations into single figures. Furthermore, following Billinton and Allen, it is useful to consider three hierarchical levels for which adequacy indices can be obtained [9]. These hierarchical levels are visualised in figure 2.1. Even though the strict unidirectional and centralised hierarchical organisation of power systems is a relic of the past due to distributed energy resources, the three hierarchical levels are nevertheless useful for a rough classification of power system models and scopes at which adequacy can be studied. Some exemplary studies concerned with each of the levels are reviewed in the following for a general overview of the reliability evaluation literature. The evaluation methods are of course mentioned hereby, however a review specifically focusing on methods and their usefulness for this thesis can be found in section 2.3. Two main groups of methods are used in reliability evaluation, which are introduced already at this point: Analytical methods to

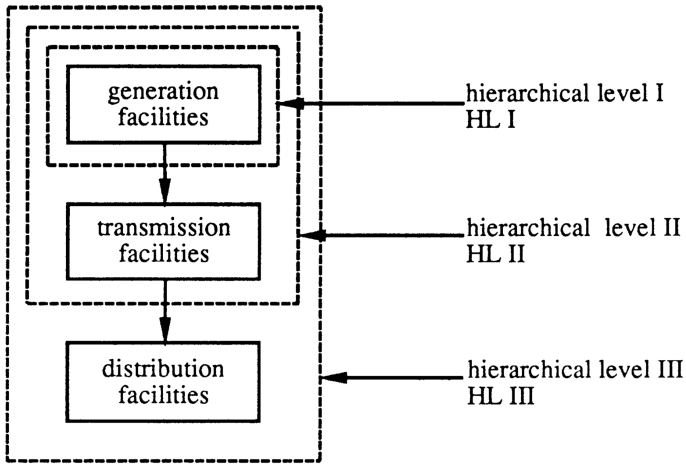


**Figure 2.1:** Hierarchical levels based on which power system models can be roughly classified. Source: Billinton and Allen [9].

directly calculate the desired indices and Monte Carlo methods which evaluate a large number of random experiments to obtain the same indices.

In **Hierarchical Level I (HLI)** studies, the total system generation is modelled and its adequacy to satisfy the total system load is assessed [9]. Therefore, the transmission and distribution networks are disregarded and usually single bus models are considered on this level. In an early study from 1970, adequacy indices are determined with an analytical approach considering outage probabilities and load uncertainties for the purpose of generation reserve planning [36]. More of the early work in probabilistic reliability evaluation is referenced in [3]. Dragoon and Dvortsov point out that the traditionally used planning reserve margin will typically result in different levels of reliability in different systems, whereas probabilistic methods are often computationally expansive. The authors instead propose the 'Z-Method' which is a simple analytical approach allowing to quickly calculate the effect of additional loads and generating units on system adequacy [31].

Assessing the adequacy of generating systems which include wind energy is a subject of more recent research. This is due to the energy transition and the system level impact that large volumes of wind energy can have. For example, an analytical approach using a frequency and duration analysis based on a Markov model of wind farms is developed in [29]. The approach allows modelling a wind farm like a multi-state conventional unit which facilitates the integration with common practices and existing models. In [10], an analytical method and a state-sampling Monte Carlo approach are compared for a multi-state wind energy conversion system model. It is shown that both approaches allow a reasonably good generation adequacy assessment.

**Hierarchical Level II (HLII)** approaches consider bulk transmission from generators to loads in composite systems. Thus, transmission network topology is taken into account here in a simplified manner. Two complementary types of indices can be calculated for HLII systems, bus or load-point indices and overall system indices [9]. This is, for example, done

in [11] where the effect of long-term load forecast uncertainty on both types of reliability indices is examined. In the study, both types of indices are found to increase with rising load forecast uncertainty. The authors furthermore highlight the usefulness of bus indices, as system wide indices can mask certain local phenomena [11]. The effects of including wind power in addition to load forecast uncertainty on composite system adequacy are investigated in [47]. To this end, the well-being analysis framework is employed which distinguishes the system states 'healthy', 'marginal' and 'at risk' and attempts to find a middle-ground between probabilistic and deterministic approaches.

For larger composite systems and when studying high outage levels with several failed components, enormous numbers of outage states have to be evaluated leading to high computation times. The adequacy equivalent method can circumvent this difficulty if the adequacy of specific parts of the network is of primary interest [58, 98]. Due to the complexity of transient phenomena, many reliability studies focus on assessing static system adequacy alone. Da Silva et al. in turn propose a framework for the integrated evaluation of system adequacy and security in composite systems. The framework distinguishes static and dynamic problems resulting from a disturbance which eventually allows to obtain composite reliability indices covering system adequacy and security [25].

The **Hierarchical Level III (HLIII)** comprises the entire power system with generation resources, as well as the transmission and distribution grids. Evidently, studying all major elements of a power system at once can result in very large and complex problems. Therefore, in practice HLIII studies often analyse the distribution grid only, but may use HLII load point indices as input [9]. HLIII load point indices reflect the adequacy of individual customers and thus the quality of service which end customers experience. In [7], an analytical approach to assess HLIII customer indices using HLII indices as input to important nodes of the distribution network is shown. The approach also allows to distinguish the relative contributions of the composite system (bulk generation and transmission) and the distribution grid to HLIII overall system indices. It is found that failures in the distribution grid contribute with 90-94% the dominant share to the system average frequency and duration of service interruption [7].

The impact of distributed energy resources on adequacy is among the topics investigated in more recent HLIII studies. A Monte Carlo approach to evaluate the adequacy of a distribution grid with customer-controlled distributed generation units is proposed in [44]. The study finds that distributed generation can improve the distribution system adequacy and thus be a viable alternative to substation expansion. In [19], a reliability model for distributed generators is developed using an analytical scenario reduction technique. Furthermore, Chen et al. address new issues in adequacy assessment relating to protection strategies, supply restoration and islanding operation which arise due to the integration of distributed generators [19]. Electric Vehicles (EVs) are another distributed energy resource with a potentially significant impact on distribution system adequacy. This impact is studied in [88] for plug-in hybrid EVs specifically. The authors present a Monte Carlo simulation approach with a load-dependant transformer ageing model to obtain a time-varying transformer failure rate. Furthermore, a business model is proposed which incentivises customers to charge their vehicles such that distribution system adequacy is improved [88].

At the end of the general overview of power system capacity planning and reliability evaluation in this section, it is possible to determine where the ANDES model is situated in this context: ANDES does not deal with the overall trade-offs faced in capacity planning, e.g.

through an optimisation procedure as in the studies discussed at the beginning of this section. Rather, it falls within the domain of reliability evaluation and its results are valuable inputs to the process of balancing the conflicting objectives of capacity planning. More specifically, the ANDES model is an HLIII adequacy assessment model. However, the distribution grid is modelled as a separate entity without considering the processes on the HLII to assess transmission and generation adequacy. This focus is of course sensible, since as a Distribution System Operator (DSO) Alliander is concerned with the adequacy of its distribution capacity. ANDES can thus be characterised as a distribution network demand model.[1] For this reason, the next section focuses specifically on literature relating to distribution network demand modelling.

## 2.2 Distribution Network Demand Forecasting and Modelling Approaches

Historically, peak load demand forecasts for distribution grid assets were often the primary ingredient of adequacy analyses and the subsequent capacity planning processes of DSOs. Thus, the general distribution network demand modelling problem was reduced to modelling peak demand only. According to McQueen et al., the maximum yearly load demand of a group of customers is often approximated using linear models and empirical observation [65]. The study remarks that, especially for residential customers, it is common practice to use the After Diversity Maximum Demand (ADMD) approach which takes into account the diversity behaviour of customer groups. The term diversity refers here to the phenomenon that the maximum demand of a group of customers is likely to be lower than the sum of the maximum demands of all individual customers because the latter are unlikely to all occur at the same time. A diversity factor is used to model that for smaller groups of customers the degree of coincidence is higher than for larger groups [65]. The ADMD approach, as described by McQueen et al., relies on measurements of the maximum yearly demand on assets and an empirically calibrated constant used in calculating the diversity factor.

Within Alliander, peak load demand forecasts also constitute the starting point of the traditional grid capacity planning process. As described by Sande et al., the forecasts are based on three types of inputs: The peak load of the previous year, expected additional demand of future spatial developments and a constant percentage of overall demand growth [81]. These peak load forecasts are, however, only made for *Onderstations* (OSs) which are at the top level in the asset hierarchy. For the underlying grid with lower voltage levels, comprising of *Middenspannings-Hoofdleidingen* (MS-HLDen), *Middenspanningsruimtes* (MSRs) and *Laagspannings-Hoofdleidingen* (LS-HLDen), only the present situation is assessed. The study details further that based on the peak load forecasts and present-day assessments, bottlenecks are identified and capacity expansion investments are decided on in a risk-based asset management scheme. The risk metric is obtained as the product of the occurrence probability and the impact of a bottleneck [81].

The ADMD method, the growth extrapolation approach of Alliander as well as similar

1 To be precise: Solar Photovoltaic (PV) generation of individual customers is modelled through negative values in the employed load profiles, distribution network net-demand model might be the more accurate term therefore. Nevertheless, the purpose remains distribution capacity adequacy assessment only.

traditional peak demand forecasting methods suffer from a common issue: They were conceived at a time of more constant developments and may not produce sufficiently accurate forecasts anymore, as fundamental changes in the energy landscape are underway [38]. The rapid uptake of sustainable energy technologies as part of the energy transition leads to new dynamics which the traditional methods, based on extrapolating present-day measurements with linear models, can fail to capture. Specifically, the decentralised nature of many sustainable energy technologies, impacting the lowest grid levels, poses difficulties for Alliander's traditional grid planning process which is centred on changes at the OS level [81]. Thus, there is a clear need for new forecasting approaches which allow a more reliable look into the near to longer-term future and go beyond forecasting peak demand alone.

When responding to this need by building an advanced distribution network demand model, it is important to consider that Low Voltage (LV) customers often exhibit higher volatility and random patterns which deviate more from the typical behaviour of aggregated demand on higher voltage levels [38]. Naturally, measured load profiles of individual customers already contain these desirable features and are, therefore, very well suited as model inputs or for model evaluation. Hoogsteen et al., for example, use the outcomes of a real world stress test on a network of 83 households, whose power consumption is measured every second, to benchmark simulation results [46]. However, as the authors in [38] point out, DSOs typically have access to half-hourly load measurements of larger, commercial customers, while for small customers measurements with high time resolution are still limited and expensive, despite the ongoing roll-out of smart meters. Alliander is no exception here, the share of unmeasured customers in the total energy consumed at the MSR level lies at approximately 80% [92]. These data limitations clearly show the necessity of suitable modelling techniques to simulate LV customer demand.

Several approaches which address this challenge can be found in the literature. As larger, commercial customers are often measured and tend to have less volatile load patterns, for the remainder of this section the focus lies mainly on modelling the more complex residential customer demand. Also does this narrower focus not limit the generality of the review substantially, as some of the modelling approaches for residential customers can be transferred to larger, commercial customers (as is done in the ANDES model, see the 'characteristic household demand profiles' approach below and section 3.1).

Whereas Swan and Ugursal provide a general review of residential energy consumption modelling [87], Giasemidis et al. review more specifically residential electricity demand modelling techniques [38]. Three major groups of modelling techniques, highlighted in bold below, can be identified whose core principles overlap considerably in both publications and are particularly relevant within the context of this thesis.[2] In the general review in [87], these three groups of techniques are classified as engineering methods which explicitly model energy consumption of end-uses on the basis of power ratings, appliance usage and physical laws. However, the scope of general residential energy consumption modelling is too wide for the given context. Therefore, in the following the modelling principles pertaining to the three groups will be considered in residential electricity demand modelling approaches employing limited smart meter measurements as in [38].

2 The other groups of the proposed classification in [87] are not as relevant in this context, since they either take estimates of the total energy consumption of the residential sector as input (top-down models) or attribute measured total building energy consumption to particular end-uses through regression techniques (bottom-up models with statistical methods).

At one extreme top-down models aim to find relations between energy consumption and characteristics of the housing sector as a whole, at the other extreme bottom-up models calculate the sectoral energy consumption by aggregating the energy consumption of individual appliances or buildings [87]. The smallest sensible unit from which a bottom-up model can be constructed is the individual appliance – by aggregating **appliance demand profiles**, household demand and eventually residential sector electricity demand can be obtained.[3] An example of such a model can be found in [78], where switch-on events of appliances result from a stochastic simulation based on active occupancy patterns and daily activity profiles which represent the likelihood of occupants to engage in certain activities using certain appliances at different times. The active occupancy modelling and the daily activity profiles rely on time use survey data. From switch-on events and power use characteristics of appliances, the overall building power consumption is obtained. A different approach to modelling the activity patterns of occupants can be found in [67], where a heterogeneous Markov chain with transition probabilities calibrated with time use survey data is employed. Furthermore, it is notable that in the study heating, ventilation and cooling appliances are modelled based on physical relations, making the model also a good example of an engineering method approach. A similar Markov chain based approach has been developed by Nijhuis et al. for research in cooperation with Alliander, respectively employing a Markov chain to model occupancy and appliance switching [71]. A focus of the work is the modelling of scenarios to enable future household load analysis. Overall, appliance demand profile models allow for a high degree of flexibility and detail, which however comes at the cost of model complexity and high input data requirements [38, 67].

The next higher level above the individual appliance, which can be used as the basic unit of a bottom-up model, is the household or single customer level. The general residential energy consumption review in [87] suggests it could be useful here to identify, model and scale up 'archetypes' of households to obtain an overall representation of the residential sector. In the more specific context of electricity demand modelling using limited monitoring data, this translates to the common approach of identifying **characteristic household demand profiles** and assigning them to unmeasured customers based on socio-economic or other household attributes. Various techniques which allow to uncover structure and cluster smart meter data lend itself to the first part of the task. Using principal component analysis, it was found in [1] that within individual household profiles the combination of weather and day specific baselines, and patterns of habitual behaviour can explain approximately 80% of electricity use. Also, Fourier transforms and Gaussian processes have proven to be suitable for the characterisation of residential demand profiles [64]. Using self organising maps [62] and self organising maps in combination with k-means and hierarchical clustering [77], groups of households with characteristic patterns in the respective demand profiles were identified. Thus, several studies show the principle feasibility of creating household groups with similar energy use behaviour and corresponding characteristic demand profiles.

The second part of the task, assigning the created characteristic profiles of household clusters to unmeasured customers, has also been investigated. Correlations between electricity consumption and socio-economic variables are found in [63]. Particularly building type, the number of bedrooms and household composition significantly influence electricity consumption within the scope of the study. However, the parameters to characterise electri-

3 In the classification of [87], this type of model falls most closely in the 'distributions' group.

city consumption (total electricity consumption, maximum demand, load factor and time of maximum demand) are six-month aggregates or averages and, as such, do not necessarily indicate how the studied socio-economic variables correlate with time-resolved load patterns. Furthermore, studies have shown that the variability in residential energy demand cannot be fully explained by infrastructural differences (see [66] and its references). Or, in other words, "families living in identically-designed homes use strikingly different amounts of energy" [66]. Another study finds that a typical socio-demographic classification used by energy companies matches only poorly with the energy behavioural clusters identified in the same study [41]. Overall, while the studies reviewed here may be too few to draw definite conclusions, they do suggest that assigning load profiles to unmeasured customers is a complex task and that it is worthwhile to verify the validity of the assignment. In addition to this challenging aspect, another important concern is that using averaged characteristic profiles for household groups reduces volatility and smoothes load extremes as well as other desired stochastic features of actual household demand profiles [38]. Since smoother individual profiles are likely to also reduce peaks and troughs on aggregate, this may lead to an underestimation of the extent to which a network is overloaded and, thus, limit the significance of results.

Remaining at the household level as the basic unit of a bottom-up model, **Monte Carlo simulation** approaches aim to address the two described drawbacks of approaches which rely on averaged, characteristic profiles. Termed 'sample' approaches in the more general review of [87], the core idea of the technique is to repeatedly and randomly sample model inputs from an actual or modelled distribution. This allows to reflect the uncertainty in model inputs, e.g. due to the variability in household attributes or unknown future developments, and to estimate the possible range of model outputs. For example in [68], low carbon technologies are randomly allocated to households in a LV network before running a power flow analysis to assess the possible range of impacts on energy losses, voltages and thermal loading.

In the more specific context relevant here, a Monte Carlo approach can consist in many repetitions of randomly assigning measured electricity consumption or generation profiles from a large pool to unmeasured customers. Through randomly sampling residential consumption profiles in combination with either PV profiles [69] or EV charging profiles [70], the range of potential impacts of these technologies on LV distribution networks has been assessed. Altogether, such Monte Carlo based approaches preserve the natural volatility of residential customer demand, since directly measured profiles are used. Also, when assigning consumption profiles randomly from an overall pool without any grouping, the difficulties to link energy behaviour to other household characteristics can be circumvented. However, the randomness of the assignments may also lead to an overestimation of volatility and network impacts, since the specifics of real households – which bound the stochasticity to some extent – are not considered [38]. Furthermore, the large number of repetitions necessary for meaningful results leads to an increased computational load.

Beyond the outlined three major approaches to the challenge of modelling residential electricity demand using limited smart meter data, naturally also approaches exist which investigate different techniques. Giasemidis et al. include, next to the household demand profiles, also the higher level of substation profiles in the set of building blocks of their model. In the framework of the study, 'buddy' profiles (measured smart meter profiles) are assigned to unmeasured customers in an optimisation procedure constrained by the mean daily demands of customers and measured aggregate demand on the substation. A genetic algorithm is used

for the optimisation procedure which imitates the principles of natural selection [38]. The authors conclude that their method performs better than comparable Monte Carlo approaches, however it requires substation monitoring data (preferably also from the MSR level and not only the OS level) which may not be available for entire distribution networks.

## 2.3 Review of Methods to Compute the Risk Metrics of Interest

The purpose of this section is to review potentially suitable methods for the efficient computation of the three risk metrics introduced in chapter 1. After the more narrow focus on distribution system demand modelling in the previous section, the focus in identifying appropriate methods to achieve the objectives of this thesis is widened here to not only consider studies concerned with distribution systems. The reason for this is that, despite the differences in the hierarchical levels described in section 2.1, the fundamental structure of adequacy assessment problems remains similar regardless of the level. In fact, the methods touched upon are even more generally applicable and have been used in a range of disciplines from engineering, natural science, economics and other fields. However, to remain close to the matter of interest in this thesis, the studies reviewed here are mostly concerned with power system applications.

### 2.3.1 Methods for Risk Metrics 1 and 2

The first two risk metrics assess the adequacy of currently installed distribution grid capacity. As such, they indicate which assets are at risk of being congested – in the present or the future, depending on whether the current baseload or future scenarios are considered. To reiterate, the questions addressed by the first two risk metrics are: (1) *For a given grid asset, what is the probability per time step to exceed a critical power level above the rated capacity?* (2) *For a given grid asset, what is the expected duration of exceeding the rated power capacity per year?* The risk metrics derived from the two questions are very similar, as discussed in detail in subsection 5.3.2. For the present-day baseload on assets modelled in this thesis, but also for the forecasted load of future years, exceeding the rated capacity or an even higher critical power level does not occur frequently. This suggests that methods to compute rare event probabilities could be well suited to compute the first two risk metrics. For this reason, the focus lies on these methods in this subsection.

The distinction between analytical methods and Monte Carlo methods in reliability evaluation has already been introduced in the general review of section 2.1. In **analytical approaches**, the system of interest is represented by a mathematical model whose equations are transformed or solved such that the desired reliability indices can be calculated [8]. Common analytical techniques used in reliability evaluation of engineering systems – therefore also beyond the specific context of power system reliability – are described by Billinton and Allan in [8] and comprise among others: the use of combinatorics; modelling the system as a network of components which can be connected in series, in parallel or in more complex ways; modelling the system with probability distributions; the use of Markov chains and processes; and frequency and duration techniques. Also the first and second order reliability methods

reviewed in [75], which are used in structural reliability to compute the failure probability of a structure, fall in this category. Essentially, the latter methods consist in Taylor series approximations of so called limit state functions, which divide safe from failure regions in the state space [76].

Billinton and Allan highlight the advantages and disadvantages of analytical approaches with respect to Monte Carlo techniques. An advantage of analytical approaches is the frequently much smaller computation time required to solve a model. One disadvantage is that analytical methods usually output expected values only (and no error quantification, or even entire probability densities). A second disadvantage consists in the simplifications that analytical models often rely on and which may limit the transferability of conclusions drawn from the model to the modelled real system [8]. The latter drawback considerably hampers the application of an analytical solution method to either of the three demand modelling approaches from section 2.2 based on aggregating simulated or measured customer demand profiles. It is not evident what form an analytical model could take that adequately captures the stochastic features of time-resolved simulated or measured demand profiles. If the demand distributions or other statistical properties of the demand profiles were used in an analytical model, the time information would be lost and it is likely that an over-simplification would result. Therefore, analytical approaches are ruled out at this point and the remainder of the subsection focuses on Monte Carlo approaches.

While analytical approaches to reliability evaluation strive for direct mathematical solutions of a model, in **Monte Carlo approaches** reliability indices are estimated by repeatedly simulating the random behaviour of a system. For each simulated random experiment, it is checked whether a failure event occurs. Then, essentially by counting the number of occurrences of an event, the various reliability indices can be estimated [8]. Formally speaking, this is commonly achieved by estimating the expectation of a simulation output variable [79]. More generally, Monte Carlo methods are a technique to compute integrals and rely on randomly sampling from a specific distribution to estimate quantities of interest [13].

The abovementioned disadvantage of Monte Carlo methods of a high computational effort becomes especially challenging when rare events are to be estimated using simple random sampling, also called Crude Monte Carlo (CMC). The reason for this is that a very large number of samples is needed to produce the rare events and an even larger number to obtain a desired level of accuracy for the estimates [13]. A common way to obtain rare event estimates more efficiently is through **variance reduction techniques**. As presented in greater detail in **??**, the accuracy of Monte Carlo estimates directly depends on the variance of the output quantities calculated for each random experiment. The general idea of variance reduction techniques is then, as the name says, to reduce this variance by exploiting knowledge of the model and its structure to reach the desired accuracy of estimates faster [79]. In the following some of the common variance reduction techniques are reviewed in rather general terms and illustrated with various power system studies in which they find application. Many of the general descriptions of the methods are based on Rubinstein and Kroese [79].

A positive or negative correlation between two random variables is used to obtain a variance reduction in the **common and antithetic variables** and the **control variables** methods.[4] A positive covariance between the two random variables is needed in the common

4 In some publications the same techniques are called the common and antithetic *variates* and control *variates* methods. Here the terminology of Rubinstein and Kroese [79] is followed.

variables method to achieve variance reduction, while the antithetic variables method relies on a negative covariance. For the control variables technique a control variable is employed which is correlated with the main variable of interest and for which the expectation is known [79]. Antithetic variables are used in [21] to achieve a variance reduction in a study of composite power system reliability with sequential Monte Carlo simulation. The core principle here is to use pairs of complementary random variables to obtain a negative correlation between two simulated experiments. In [12], composite system adequacy is studied with a sequential Monte Carlo simulation approach as well, while in addition to antithetic variables also control variables are investigated. In the study, HLI indices with a known expectation are used as control variables to achieve a variance reduction in obtaining the same indices for HLII.

**Conditional Monte Carlo** and **stratified sampling** are two closely related methods. The conditional Monte Carlo method is applicable if it is possible to analytically compute the expectation of the sample output quantity of interest conditional on another random variable. By sampling this other random variable and using it to compute the conditional expectation of the actual output quantity of interest, the variance in the latter is reduced. In stratified sampling, the sample space is partitioned into disjoint subregions called strata and a prescribed number of samples is taken from each stratum [79]. The variance reduction principle underlying stratified sampling is to prevent the 'clumping' of samples in the sample space. This is sensible because samples which are close to each other do not add much new information in the estimation process [52].

**Latin hypercube sampling** is a particular stratified sampling technique applicable to multidimensional distributions. The approach consists in dividing the sample space such that strata of equal marginal probability are obtained for each input variable. One sample is taken from each stratum which assures that all portions of the distributions of input variables are well represented [61]. Furthermore, by random shuffling it is assured that the input variables are uncorrelated [22]. Latin hypercube sampling is applied in [83] on the two random input variables renewable generation and system load to sample system states for which reliability indices are calculated. The method is found to be as accurate as CMC, while achieving the desired variance reduction resulting in less computation time. Composite system reliability is investigated with a sequential simulation approach in [84]. Latin hypercube sampling is hereby employed to obtain the time duration of each system state resulting in output time sequences which are more representative of the system's sample space compared to sequences obtained with CMC.

The **multilevel Monte Carlo** method is suitable for problems that can be approached with models of different hierarchical levels, which allow estimating model output quantities of interest with rising precision. By means of a telescopic sum, the quantity of interest can be decomposed into a rough estimate to which refinements with increasing precision are added [90]. Computational effort can be saved compared to CMC because the models producing coarser estimates require less computation time, while the number of samples necessary decreases with increasing level. Therefore, most of the samples are taken on the lower, computationally cheaper levels, whereas on the higher, computationally expansive levels only few samples are needed [48]. Huda and Živanović demonstrate the effectiveness of the approach in computing distribution grid reliability indices, where failure and repair processes are modelled through stochastic differential equations. The hierarchical levels consist in approximations of these stochastic differential equations on increasingly finer discretisation grids [48]. Tindemans

and Strbac apply the multilevel Monte Carlo methods to two exemplary adequacy assessment problems: Firstly, a composite HLII system with generation and transmission facilities, which is simplified to a single bus HLI system considering generation only to obtain a coarse lower level. Secondly, a system with storage units where the levels consist in storage dispatch models of increasing computational complexity [90].

The idea of **Importance Sampling (IS)** is to choose a sampling or biasing distribution which favours important regions in the state space [79]. Thus, the probabilities of drawing samples are changed such that more samples come from these regions of interest. In the context of rare event estimation, this results in producing the rare events more frequently as compared to random sampling. Through importance weights, also called likelihood ratio, each sample's contribution is weighted such that the overall estimate remains unbiased [13]. The increased frequency of rare events in combination with the down-weighting of their contribution reduces the variance among individual samples, resulting in the desired increase in estimation efficiency.

An early study in which IS finds application is [24], where the reliability of the Brazilian hydro-dominated composite system is assessed. System states in terms of outages, load levels and state of the hydrological reservoirs are sampled therein. The results of a subsequent power flow analysis and load shedding minimisation are used to calculate reliability indices. IS is applied by biasing the sampling towards those states which lead to load shedding with higher probability. Composite system reliability is also assessed in [72], where system failure rates in terms of customer disconnection are estimated using IS. The sampling scheme is distorted to favour extreme conditions where the loss of supply becomes more likely by introducing bias factors to increase line and generator failure rates.

For IS to be effective, choosing an appropriate importance sampling distribution for a given problem is crucial. Biondini highlights that choosing a bad sampling distribution can aggravate the problem and increase the variance of the IS estimator with respect to the CMC estimator [13]. Exploiting the knowledge regarding the structure of a specific problem can help in identifying a suitable importance sampling distribution. However, sometimes such knowledge is not available or it is desirable to use a more formalised and self-tuning approach. The **Cross-Entropy (CE) method** is such an approach and has gained increasing popularity in recent years. The general idea of the method is to minimise the Kullback–Leibler divergence between the (unknown) optimal importance sampling distribution and the actually employed sampling distribution, while using sampling to estimate a solution to the minimisation problem [79].

Da Silva et al. study generation capacity reliability evaluation in a single node HLI model using the CE method. The unavailability of generating stations with independent units is modelled using a binomial distribution. The CE method is then used to automatically tune the unit unavailabilities, leading to a distortion of state space probabilities that makes generation inadequacy more frequent [26]. CE-based IS is applied to study composite HLII system reliability in [91]. As in [26], the unavailabilities of generators are modelled using a binomial distribution and are automatically tuned. The same principle is applied to transmission lines. In addition, IS sampling is also performed for load states employing a truncated Gaussian density whose parameters are CE-optimised. An extension of the CE method is proposed in [15], which can take the correlation between random variables into account and uses a multinomial distribution to model multi-state conventional units and wind turbines. The approach allows

to counteract the efficiency loss of conventional CE-optimised IS that occurs when random variables are correlated. The effectiveness of the approach is demonstrated for a composite system with wind generation where wind speeds and loads are highly correlated [15].

For the sake of completeness, it should be mentioned that there are other variance reduction techniques and related methods with a more general purpose, which are however applicable to different kinds of problems than the problem in the focus here. Among these are **sequential Monte Carlo** as well as **Markov chain Monte Carlo** methods. Sequential Monte Carlo methods (to be distinguished from time-sequential Monte Carlo simulation), which also comprise the sequential variant of IS, are useful in Bayesian statistics for sequentially updating posterior distributions, as soon as new information arrives [30]. Markov chain Monte Carlo techniques allow approximate sampling from arbitrary distributions which do not have to be normalised. This is particularly useful for random variables with possibly nonlinear interdependencies [79]. As both classes of methods are not of direct relevance for this thesis, they are not reviewed more closely here.

After the overview of potentially suitable methods for the computation of metrics 1 and 2 in this subsection, it is now possible to compare and select those which appear most promising for further investigation. Analytical approaches were ruled out already above as not applicable to the type of model in consideration, Monte Carlo methods are thus the logical choice. Furthermore, due to the rare event character of the first two metrics, it seems sensible to make use of a variance reduction technique. Rubinstein and Kroese highlight the greater variance reduction potential of IS as compared to the other variance reduction techniques reviewed above (excluding sequential Monte Carlo and Markov chain Monte Carlo): IS can lead to "dramatic" variance reduction, sometimes in the order of millions, especially when estimating rare event probabilities. With the other techniques typically an up to 10-fold variance reduction can be achieved [79]. Multilevel Monte Carlo may be an exception here, as Tindemans and Strbac find a more than 2000-fold speedup in one case using the technique [90].

For the three distribution network demand modelling approaches based on customer demand profiles identified in section 2.2, it is not evident how a hierarchy of models with one or several computationally cheaper coarse models could be constructed. This suggests that multilevel Monte Carlo is not a natural choice here. IS, in turn, appears very promising for the problem at hand, due to its suitability for the estimation of rare event probabilities and its big variance reduction potential. Furthermore, considering the heterogeneity of distribution grid assets, the possibility of automating the search for a suitable importance sampling distribution with the CE method seems very relevant. For these reasons, IS and the CE method are selected for further investigation in this thesis.

### 2.3.2 Methods for Risk Metric 3

While the first two risk metrics indicate which assets are at risk of being affected by congestion problems in the present or the future, the purpose of the third risk metric is to determine how much asset capacity would be necessary to resolve potential congestion problems. To this end, the question addressed by the third risk metric is: *How high can the maximum demand (and how low can the minimum demand) on a given asset become?* The nature of this question is rather different than that of the questions relating to other risk metrics. It does not lend itself to be formulated as a typical Monte Carlo problem of estimating the expectation of a

simulation output. This is due to the focus on the very extremes of the demand distribution. A suitable method which has the relevant focus, devised to examine distribution extremes, is reviewed in this section.

Peak load demand forecasts are an important type of analysis feeding into the capacity planning processes of DSOs. As described at the beginning of section 2.2, often linear models calibrated with empirical observation were or are still used for this purpose, e.g. the ADMD method or the peak demand growth extrapolation approach traditionally used within Alliander. While these methods may work sufficiently well in practice under certain conditions (or at least they did in the more predictable pre-energy transition world of the past), the more mathematically sound and general approach of dealing with distribution maxima and minima is **Extreme Value Theory (EVT)**. Jacob et al. underline the analogous character of EVT and classical central limit theory: While the latter is concerned with the behaviour of 'normal' events in the asymptotic limit of large sample sizes, EVT is concerned with the behaviour of 'extreme' events in the asymptotic limit. Being an asymptotic theory, the objective of EVT is to extrapolate beyond the observations in a sample to make predictions about extreme events which may not have been observed before [51]. There are two main methods of inferring extreme characteristics of the overall population from a given sample – the Block Maxima (BM) and the Peak Over Threshold (POT) method. The BM approach uses the maxima or minima of large, equally long blocks of observations for inference. The POT technique performs inference based on all observations which exceed a high threshold. Depending on the method, the parameters of two different families of distributions are estimated, the Generalised Extreme Value (GEV) distribution for the BM method and the Generalised Pareto Distribution (GPD) for the POT method [51]. Being a general method, EVT has found application in many domains, including the analysis of power systems. In the following, some studies which fall in the latter domain are reviewed.

Belzer and Kellogg use EVT for an analysis of uncertainty in peak load forecasts. To this end, a daily peak load model is empirically calibrated and historical seasonal peak loads are simulated. These seasonal peak loads are then used to estimate the parameters of the GEV distribution, from which the peak load magnitudes corresponding to various probabilities of exceedance are obtained [6]. Wilson and Zachary demonstrate in [97] how the POT method can be used to estimate the demand-net-of wind distribution, which is subsequently employed in a HLI capacity adequacy assessment to obtain reliability indices. The motivation to model the demand-net-of wind distribution is that there may be a correlation between demand and wind which should be accounted for. EVT is advantageous in this context, as observations of demand and wind can be used directly without any further assumptions regarding their interdependence. The EVT-based approach is found to perform better than two alternative approaches, based on either the empirical demand-net-of-wind distribution or on the assumption of independence between wind and demand [97].

Usually, the assumption is made that the process generating the observations to be analysed with EVT is stationary [51]. However, it is also possible to consider the non-stationary case and allow the parameters of the extreme value distributions to change. In [20], power loss during blackouts is modelled with a non-stationary GEV approach. The authors argue that the approach is sensible because stationarity should not be assumed when considering the long-term evolution of power grids. To introduce the non-stationarity, the GEV parameters are allowed to vary linearly with time. For the case study China, it is found that the scale

parameter of the GEV distribution of the yearly maximum power loss changes significantly in the period 1981-2011 [20]. Substation annual maximum demand is forecasted in [59], similarly employing a non-stationary GEV. More specifically, the influence of customer count, average demand and installed PV capacity is accounted for through a Poisson point process model which is linked to the non-stationary GEV. The advantage of the approach is that energy consumption trends are integrated into the maximum demand forecasts, while usually both quantities are forecasted separately by utilities which can lead to inconsistent results [59].

To sum up, in contrast to metrics 1 and 2 where in principle various methods could be applied, for metric 3 the options are fewer and EVT, being a well developed and mathematically sound theory, is the method of choice. However, within the framework of EVT, the question remains whether the BM or the POT approach is more suitable in the context of this thesis. This question is addressed in chapter 4, after the relative merits of each approach have been considered in more detail.

## 2.4 Conclusion

In section 2.1 of this chapter, a general overview of the power system reliability literature was given. Exemplary studies were reviewed for each of the three hierarchical levels at which power system reliability can be evaluated. The general overview allowed classifying ANDES as an HLIII adequacy assessment model and, more specifically, as a distribution network demand model, since the distribution grid is modelled as a separate entity without considering HLII generation or transmission. Evidently, this is a sensible modelling scope for a DSO.

Due to this scope of ANDES, the distribution network demand forecasting and modelling literature was given a closer look in section 2.2. Hereby, three major groups of demand modelling approaches relying on limited smart meter measurement data as input were identified: appliance demand profiles, characteristic household demand profiles and Monte Carlo simulation approaches. The thread of distribution network demand modelling is picked up again in the following chapter 3, where the ANDES model is discussed in detail. On the basis of the reviewed groups of demand modelling approaches, an appropriate approach for this thesis is then identified.

Finally, section 2.3 focused on potentially suitable methods to obtain the risk metrics in the focus of this work. Exemplary studies from power system reliability and related fields were reviewed to illustrate how the discussed methods were applied previously in that context. IS, potentially in combination with the CE method, was found most promising for the efficient computation of metrics 1 and 2. For metric 3, EVT is the relevant approach from the toolbox of statistics. These methods will be presented and specified in more detail in chapter 4.

# 3 The ANDES Model and Demand Model Requirements

The purpose of this chapter is to describe and discuss the Advanced Net DEcision Support (ANDES) model in detail, in order to then specify requirements for the demand modelling approach to be adopted in this thesis. The description of ANDES in section 3.1 is broken down into a description of the input data used (subsection 3.1.1) and the model's workflow (subsection 3.1.2), based on which the assumptions underlying the model are uncovered (subsection 3.1.3). In section 3.2, these modelling assumptions are firstly discussed (subsection 3.2.1). Subsequently, requirements for the demand modelling approach are specified (subsection 3.2.2). The threads of this chapter are picked up in chapter 5, where the requirements specified here form the basis of developing and implementing the demand model.

## 3.1 The ANDES Model

The ANDES model is a detailed and complex application in continuous development, covering the entire Alliander grid with approximately 3 million customers and having a forecasting horizon of 40 years. In a nutshell, the working principle of the model is to assign load profiles to all customers and to sum them up in order to obtain yearly load forecasts for each grid asset. Possible futures are modelled in five technology diffusion scenarios and taking into account current spatial planning information. To assess the impact of these possible futures on the Alliander grid, technology and spatial planning load profiles are added to baseload profiles which model the present.

Providing a comprehensive description of all aspects of the ANDES model would not serve the purpose of this research. The aim of the following description is therefore rather to give a condensed overview of the model's input data and workflow which will subsequently allow to consider the key modelling assumptions.[5] The ANDES version which the description and the demand model developed for this thesis in chapter 5 refer to, is ANDES release 1.8 from January 2020 (before the year transition to 2019 data).

### 3.1.1 Input Data

Before looking at the model workflow, it is useful to consider the various types of input data employed. These are network topology data, spatial planning (*planologie*) data and strategic analysis scenarios as well as various sets of electrical load profiles. The load profiles contain quarter-hourly average power consumption values for one year, which corresponds to time series with 35,040 points. The main distinction to make here is between load profiles used to construct the baseload of the *present* and load profiles used to forecast the *future*. Note that baseload refers here to the entire load of the present to which technology load profiles are

5 Large parts of the description rely on internal Alliander documents which cannot be cited. Where available, publications are used.

added for modelling the future. Therefore, within the ANDES model documentation and this thesis the term is used in a slightly different sense than the usual sense of a constant load level which normally is not undercut. In the following, the employed inputs are described more closely:

- **Network topology data** specifying which customers are connected to which assets of different levels in the asset hierarchy and thus allowing to recreate the topology of the Alliander grid. A major part of the network has a radial topology, while a smaller part of the network is meshed. In the latter case, the load of customers connected to multiple assets of the same level is distributed evenly between them.
- **Present – Load profiles** to create the baseload of the present:
  1. **Telemetry data:** For many large consumers, telemetry data is available which records quarter-hourly averaged power measurements. Whenever a telemetry profile with direct measurements of a large customer's power consumption is available, the customer is assigned that profile.
  2. **Profiles based on KvK segment membership:** Dutch businesses can be classified according to the Standard Industrial Classification (*Standaard Bedrijfsindeling*) of the Netherlands Chamber of Commerce, *Kamer van Koophandel* (KvK) [18]. On a rather high level of aggregation 20 KvK segments can be identified representing major industrial divisions. For each KvK segment, the available telemetry data of the segment is averaged and normalised to obtain KvK segment profiles. If no telemetry data is available for a large consumer, it is checked whether the consumer can be assigned to a KvK segment. If that is possible, the normalised KvK profiles are scaled with the consumer's known *Standaard Jaarverbruik* (SJV), to approximate the consumer's unknown load profile.
  3. **Smart meter data:** For a fraction of small customers, smart meter data is available which, similarly to the telemetry data, records quarter-hourly averaged power measurements. Whenever smart meter data is available, it is directly used for the baseload forecast of the respective small customer.
  4. **Profiles based on smart meter data clustering:** For the majority of small customers, smart meter measurements are not available. To cope with the lack of data, the available smart meter data is clustered with an unsupervised learning algorithm and for each cluster normalised average profiles are calculated. A number of socio-economic characteristics of the customers in each cluster is then employed to train a supervised classification algorithm. Using this algorithm, the clusters small customers without a smart meter most likely belong to are predicted. These customers are then respectively assigned the normalised profile of their predicted cluster which is multiplied by the known SJV to approximate the unknown load profiles of these customers.
  5. **NEDU profiles:** The association *Nederlandse Energiedatauitwisseling* (NEDU) provides each year a set of average load profiles for different connection categories [93]. Due to the averaging, the NEDU profiles are less useful for the lowest grid asset level which is why the smart meter data clustering method has been developed. However, for large customers which cannot be assigned a KvK segment, the NEDU profiles of the respective connection categories are used. Also, for a negligible number of small

customers (less than 10 cases) NEDU profiles are used. Furthermore, traffic and street lights, which are small connections but not small private customers, fall within the connection category E4A and are assigned the corresponding NEDU profile. Again, the known SJV of the customers in question is used to scale the normalised NEDU profiles.

- **Future – Spatial planning data and load profiles** which can optionally be used for the future load forecasts. Enabling this options means that available data on spatial planning developments, entailing the connection of new customers or the strengthening of existing connections at some point in the future, is used. Different types of developments are distinguished and linked to corresponding spatial planning load profiles. These load profiles are scaled with the magnitude of the expected developments and are added to the baseload forecast of the assets which are projected to be affected. To account for the possibility that a spatial planning project is not realised, an average success rate of 70% is assumed and modelled through an additional scaling factor of 0.7. The main categories of spatial planning developments distinguished are urban developments and prospective large customer developments. Within these categories, several subcategories exist which can also differ in terms of the method used to scale the respective spatial planning profiles with the magnitude of the development:
  1. **Urban developments (*gebiedsontwikkelingen*):** The subcategories here are developments concerning real estate, offices, public facilities, greenhouses, industry and logistic. Within these a further differentiation may exist which determines the spatial planning load profile assigned to a particular urban development. For example, in the case of a real estate developments, the projected heating technology determines which load profile is assigned. Furthermore, the method to scale the load profiles to the magnitude of an urban development differs for the subcategories. The scaling is either done based on the number of houses to be build (for real estate), the number of square meters in new buildings (for offices and public facilities) or the number of hectares covered (for greenhouses, industry and logistic).
  2. **Prospective large customer developments (*klantontwikkelingen*):** While the urban developments always affect multiple connections, large customer developments only concern one connection. The following subcategories are formed to determine load profile assignment, depending on the business area a large customers engages in: agriculture, data centres, greenhouses, industry, offices, public facilities, charging stations, logistics, thermal power plants, hydroelectric power plants, wind parks, solar farms and other. For all subcategories the peaks of the assigned load profiles are scaled to match the capacity requested by the prospective customers.
- **Future – Strategic analysis scenarios and technology load profiles** are employed to predict the future impact of sustainable technologies on the Alliander grid which diffuse in the course of the energy transition and lead to an increasing electrification. Specifically, Alliander expects Solar Photovoltaic (PV), Electric Vehicles (EVs) and Heat Pumps (HPs) to have the largest impact on the electricity network load [81], therefore these technologies are modelled. The technology diffusion model results in predictions per customer of the installed amounts of the three technologies for each strategic analysis scenario. As with the spatial planning profiles, the technology load profiles are scaled and added to the baseload profile of a customer. While the strategic analysis scenarios and the technology

diffusion model are described more closely in the next subsection 3.1.2, below the origin and subcategories of the technology load profiles are outlined:

1. **Solar Photovoltaic:** The PV profile is based on 80 measured load curves of installations in the region around Utrecht retrieved from [73]. The proximity of the installations was chosen in order to preserve potential spatial correlations. The measured load curves are averaged to obtain the PV technology profile. The profile is scaled to match the predicted peak capacity of a future PV installation. The values of the profile are negative, since feed-in can be seen as a negative load. This allows summing the PV profile with other profiles, as usually.
2. **Electric Vehicles:** Four profiles are used to model the charging behaviour of EVs in four characteristic locations: at home, at work, as a guest (e.g. of a shop) and at fast charging stations. The profiles reflect time characteristics, such as a charging peak in the morning at work and in the evening at home during week days. Based on the findings of a *Centraal Bureau voor de Statistiek* (CBS) mobility study from 2015 [17], the yearly amount of kilometres per car is assumed to be 15,200 km. The energy required for this distance is redistributed among the charging locations according to a fixed ratio, based again on the results of [17] and on further assumptions. Therefore, the four charging profiles are scaled relatively to each other according to the mentioned fixed ratio and such that the overall energy consumed corresponds to the energy needed for the yearly amount of kilometres. For the future scenarios, the profiles are multiplied with the number of EVs predicted in the technology diffusion model for a given asset.
3. **Heat Pumps:** The HP profile is generated using a model developed by the *Energieonderzoek Centrum Nederland* (ECN) for the Flexnet project [34], which has been adapted such that it can generate HP profiles for different situations. Specifically, an average household has been assumed in terms of the input variables energy label, renovation year, living area, HP capacity and presence of a cooling functionality to generate the profile used in ANDES. For the future scenarios, the profile is scaled with the predicted number of HPs for an asset in question.

### 3.1.2 Model Workflow

In this subsection, the overall workflow of the ANDES model is given a closer look and it is described how the model inputs detailed in the previous subsection are used to produce the model outputs. Sande et al. distinguish in [81] seven main steps in the ANDES model workflow which are shown in figure 3.1. The following description is based on [81] and oriented around the seven steps. However, since the article was published in 2017, it is not up-to-date in all aspects – the few major changes to the model since 2017 are included based on internal Alliander documents.

In the first step, **strategic analysis scenarios** are devised regarding the future adoption of PV, EVs and HPs which are the technologies Alliander expects to have the largest impact on its grid. The strategic analysis takes into account the vision of the Dutch government for the energy transition, the regulatory framework and the economic side of the three technologies. The outcomes of the strategic analysis for various scenarios are used to determine the parameters of S-curves which model the diffusion dynamics of each technology on aggregate following the Fisher-Pry model. Thereby, technology diffusion is assumed to occur first slowly,

accelerate and then reach a turning point after which the diffusion speed decreases again. Originally, four scenarios of the most relevant combinations of low, middle and high diffusion pathways for each technology were created. Currently, five combination scenarios are used which feature a more detailed technology breakdown than the original four scenarios with subcategories within the three major technologies (e.g. various types of PV installations).

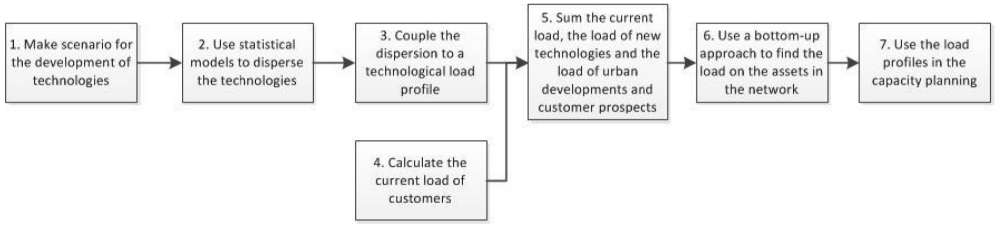


**Figure 3.1:** Schematic of the main steps of the ANDES model calculations. Source: Sande et al. [81].

Secondly, a **technology dispersion** step follows during which the yearly aggregate installed capacities from the first step are used to predict whether individual customers will install a technology under a given scenario. To accomplish this prediction, statistical models are trained with current socio-economic data which is linked to data on the currently installed capacities of the technologies. The statistical models allow determining probabilities for each customer to posses an EV, a PV system or a HP in a given year of the forecasting period. In a Monte Carlo simulation, the aggregate installed capacities of the technologies are then dispersed among individual customers according to the respective adoption probabilities. To arrive at single predictions for each year and future scenario, the capacity values per customer and technology resulting from 100 Monte Carlo simulation runs are averaged.

In the third step, **technology load profiles** are generated based on measured data for PV and EVs, and based on model outputs in the case of HPs. The details of how the average profiles for each technology are obtained can be found in the previous subsection 3.1.1. Subsequently, the technology load profiles are assigned to customers to which a technology was allocated in the technology dispersion step. PV profiles are scaled to match the predicted peak capacity of the installation in question, while EV and HP profiles are multiplied with the predicted numbers of EVs and HPs, respectively.

The fourth step consists in generating the **present-day baseload** for each customer. For a fraction of large and small customers, measured load profiles are available from telemetry or smart meter measurements. In these cases, the measured load curves are directly used as the baseload of the respective customer. Otherwise, three types of categorised average profiles are used to model the present-day load curves of the large proportion of unmeasured customers. In all cases, the average profiles are scaled to match the known SJV of the customers to be modelled. The methods of obtaining the average profiles, their categorisation as well as the methods to assign the average profiles to customers were given a closer look in the previous subsection 3.1.1.

In the fifth step, the **future load profile** of each customer is obtained by adding the load profiles of the technologies whose implementation is predicted to the customer's baseload profile. In step six, the future load profiles of all customers connected to an asset are summed up to obtain future asset load profiles. The Alliander network is thus recreated using a **bottom-up**

**approach**. This step relies on the network topology data specifying which customers are connected to which assets of different levels in the asset hierarchy. Additionally, the expected load from future urban and prospective large customer developments can be taken into account at the *Onderstation* (OS) level. This is achieved through spatial planning load profiles which are added to the obtained future asset load profiles. Details on the categorisation and further specifics of the spatial planning load profiles can be found in the previous subsection 3.1.1.

Finally, in step seven, the resulting load profile forecasts for all assets in the Alliander grid can be employed for **capacity planning** purposes. A key result of the current ANDES model are the forecasted load peaks and troughs and how they are composed (e.g. in terms of the shares of baseload, future technology and spatial planning profiles). For each asset, the load extremes can be compared to the asset's rated power capacity in order to identify where congestion problems are likely to arise and where grid extension may be needed.

### 3.1.3 Modelling Assumptions

After the relatively detailed look at the input data and the workflow of the ANDES model in the previous two subsections, it is now possible to take a step back and consider the key assumptions underlying the model. Naturally, more assumptions regarding the sources and pre-processing of input data, the clustering and creation of average profiles, the policy scenarios, the technology dispersion submodel etc. have been made (some of which are mentioned above or can be inferred from the model description). However, it is not the purpose of this section to unearth all assumptions in detail, but rather to focus on key assumptions and overall modelling choices. These are listed in the following:

1. **Simplified physics:** The physics of the Alliander network are modelled in a simplified way through the addition of power consumption profiles (which can have negative values for customers with a PV system) to approximate the load profile of grid assets on various levels. Grid losses, voltages and currents are therefore not explicitly modelled or taken into account.

2. **Monitored customers:** Whenever direct measurements of past load profiles for small or large customers are available, these are used to model the baseload demand of the respective customers. Thus, it is implicitly assumed that past behaviour of these customers is most suitable to forecast their future baseload behaviour.

3. **Unmonitored customers:** The baseload demand of unmonitored small and large customers is modelled through average profiles which are assigned based on classification memberships. Firstly, it is hereby assumed that separable categories or clusters of characteristic energy demand behaviour exist which can be modelled through average load profiles. Secondly, it is assumed that the average profiles can be reasonably well assigned based on socio-economic attributes (for small customers), industrial classification properties (for large customers) or electrical connection categories (if no other classification is available).

4. **Scaling of demand profiles:** To make the baseload model of unmonitored customers more realistic, the assigned average profiles are scaled such that the total yearly energy consumption of the profile matches the known SJV of these customers.

5. **Constant baseload**[6]**:** The baseload demand of customers remains constant during the entire 40-year forecasting period of ANDES, changes in the future demand are due the adoption of the three modelled technologies.

6. **Three modelled technologies:** Future energy transition scenarios are only modelled for the low-carbon technologies PV, EVs and HPs. Other promising technologies and approaches which could play an important role in the future, such as storage, demand-side management and energy efficiency improvements, are not taken into account.

7. **Deterministic high-level technology diffusion scenarios:** The future development of the total installed capacities of the three considered low-carbon technologies is modelled through deterministic scenarios which consider policy, regulatory and economic aspects.

8. **Deterministic technology dispersion:** The assignment of the technologies to individual customers is based on a probabilistic Monte Carlo step. However, to arrive at single predictions, the capacity values obtained from 100 Monte Carlo runs are averaged. This means that the future development of installed capacities at the customer level is also modelled in a deterministic way. Another implication here is that fractional capacity values are used, e.g. customers are assumed to install a fraction of a PV panel.

9. **No year-on-year variability:** Year-on-year variability due to weather fluctuations, economic cycles and other drivers is not taken into account, as only one set of future technology profiles is used.

10. **Constant network topology:** The network topology remains in the current state during the entire 40-year forecasting period of ANDES and is therefore assumed to be constant.

## 3.2 Requirements for the Demand Modelling Approach

The purpose of this section is to formulate requirements which the demand modelling approach should fulfil. To do so, the assumptions underlying the ANDES model detailed in the previous subsection 3.1.3 are now examined. This allows to identify which changes to the model could lead to more realistic assumptions and model results. On that basis, requirements are specified such that the demand model aims to replicate many aspects of ANDES, while proposing changes to others.

### 3.2.1 Discussion of the ANDES Modelling Assumptions

Due to the capacity planning and long-term focus as well as the model size of ANDES, assumption 1 (**simplified physics**) is quite reasonable. Carrying out detailed power flow calculations for all grid assets over 40 years and for various scenarios would be very computationally expensive and is not necessary for the assessment of network capacity bottlenecks. Also assumption 7 (**deterministic high-level technology diffusion scenarios**) is sensible in this

6 Note that baseload refers here to the entire load of the present to which technology load profiles are added for modelling the future. Therefore, within the ANDES model documentation and this thesis the term is used in a slightly different sense than the usual sense of a constant load level which normally is not undercut.

context. Long-term technology adoption speed and future policy cannot reliably be described by probability distributions [45], therefore a scenario analysis based on currently implemented and likely future policy pathways, the economic context as well as other societal factors is a suitable approach here.

Assumptions 6 (**three modelled technologies**) and 10 (**constant network topology**) in turn are unlikely to hold in reality, which is also highlighted by Sande et al. who are part of the ANDES development team [81]. It is probable that also other technologies apart from PV, EVs and HPs will play a role in the Dutch energy transition and similarly the grid topology is likely to experience some changes. However, it is also probable that the three chosen low-carbon technologies will have the highest grid impact and the grid topology is unlikely to change dramatically. Thus, the current assumptions are still justifiable, especially when taken into account in the interpretation of the results.

Assumptions 5 (**constant baseload**) and 9 (**no year-on-year variability**) are related and will almost surely not hold. Changing weather conditions, for example, have the potential to drive changes in baseload consumption and PV production which are not captured by constant baseload and technology profiles. Also, it is possible that there are long-term trends in household energy consumption [81], e.g. due to gradual energy efficiency improvements. Similarly, assumption 8 (**deterministic technology dispersion**) is questionable because much uncertainty is connected with the precise future locations and capacities of the considered technologies. Furthermore, using fractional, average capacity values is unrealistic and may lead to an underestimation of grid impacts, since an installed technology will have the full capacity of one unit. Many of these concerns could be addressed by making the ANDES model more probabilistic, since then the uncertainties relating to baseload consumption, year-on-year variability and technology dispersion could be explicitly quantified.

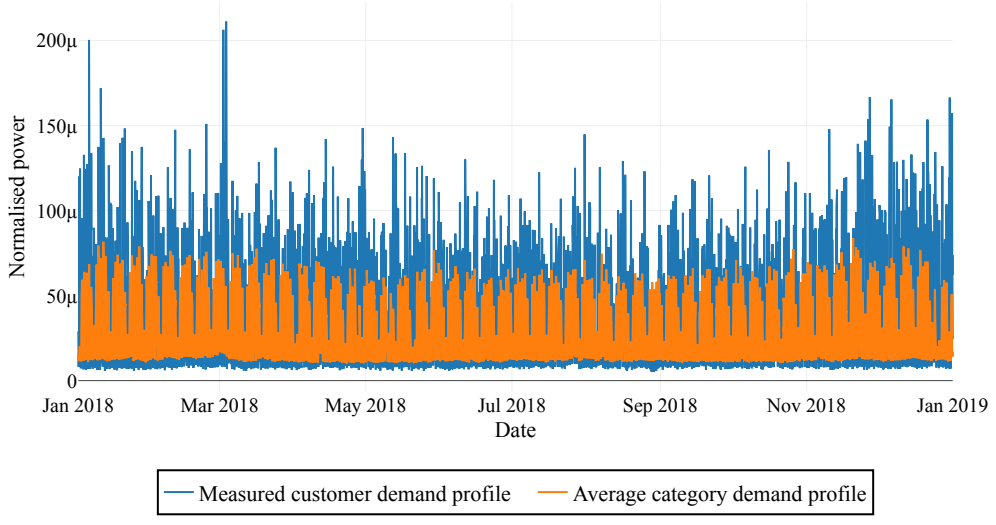


**Figure 3.2:** Comparison of a measured customer profile with the average profile of the category the customer is assigned to. Both profiles are normalised.

Assumption 3 (**the modelling of unmonitored customers**) appears to be the most problematic assumption. When looking at the description of the ANDES model in section 3.1 through the lens of the literature review in section 2.2, the ANDES model can be placed quite clearly in the second group of major modelling techniques employing 'characteristic household demand profiles'. This means that also the particular issues of the second group of techniques, identified in the literature review in section 2.2, are of concern: Firstly, the reservations regarding the validity of assigning energy-behavioural average category profiles to unmonitored customers based on socio-demographic customer attributes. Secondly, the likely underestimation of load peaks and troughs due to the usage of average profiles with smoother extreme features. The first issue is not in the focus of this thesis, as already much previous research and effort within Alliander have gone into the development of a method to assign average category profiles to unmeasured customers. The second issue remains a concern, especially on the Medium Voltage (MV) and Low Voltage (LV) asset levels where the variability and extreme features of individual customers weigh more in aggregates of typically much fewer customers, as compared to the OS level. Figure 3.2 illustrates the problem by contrasting a measured demand profile of an individual customer with the averaged category profile of the category the customer belongs to. Both profiles are normalised in the figure. It can be seen that in this case indeed the measured profile is much more spiky than the corresponding category profile.

In light of these issues regarding assumption 3, the remaining assumptions 2 (**the modelling of monitored customers**) and 4 (**scaling of demand profiles**) can be well justified. Using direct measurements for customers where they are available means circumventing the profile assignment and peak underestimation issues. Of course, there is no guarantee that past behaviour of a customer reflects its future behaviour. But using direct measurements, which contain the natural volatility and stochasticity of demand, is likely to be much more accurate than an averaged category profile. The scaling of averaged category profiles for unmeasured customers is also reasonable, since it assures that at least the present total yearly consumption, SJV, of unmonitored customers is modelled accurately.

### 3.2.2 Specification of the Demand Model Requirements

On the basis of the discussion of the modelling assumptions of ANDES in the previous subsection and considering how the scope of this thesis is defined, it is now possible to specify requirements for the demand model. The overarching requirement is to stay close to the current ANDES model where it is sensible and to only change the modelling of those aspects where underlying assumptions are most problematic. This requirement is directly related to the objective of this research to provide useful input for the further development of ANDES. In accordance with these considerations, the requirements are:

1. The core modelling principle of ANDES – summation of power consumption profiles of customers to obtain asset load profiles – should remain the same. Apart from fulfilling the overarching requirement in this central aspect, this also makes much sense because the way in which power system physics are simplified (assumption 1) is reasonable for long-term capacity planning purposes.

2. As mentioned above, the formation of energy-behavioural categories and the assignment

of customers to these categories has already been researched intensively at Alliander and picking up these threads is not in the scope of this project. Therefore, the demand model should remain within the already existing framework of customer categories used in ANDES.

3. Certain aspects of how the future is modelled were found to be justified (assumption 6: three modelled technologies, assumption 7: deterministic high-level technology diffusion scenarios, assumption 10: constant network topology), while other aspects appear to be more problematic (assumption 5: constant baseload, assumption 8: deterministic technology dispersion, assumption 9: no year-on-year variability). However, how the future development of technologies and network topology are modelled is not within the scope of this thesis and may be a topic of future research. Thus, the demand model is required to focus on how the baseload is modelled (which has implications for modelling the present and the future).

4. The principle requirement regarding the baseload demand modelling is to capture the stochastic and volatile nature of real customer demand as accurately as possible. Hereby, it is important to strike a reasonable balance between introducing too little and too much variability in the resulting asset demand, to avoid under- or overestimating network impacts.

## 3.3 Conclusion

In section 3.1, a close look was taken at the input data and the workflow of the ANDES model in order to uncover key modelling choices and assumptions. In the discussion of these in subsection 3.2.1, it became clear that while many assumptions can be well justified, especially modelling the demand of unmonitored customers through smoother average profiles diminishes the desirable stochastic variability of real customer demand. In subsection 3.2.2, requirements for the demand model were then formulated, taking into account the discussion of the ANDES modelling assumptions and the scope of this thesis. To improve the modelling of unmonitored customers, an important requirement for the demand model is to introduce more stochasticity and variability in the baseload demand. Based on the requirements formulated here, the demand model is developed and specified in chapter 5.

# 4 Methodology for the Statistical Analysis of the Demand Model

In this chapter, the sampling and inference methods for the statistical analysis of the demand model outputs selected in the literature review of section 2.3 are given a closer look. The general descriptions provided in the review will now be specified in mathematical terms for the methods identified as most promising to evaluate the risk metrics of interest: Importance Sampling (IS), potentially in combination with the Cross-Entropy (CE) method, and Extreme Value Theory (EVT). IS and the CE method fall in the broad class of Monte Carlo methods. Therefore, in section 4.1 firstly the conventional Monte Carlo approach is described, before turning to the mentioned more specific methods of particular interest here. In section 4.2, initially a comparison of the two most common approaches in EVT is given. The comparison serves to select the most suitable approach for the context of this work which is subsequently described in more detail. How each of the methods described in this chapter is applied to the specific problems in the focus of this thesis is the subject of chapter 6.

## 4.1 Monte Carlo Methods

The generic methodological background of the Monte Carlo methods identified as promising to address the risk metrics 1 and 2 is described in this subsection. The material presented in subsection 4.1.1 and subsection 4.1.2 is based on Rubinstein and Kroese [79], unless other references are cited. Subsection 4.1.3 relies on various references which are cited individually. Before beginning the main body of this section, the following pre-considerations illustrate how the notation used here relates to the demand model.

Several requirements for the demand model were formulated in subsection 3.2.2. From the requirement to introduce more stochasticity and variability in the modelling of baseload customer demand, it is clear that a *probabilistic* demand model needs to be set up. This is the first piece of information needed to proceed in this chapter. However, since the demand model has not been specified mathematically yet, for the time being the generic random vector $\boldsymbol{X}$ is used to denote all uncertain input variables of the model. The input variables determine the state the modelled system assumes, whereas the sample space $\Omega$ is the set of all possible states the system can assume. A *specific* set of inputs is written as $\boldsymbol{x}$, since it is a concrete realisation of the random input vector $\boldsymbol{X}$, drawn from the sample space $\Omega$.

The second important piece of information to proceed here is that in the definition of the risk metrics impact functions are used. These quantify the contribution or impact of each sampled system state to a risk metric. Since the risk metrics have also not been specified yet, a generic impact function $M(\boldsymbol{x})$ is used throughout this chapter. Formally, $M(\boldsymbol{x})$ is a function defined on the sample space $\Omega$ of the model. This function $M(\boldsymbol{x}) : \Omega \rightarrow \mathbb{R}$ assigns a numerical outcome to every specific system state $\boldsymbol{x} \in \Omega$. For a random system state $\boldsymbol{X}$, also the resulting impact $M(\boldsymbol{X})$, denoted for brevity as $M$, is a random variable. The probabilistic behaviour of the model then arises by assigning probabilities to all possible states [90].

Risk metrics 1 and 2, represented by the generic risk metric $r$, are both of the form $r = \mathbb{E}[M(\boldsymbol{X})]$. Theoretically, $r$ can be calculated by evaluating

$$r = \mathbb{E}[M(\boldsymbol{X})] = \sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \,, \tag{4.1}$$

where $f(\boldsymbol{x})$ is a Probability Mass Function (PMF), assigning a probability to each specific state $\boldsymbol{x}$.[7] However, this option quickly becomes computationally infeasible for systems of practical relevance with big sample spaces. As will become clear in subsection 5.1.4, this is also the case for the demand model considered here whose sample state contains a large number of states. The Monte Carlo methods presented in this section allow estimating $r$ with a reasonable computational effort.

### 4.1.1 The Conventional Monte Carlo Estimator and Confidence Intervals

As introduced in section 2.3 of the literature review, Monte Carlo approaches rely on repeatedly simulating the random behaviour of a system and evaluating the quantities of interest obtained in each random simulation experiment. In the conventional Monte Carlo approach simple random sampling is used. Emphasising the high computational cost which simple random sampling requires in many cases, the approach is also termed Crude Monte Carlo (CMC). For simplicity however, in the following Monte Carlo (MC) refers to the conventional approach based on simple random sampling. Supposing that in repeated runs of the simulation experiment $n$ independent and identically distributed (i.i.d.) system states $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are sampled, for which the impact function $M(\boldsymbol{x})$ is evaluated resulting in the i.i.d. random variables $M(\boldsymbol{X}_1), M(\boldsymbol{X}_2), \ldots, M(\boldsymbol{X}_n)$. The MC estimator to approximate the risk metric of interest $r$ is then

$$\hat{R}_{MC} = \frac{1}{n} \sum_{i=1}^{n} M(\boldsymbol{X}_i) \,. \tag{4.2}$$

This means that instead of evaluating the sum over the entire sample space in equation (4.1), only the mean of a sample drawn from the random variable $M(\boldsymbol{X})$ has to be calculated. The estimator is unbiased because its expectation is equal to the quantity of interest, $r = \mathbb{E}[\hat{R}_{MC}]$. Furthermore, the law of large numbers guarantees that $\hat{R}_{MC}$ converges to $r$ as the sample size goes to infinity. Following the notation in [90], estimators $\hat{R}$ are capitalised to underline that they are random variables. An estimate, in turn, obtained in a *particular* batch of simulation experiments, is a simple number and therefore denoted as $\hat{r}$. An estimate $\hat{r}$ can, therefore, also be pictured as a random draw from an estimator $\hat{R}$. The MC estimator describes the procedure to obtain an estimate, thus

$$\hat{r}_{MC} = \frac{1}{n} \sum_{i=1}^{n} M(\boldsymbol{x}_i) \,, \tag{4.3}$$

where the small letter $\boldsymbol{x}_i$ denotes a *particular* state sampled to evaluate $M(\boldsymbol{x})$ in a specific batch of $n$ simulation experiments.

7 Throughout this section the relations for the discrete case are shown, the relations for the continuous case can be obtained by replacing the sums with integrals whenever expectations are calculated analytically.

The MC estimator describes a procedure to obtain a point estimate. However, due to being an *estimate*, it is important to also quantify its accuracy. This is usually accomplished through confidence intervals which are reported next to point estimates. The reasoning behind the computation of confidence intervals starts with the central limit theorem, as a result of which for $n \to \infty$ and a finite variance $\sigma_M^2$ of the random variable $M = M(\boldsymbol{X})$, the MC estimator

$$\hat{R}_{MC} \xrightarrow{d} \mathcal{N}\left(r, \sigma_M^2/n\right) , \tag{4.4}$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\mathcal{N}(\mu, \sigma^2)$ the normal distribution with expectation $\mu$ and variance $\sigma^2$ [28]. The variance $\sigma_M^2$ is usually not known and can be estimated through the sample variance

$$S_M^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( M(\boldsymbol{X}_i) - \hat{R}_{MC} \right)^2 , \tag{4.5}$$

Thus, for large $n$, $\hat{R}_{MC}$ as well as the estimation error $\hat{R}_{MC} - r$ are approximately normally distributed, which means that approximate confidence intervals of the following form can be constructed

$$\left( \hat{R}_{MC} \pm z_{1-\alpha/2} \frac{S_M}{\sqrt{n}} \right) , \tag{4.6}$$

where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution $\mathcal{N}(0, 1)$ and $\alpha$ the confidence level.[8]

Two further useful measures for quantifying the accuracy of an estimate are the standard error *SE* and the coefficient of variation or relative error *RE*. The standard error of the estimator $\hat{R}_{MC}$ is the standard deviation of its sampling distribution. As a result of the central limit theorem, this sampling distribution is approximately normal for large $n$ and the standard error can be estimated using the sample standard deviation $S_M$ as follows:

$$SE_{\hat{R}_{MC}} = \sqrt{\mathbb{V}\mathrm{ar}[\hat{R}_{MC}]} \approx \frac{S_M}{\sqrt{n}} . \tag{4.7}$$

The relative error measures the variability of an estimator in relation to its magnitude. It is defined as the ratio of the standard error and the expectation of the estimator, which can be estimated as shown:

$$RE_{\hat{R}_{MC}} = \frac{\sqrt{\mathbb{V}\mathrm{ar}[\hat{R}_{MC}]}}{\mathbb{E}[\hat{R}_{MC}]} \approx \frac{S_M}{\sqrt{n} \cdot \hat{R}_{MC}} . \tag{4.8}$$

All three ways of quantifying the accuracy of $\hat{R}_{MC}$ show the typical convergence rate of conventional MC of the order $n^{-1/2}$. This convergence rate entails that each additional significant digit requires increasing $n$ by a factor of 100, thus computing 100 times more simulation experiments [5]. This shows why estimating rare event probabilities – which require large

**8** As the variance is estimated through the sample variance, it would be most correct to use Student's t-distribution instead of the normal distribution [28]. However, as the t-distribution converges to a normal distribution for $n > 30$ and $n$ is usually much larger in MC simulations, this difference is not of practical relevance and it is safe to use the quantiles of the normal distribution to construct confidence intervals.

sample sizes already for estimates with low accuracy – can become very computationally expansive with conventional MC.

### 4.1.2 Importance Sampling

In the literature review of section 2.3, IS was identified as an especially effective variance reduction technique for problems involving the estimation of rare event probabilities. The intuition behind variance reduction becomes clear when considering equations (4.6)–(4.8): If a technique is able to reduce $\sigma_M^2$ and consequently also $S_M^2$, the standard error and the relative error will be lower for the same sample size $n$ compared to the case without variance reduction. This allows computing estimates of higher accuracy for a given sample size, or equivalently estimates of comparable accuracy with a smaller sample size.

The idea of IS is to achieve variance reduction by choosing an alternative sampling distribution which favours important regions in the sample space. In the context of rare event estimation, a well chosen sampling distribution will result in producing the rare events of interest with a higher frequency. This allows faster estimation of the quantity of interest. However, the bias introduced in the sampling through the use of the alternative distribution must be corrected adequately. How this can be accomplished is presented in this subsection.

As previously, the expectation of $M(\boldsymbol{X})$ is the quantity of interest which can theoretically be calculated by evaluating

$$r = \mathbb{E}_f[M(\boldsymbol{X})] = \sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \,, \tag{4.9}$$

where the PMF $f$ is the original sampling distribution and the subscript $f$ in $\mathbb{E}_f[\cdot]$ highlights that the expectation is taken with respect to the original sampling distribution. Now, an alternative sampling distribution with PMF $g$ is introduced, which may only become zero when also $f$ is zero. By multiplying equation (4.9) with $g(\boldsymbol{x})/g(\boldsymbol{x}) = 1$, one obtains

$$r = \sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x}) = \mathbb{E}_g\left[M(\boldsymbol{X}) \frac{f(\boldsymbol{X})}{g(\boldsymbol{X})}\right], \tag{4.10}$$

whereas $\mathbb{E}_g[\cdot]$ highlights that now the expectation with respect to the alternative sampling distribution with PMF $g$ is computed. The alternative sampling distribution is also called importance or biasing distribution, due to biasing the sampling with respect to the uniform original distribution.

Estimating the expectation specified in equation (4.10) can be done in analogous manner to how the expectation of equation (4.9) is estimated with the conventional MC estimator of equation (4.2). Let $M(\boldsymbol{X}_1), M(\boldsymbol{X}_2), \ldots, M(\boldsymbol{X}_n)$ be a random sample taken from $g$ whose elements are i.i.d. random variables with distribution $g$. Then, the IS estimator takes the form

$$\hat{R}_{IS} = \frac{1}{n} \sum_{i=1}^{n} M(\boldsymbol{X}_i) \frac{f(\boldsymbol{X}_i)}{g(\boldsymbol{X}_i)} \,. \tag{4.11}$$

Sampling from $g$ instead of $f$ introduces a bias which ideally leads to a higher occurrence of

the events of interest. The ratio of the distributions,

$$W(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\,, \tag{4.12}$$

ensures then that the estimator is unbiased by giving appropriate weights to each $M(\boldsymbol{X}_i)$ in equation (4.11). This explains the terms importance weight or likelihood ratio for $W$. If the distribution $g = f$, the likelihood ratio $W = 1$ and the IS estimator becomes the conventional MC estimator.

The central limit theorem applies in analogous manner to the IS estimator as it does to the MC estimator. Therefore, using the estimated sample variance

$$S_M^{2\,'} = \frac{1}{n-1}\sum_{i=1}^{n}\left(M(\boldsymbol{X}_i)\cdot W(\boldsymbol{X}_i) - \hat{R}_{IS}\right)^2, \tag{4.13}$$

confidence intervals, standard errors and relative errors can again be estimated based on approximate normality for large sample sizes. If the distribution $g$ is well chosen, then $S_M^{2\,'} < S_M^2$, where $S_M^2$ is the conventional MC sample variance from equation (6.7). In analogy to equations (4.6)–(4.8), approximate confidence intervals for the confidence level $\alpha$ take the form

$$\left(\hat{R}_{IS} \pm z_{1-\alpha/2}\frac{S_M'}{\sqrt{n}}\right), \tag{4.14}$$

standard errors

$$SE_{\hat{R}_{IS}} = \sqrt{\mathbb{V}\mathrm{ar}[\hat{R}_{IS}]} \approx \frac{S_M'}{\sqrt{n}}\,, \tag{4.15}$$

and relative errors

$$RE_{\hat{R}_{IS}} = \frac{\sqrt{\mathbb{V}\mathrm{ar}[\hat{R}_{IS}]}}{\mathbb{E}[\hat{R}_{IS}]} \approx \frac{S_M'}{\sqrt{n}\cdot\hat{R}_{IS}}\,. \tag{4.16}$$

### 4.1.3 The Cross-Entropy Method

As mentioned in the literature review of section 2.3, it is important to choose the right IS distribution for an efficient estimation of the desired output quantities. An inappropriate choice for the IS distribution can even lead to a higher variance compared to conventional MC. The theoretically optimal IS distribution takes the form [13]:

$$g^*(\boldsymbol{x}) = \frac{M(\boldsymbol{x})\cdot f(\boldsymbol{x})}{\sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x})\cdot f(\boldsymbol{x})} = \frac{M(\boldsymbol{x})\cdot f(\boldsymbol{x})}{r}\,. \tag{4.17}$$

As the above equation shows, obtaining the optimal IS distribution requires knowing the quantity $r$, which is to be estimated, already in advance. Therefore, $g^*$ is not of direct practical use. However, indirectly it can be very useful for finding a good IS distribution, as the IS distribution can be required to be 'close' to the optimal distribution. In the CE method, the Kullback-Leibler divergence is used as a measure of how different two PMFs or, for the

continuous case, Probability Density Functions (PDFs) are.[9] The Kullback-Leibler divergence is also called the cross-entropy between two probability distributions, which justifies the name of the CE method. The fundamental idea of the method is to minimise the Kullback-Leibler divergence between the actually used IS distribution $g$ and the optimal distribution $g^*$ which is not directly accessible [13]. For the discrete case, the Kullback-Leibler divergence between these distributions is defined as [60]:

$$\mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v})) = \mathbb{E}_{g^*}\left[ln\left(\frac{g^*(\boldsymbol{X})}{g(\boldsymbol{X};\boldsymbol{v})}\right)\right] = \sum_{\boldsymbol{x}\in\Omega} ln\left(\frac{g^*(\boldsymbol{x})}{g(\boldsymbol{x};\boldsymbol{v})}\right)\cdot g^*(\boldsymbol{x})\,, \tag{4.18}$$

where $g^*$ is the optimal IS distribution and $g(\cdot;\boldsymbol{v})$ a parametric distribution with parameter vector $\boldsymbol{v}$. The logarithm in equation (4.18) can be rewritten as

$$\mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v})) = \sum_{\boldsymbol{x}\in\Omega} ln(g^*(\boldsymbol{x}))\cdot g^*(\boldsymbol{x}) - \sum_{\boldsymbol{x}\in\Omega} ln(g(\boldsymbol{x};\boldsymbol{v}))\cdot g^*(\boldsymbol{x})\,. \tag{4.19}$$

From equation (4.19) it becomes apparent that only the second term on the right-hand side is relevant for finding the parameter vector $\boldsymbol{v}$ which minimises the Kullback-Leibler divergence $\mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v}))$. This is due to the fact that the first term does not contain $g$ and is thus invariant with respect to the parameter vector $\boldsymbol{v}$. When substituting the expression for the optimal IS density from equation (4.17) for $g^*(\boldsymbol{x})$, the relevant optimisation problem is obtained [37]:

$$\begin{aligned} arg\ min_{\boldsymbol{v}}\ \ \mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v})) &= arg\ max_{\boldsymbol{v}} \sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x})\cdot ln(g(\boldsymbol{x};\boldsymbol{v}))\cdot f(\boldsymbol{x}) \\ &= arg\ max_{\boldsymbol{v}}\ \mathbb{E}_f[M(\boldsymbol{X})\cdot ln(g(\boldsymbol{X};\boldsymbol{v}))]\,. \end{aligned} \tag{4.20}$$

The IS density $g(\cdot;\boldsymbol{v})$ resulting from this minimisation is called the *near-optimal* IS density. The expectation $\mathbb{E}_f[\cdot]$ suggests that it is possible to estimate the right-hand side for a given $\boldsymbol{v}$ by randomly sampling from $f$. To make this estimation more efficient, it is possible to use IS at this point by introducing an alternative sampling distribution $g(\cdot;\boldsymbol{w})$ from the same parametric family as $g(\cdot;\boldsymbol{v})$, but with a different parameter vector $\boldsymbol{w}$. The optimisation program is changed accordingly and now takes the form [37]:

$$\begin{aligned} arg\ min_{\boldsymbol{v}}\ \ \mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v})) &= arg\ max_{\boldsymbol{v}} \sum_{\boldsymbol{x}\in\Omega} M(\boldsymbol{x})\cdot ln(g(\boldsymbol{x};\boldsymbol{v}))\cdot W(\boldsymbol{x};\boldsymbol{w})\cdot g(\boldsymbol{x};\boldsymbol{w}) \\ &= arg\ max_{\boldsymbol{v}}\ \mathbb{E}_{\boldsymbol{w}}[M(\boldsymbol{X})\cdot ln(g(\boldsymbol{X};\boldsymbol{v}))\cdot W(\boldsymbol{X};\boldsymbol{w})]\,, \end{aligned} \tag{4.21}$$

where $\mathbb{E}_{\boldsymbol{w}}[\cdot]$ denotes that the expectation is taken with respect to the alternative sampling distribution $g(\cdot;\boldsymbol{w})$ and the likelihood ratio

$$W(\boldsymbol{x};\boldsymbol{w}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x};\boldsymbol{w})} \tag{4.22}$$

9 The Kullback-Leibler divergence is not an actual distance measure, because it is not symmetric in its arguments [13]. Therefore, 'close' is not meant in the direct sense of a distance metric here.

is the ratio of the original sampling distribution and the biasing distribution. IS can now be applied to approximate $\mathbb{E}_{\boldsymbol{w}}[\cdot]$ by taking an i.i.d. sample $M(\boldsymbol{X}_1), M(\boldsymbol{X}_2), \ldots, M(\boldsymbol{X}_n)$ from the biasing distribution $g(\cdot;\boldsymbol{w})$ and evaluating [37]:

$$arg\ min_{\boldsymbol{v}}\ \ \mathcal{D}_{KL}(g^*,\ g(\cdot;\boldsymbol{v})) \approx arg\ max_{\boldsymbol{v}}\ \frac{1}{n}\sum_{i=1}^{n} M(\boldsymbol{X}_i)\cdot ln(g(\boldsymbol{X}_i;\boldsymbol{v}))\cdot W(\boldsymbol{X}_i;\boldsymbol{w})\ . \tag{4.23}$$

As noted by Rubinstein and Kroese, the right hand side of equation (4.23) is convex and differentiable for typical use cases of the CE method. Particularly, analytical solutions are available if the distribution to be CE optimised belongs to an exponential family. In these cases, analytical solutions to the optimisation problem can be obtained by solving the following system of equations for $\boldsymbol{v}$ [79]:

$$\frac{1}{n}\sum_{i=1}^{n} M(\boldsymbol{X}_i)\cdot W(\boldsymbol{X}_i;\boldsymbol{w})\cdot \nabla_{\boldsymbol{v}}\ ln(g(\boldsymbol{X}_i;\boldsymbol{v})) = 0\ , \tag{4.24}$$

where $\nabla_{\boldsymbol{v}}$ is the gradient with respect to $\boldsymbol{v}$. As equation (4.24) is sampling-based, its solution for $\boldsymbol{v}$ also results in a sampling-based updating formula. Thus, the CE method operates at the intersection of optimisation and estimation – if an analytical solution to the optimisation problem exists, it can be approximated with a sampling approach.

When estimating rare event probabilities, the optimal IS distribution is often very different from the original sampling distribution. This of course is related to the fact that the original distribution has usually low values in the sample space region of interest. Therefore, as Biondini underlines, determining the near-optimal biasing distribution is a rare event estimation problem itself which requires a large sample size if no knowledge of a suitable biasing distribution $g(\boldsymbol{x};\boldsymbol{w})$ is available in advance and conventional MC has to be used. This difficulty can be circumvented by using a sequence of intermediate target regions which cover a gradually shrinking portion of the sample space. These intermediate regions allow estimating biasing distributions which gradually approach the near-optimal biasing distribution [13]. The general approach of accomplishing this is described together with the developed CE algorithms in chapter 6, since it can then directly be illustrated with how it is applied in this work.

## 4.2 Extreme Value Theory

In the literature review of subsection 2.3.2, EVT was identified as the most suitable approach to obtain metric 3, as the metric is concerned with extreme load values. The two main methods of inferring extreme characteristics of the overall population from a given sample – the Block Maxima (BM) and the Peak Over Threshold (POT) method – have also been introduced already in subsection 2.3.2. Next to a general overview of EVT, the relative merits of both approaches are compared in subsection 4.2.1. The comparison allows to justify why the BM method is chosen for the given context of this thesis. In subsection 4.2.2, the BM method is then considered in more detail and mathematically specified.

### 4.2.1 General Overview and Comparison of the Block Maxima and the Peak Over Threshold Methods

The analogous character of EVT and classical central limit theory was already emphasised in the literature review of subsection 2.3.2: While the latter is concerned with the behaviour of 'normal' events and their cumulative effect in the asymptotic limit of large sample sizes, EVT is concerned with the behaviour of 'extreme' events in the asymptotic limit. In more mathematical terms, the central limit theorem states what form the limiting distribution of sums of random variables takes, allowing therefore the modelling of cumulative effects. EVT also makes statements about limiting distributions, however here the limiting distributions of extreme values of random variables are in the focus. Being an asymptotic theory, the objective of EVT is to extrapolate beyond the observations in a sample. This allows to make predictions about extreme events which may not have been observed before [51].

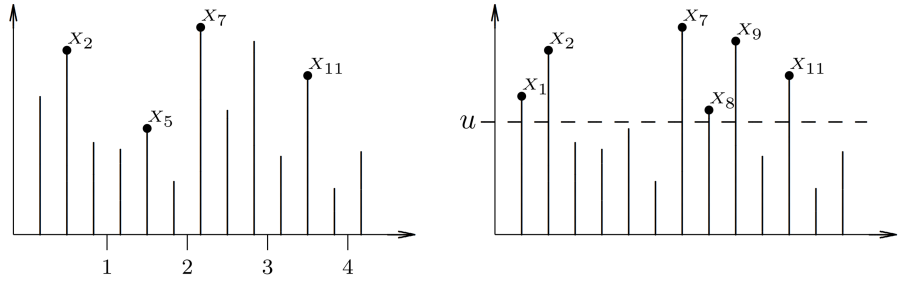


**Figure 4.1:** Visual comparison of how extreme values are identified in the BM approach (left panel) and in the POT approach (right panel). Source: Gilli and Këllezi [40].

Given a sample of empirical or simulated observations, the question arises which values can be considered extreme and used for inferring the appropriate limiting extreme value distribution. The two main definitions are visualised in figure 4.1: The definition used in the BM method is shown in the left panel of the figure, here the observations are divided in large blocks of equal lengths whose respective maximum (or minimum) values are regarded as extreme. In the example given in the figure, four periods with three observations respectively are distinguished (the low number of observations per period just serves illustrative purposes, in practice more than three observations per period should be used), thus observations $x_2$, $x_5$, $x_7$ and $x_{11}$ are the block maxima.

The definition used in the POT method is shown in the right panel of figure 4.1, where all observations exceeding the pre-defined high threshold $u$ are considered extreme. As the example shows, this leads to an overlapping, but different set of extremes as compared to the set resulting from the BM definition, comprising observations $x_1$, $x_2$, $x_7$, $x_8$, $x_9$ and $x_{11}$. Intuitively, considering these related but distinct definitions, it can be expected that each definition will also be associated with distinct limiting distributions. This is indeed the case and EVT derives the appropriate limiting distributions for both definitions: the Generalised Extreme Value (GEV) distribution for the BM approach and the Generalised Pareto Distribution (GPD) for the POT approach [40].

Both methods have their respective advantages and drawbacks which are considered in the following:

- The POT method considers all observations above the threshold, while the BM method may miss some relevant observations if two or more high observations occur within the same block. Therefore, the POT method appears to make better use of the available, potentially scarce, observations. Particularly, the POT method has been found to be more efficient in many cases if the number of exceedances is larger than the number of blocks on average [35].
- However, for the POT method it is necessary to determine a suitable high threshold. The trade-off here is that the theory requires a sufficiently high threshold, but increasing the threshold reduces the number of observations left for estimating the limiting distribution. Graphical diagnostics are available which can be used for manually finding a good balance, however automatic threshold selection was an unresolved issue in 2006 when [40] was published. In [86], a more recent study from 2017, a promising automatic threshold estimation procedure using the goodness of fit p-value has been proposed. Nevertheless, it appears that automated threshold selection for the POT method is still a matter of ongoing research on which no overall consensus has been reached yet.
- The BM method can be preferable if observations do not entirely fulfil the i.i.d. assumption which the theory underlying both methods requires. Typical examples where the i.i.d. assumption is violated are observations with seasonal periodicity or short range dependence. This is due to the fact that in such cases observations within a year or close in time cannot be considered independent anymore. However, if it is possible to choose blocks such that there is only dependence within blocks but not between blocks, independence of the block maxima can be assumed [35].
- In practice, block maxima may be the only available observations and in many applications the block periods appear naturally. The POT method, in turn, offers greater flexibility if changes in block size are necessary but difficult to realise [35].
- Overall, both methods seem to have a comparable performance for large sample size [35]. Caires finds that for more than 200 years of data, the estimation accuracies of both methods are comparable and rather good [16].

Against the backdrop of the merits and limitations of the BM and POT methods, it is now possible to select the more suitable method for the context of this thesis. For Alliander, theoretically it is desirable to analyse the distribution of demand extremes for each of its grid assets. This means that a large number of extreme value distributions has to be estimated and due to that a method which can be robustly automated is preferred. This is a strong argument against the POT method for which it is challenging to automatically find appropriate thresholds.

Furthermore, as described in chapter 3, the Advanced Net DEcision Support (ANDES) model produces yearly asset load curves as an output. Firstly, this means that annual blocks appear very naturally whose extreme values can be used directly. Secondly, asset load curves typically show some seasonality and short range dependence. The seasonality is due to the

yearly weather cycle which leads to seasonally varying consumption patterns and renewable energy production. Short range dependence arises because, despite the possibility of spikes and relatively sudden changes, aggregated load curves often have similar magnitudes in adjacent time steps. Thus, independence within yearly load curves cannot be assumed, whereas it is reasonable to assume independence for yearly peak load values. Finally, when using the Monte Carlo methods described earlier in this chapter, it is very likely that simulating more than 200 realisations of each scenario year is required for a sufficient accuracy of the estimates. Therefore, the data use efficiency advantage of the POT method for small sample sizes is not very relevant here. To sum up, for the given use case the above arguments clearly suggest that applying the BM method is more straightforward and intuitive. Therefore, only the BM method is described in greater detail in the following subsection.

### 4.2.2 The Block Maxima Method

In this subsection, the BM method is described based on Gilli and Këllezi [40] and Jacob et al. [51]. The notation used in this subsection relates to the demand model in the following way: In the given context of this thesis, using realisations of yearly asset load demand obtained from the demand model as blocks is the most intuitive choice. Therefore, the random variable $Y_i$ used below represents the magnitude of random load demand on an asset of a particular yearly realisation $i$. Furthermore, $z$ is the annual maximum or minimum demand magnitude of a particular yearly realisation and $Z$ the counterpart of a random realisation.

As mentioned above, one of the fundamental objectives of EVT is to derive limiting distributions of extreme values of random variables. The limiting distributions for block maxima are given by the Fisher–Tippett–Gnedenko theorem, also known as extreme value theorem, which is a central theorem in EVT:

▶ **Theorem 4.1 (Fisher–Tippett–Gnedenko).** *Let $Y_1, Y_2, ..., Y_n$ be a sequence of i.i.d. random variables and $M_n = max\{Y_1, Y_2, ..., Y_n\}$ the sequence of the maximum values of these random variables. If constants $a_n > 0$ and $b_n \in \mathbb{R}$ and a non-degenerate distribution function $G$ exist such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G\,, \tag{4.25}$$

*where $\xrightarrow{d}$ denotes convergence in distribution, then $G$ must belong to one of the three families of extreme value distributions:*

$$\text{Fréchet:} \qquad \Phi_\alpha(z) = \begin{cases} 0\,, & z \le 0 \\ e^{-z^{-\alpha}}, & z > 0 \end{cases} \quad \alpha > 0\,, \tag{4.26}$$

$$\text{Gumbel:} \qquad \Lambda_\alpha(z) = e^{-e^{-z}}, \quad z \in \mathbb{R} \tag{4.27}$$

$$\text{Weibull:} \qquad \Psi_\alpha(z) = \begin{cases} e^{-(-z)^{\alpha}}, & z \le 0 \\ 1\,, & z > 0 \end{cases} \quad \alpha > 0\,. \tag{4.28}$$

◀

The above theorem equally applies to minima, due to the relation $min\{Y_1, Y_2, ..., Y_n\} = -max\{-Y_1, -Y_2, ..., -Y_n\}$. The distribution functions for minima can be obtained through the

transformation $G_{min}(z) = 1 - G_{max}(-z)$ and by adjusting the above piece-wise definitions appropriately. For the case of block maxima, the probability density functions of the Fréchet, Gumbel and the Weibull distributions are visualised in figure 4.2.

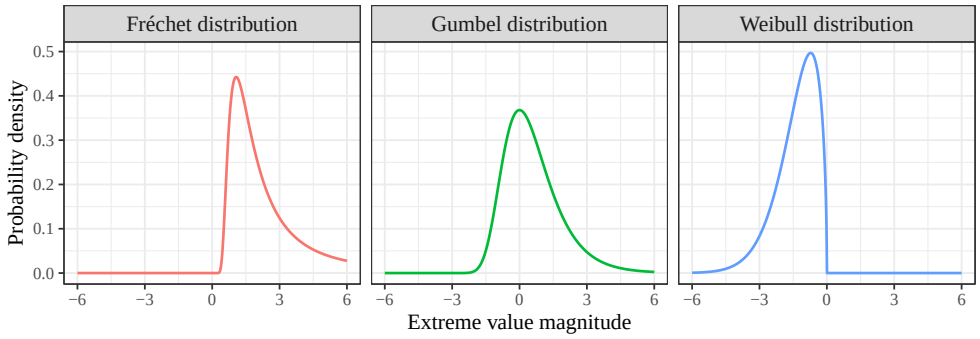


**Figure 4.2:** Probability densities of the Fréchet, Gumbel and Weibull distributions. For the Fréchet and the Weibull distribution the appearance for the parameter value $\alpha = 1.5$ is shown.

The Fréchet distribution has a heavy tail, which means that it decays more slowly than the exponential distribution. Therefore, it is suitable to model phenomena whose maxima have a tendency for extreme outliers. The Gumbel distribution has an exponentially decaying tail and can therefore be used to model phenomena whose maxima have thin tailed distributions. Lastly, the Weibull distribution is capable of modelling phenomena whose maxima have a finite upper bound.

It is possible to nest all three distribution families in one family, which is known as the von Mises-Jenkinson parameterisation or the GEV distribution. Its cumulative distribution function takes the form:

$$G_\xi(z) = \begin{cases} exp(-(1+\xi z)^{-1/\xi}) & \xi \neq 0, \quad 1+\xi z > 0 \\ exp(-e^{-z}) & \xi = 0 \end{cases} \tag{4.29}$$

where $\xi$ is called the shape parameter. As the name suggests, the shape parameter allows transitioning between the three distribution families. The parameter is obtained by setting $\xi = \alpha^{-1}$ for the Fréchet distribution, $\xi = 0$ for the Gumbel distribution and $\xi = -\alpha^{-1}$ for the Weibull distribution. Therefore, the Gumbel distribution can be seen as the limit case between the Fréchet and the Weibull distributions.

The GEV is of great practical use, since generally it is not known beforehand which type of limiting distribution a series of block maxima has. Using maximum likelihood or other estimators, the shape parameter can be estimated allowing to determine the most likely distribution type given a specific series of maxima. In practice, also the normalising constants $a_n$ and $b_n$ from equation (4.25) are unknown and therefore the three parameter representation

$$G_{\xi,\sigma,\mu}(z) = G_\xi\left(\frac{z-\mu}{\sigma}\right), \quad \text{where} \quad \begin{cases} z > \mu - \frac{\sigma}{\xi} & \xi > 0 \\ z \in \mathbb{R} & \xi = 0 \\ z < \mu - \frac{\sigma}{\xi} & \xi < 0 \end{cases} \tag{4.30}$$

is used. The scale parameter $\sigma$ and the location parameter $\mu$ replace hereby the unknown normalising constants $a_n$ and $b_n$ and can be estimated together with the shape parameter $\xi$.

Eventually, not the GEV distribution parameters themselves, but rather the distribution quantiles allow drawing conclusions for a given set of observations. The GEV distribution can be inverted which allows obtaining an explicit formula for its quantile function $Q(p)$. The quantile function returns the block maximum magnitude $z$ such that $Pr(Z \leq z) = p$. Therefore, it allows obtaining the block maximum magnitude which will not be exceeded with probability $p$. The quantile function of the GEV has the form

$$Q_{\xi,\sigma,\mu}(p) = G^{-1}_{\xi,\sigma,\mu}(p) = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - (-ln(p))^{-\xi}\right) & \xi \neq 0 \\ \mu - \sigma \cdot ln(-ln(p)) & \xi = 0\ . \end{cases} \tag{4.31}$$

Based on the quantile function and estimates $\hat{\xi}, \hat{\sigma}, \hat{\mu}$ of the GEV parameters $\xi, \sigma, \mu$, an estimate for the commonly used return levels

$$\hat{RL}^k = Q_{\hat{\xi},\hat{\sigma},\hat{\mu}}(p = 1 - \tfrac{1}{k}) \tag{4.32}$$

can be obtained, where $k$ denotes the return period in blocks. The return level $\hat{RL}^k$ is the magnitude which the block maxima will exceed once on average during any period of length $k$. In the context of this thesis with a block length of one year, a return level $\hat{RL}^{100} = 550$ kW would for example mean that a peak load value of 550 kW is only exceeded once every 100 years *on average*. The corresponding probability $p = 1 - \frac{1}{100} = 0.99$, meaning that in 99% of the years 550 kW are not exceeded – which is equivalent to the information the return level provides. However, return levels may be seen as a somewhat more tangible representation to answer the question how high a maximum value (or how low a minimum value) can become.

# 5 Specification of the Demand Model and the Risk Metrics

In this chapter, the demand modelling approach adopted within this thesis is presented in detail and the risk metrics, which are the quantities of interest to be obtained from the demand model, are specified. Firstly, the methodological core of the demand model is motivated and mathematically described in section 5.1. Secondly, section 5.2 deals with all implementation related aspects, especially those concerning the pre-processing of the input data used for the model. Lastly, in section 5.3 the choice of risk metrics is discussed and the risk metrics are specified in mathematical terms.

## 5.1 Description of the Demand Model

Before specifying the adopted demand modelling approach, the relevant previous work by Valckx et al. [92] is briefly described in subsection 5.1.1. This allows to characterise in subsection 5.1.2 in which aspects the demand modelling approach of this research is in continuity with the previous work and where it differs from it. Besides the specification of the demand model in subsection 5.1.3, also the demand model's sample space is given a look in subsection 5.1.4.

### 5.1.1 Previous Work

The usage of average category load profiles, which are often smoother and show less stochastic features than individual customer profiles, can lead to the underestimation of demand peaks and troughs. This issue has already been identified in the literature review of section 2.2 and discussed in the context of the Advanced Net DEcision Support (ANDES) model in subsection 3.2.1. Furthermore, it was also investigated in previous thesis work on the ANDES model by Valckx et al. [92]. The simulation approach of [92] and the results obtained with it are roughly outlined in this subsection, due the relevance of both for the demand model of this thesis.

For each category of small and large customers used in ANDES, a pool of 100 normalised profiles is formed (obtained from smart meter measurements for small customers and from telemetry measurements for large customers). The category membership of an unmonitored customer to be modelled determines then which pool a profile for that customer is drawn from. The normalised profiles are scaled to match the known yearly consumption, *Standaard Jaarverbruik* (SJV), of each unmonitored customer – analogous to the scaling done in ANDES. The random assignment and scaling process is repeated 100 times for all customers connected to an asset in question, such that 100 yearly asset load curves are obtained which allow a quantification of baseload demand uncertainty. This makes the method a Monte Carlo simulation approach which can be placed in the third major group of techniques in the grouping of the literature review in section 2.2.

The Monte Carlo (MC) simulation approach was benchmarked in two ways against the

ANDES approach which relies on assigning scaled average category profiles to unmonitored customers [92]: Firstly, by using measured *Middenspanningsruimte* (MSR) load curves and, secondly, by simulating load curves of 'virtual' MSRs. The idea of the latter benchmarking is to aggregate smart meter profiles of measured customers to create reference load curves of 'virtual' MSRs with similar customer structures as real MSRs. Supposing then that only the customer categories and the SJVs of the virtual MSR customers are known (as would be the case with unmonitored customers), the ANDES method and the MC simulation method were applied. Then it was evaluated how much they respectively deviated from the reference load curve of the virtual MSRs.

The principle finding from both comparisons was that while ANDES significantly underestimates peak loads, the MC simulation approach overestimates them to some extent. Therefore, the MC simulation approach constitutes an improvement with respect to the ANDES approach, however it suffers from the slight overestimation issue. It is conjectured in [92], that the overestimation in the MC simulation may be caused by profiles with an overall low yearly consumption which have some extreme peaks. If these are assigned to unmonitored customers with a regular to high yearly consumption, the scaling leads to an extreme amplification of these peaks which drive up the overall asset peak demand.

### 5.1.2 Motivation for the Adopted Demand Modelling Approach

On the basis of the requirements for the demand model formulated in subsection 3.2.2 and the results from the previous thesis work of Valckx et al. [92], the demand modelling approach adopted in this thesis can now be motivated. Firstly, for better readability, the demand model requirements are recapitulated in brief form:

1. The core modelling principle of ANDES – summation of power consumption profiles of customers to obtain asset load profiles – should also be used in the demand model.

2. Due to how the scope of this thesis is defined, the demand model should remain within the already existing framework of customer categories used in ANDES.

3. Again, due to how the scope of this thesis is defined, the demand model is required to focus on how the baseload is modelled (which has implications for modelling the present and the future).

4. The principle requirement regarding the baseload demand modelling is to capture the stochastic and volatile nature of real customer demand as accurately as possible. Hereby, it is important to strike a reasonable balance between introducing too little and too much variability in the resulting asset demand, to avoid under- or overestimating network impacts.

MC simulation approaches based on repeatedly and randomly assigning measured customer demand profiles from a profile pool to unmeasured customers, preserve the natural volatility and stochastic features of real customer demand (see the literature review of section 2.2). Therefore, they are well suited to address the fourth requirement. Taken together with the other requirements, this means that an approach similar to the one in [92] is appropriate. Evidently, this also makes sense due to the results of Valckx et al. that peak load demands

are estimated with greater accuracy compared to the average category profile approach of ANDES.

However, as described in the previous subsection, the approach in [92] suffers from overestimating peak loads to some extent. This may be seen as related to the last part of the fourth requirement, to strike a balance between over- and underestimating demand variability. In light of these considerations, the adopted demand modelling approach of this thesis extends the previous approach and differs from it in certain aspects:

- While in the previous work no differentiation regarding the yearly consumption of customers within categories is made, a more fine-grained modelling of this aspect is introduced here. This is accomplished through an additional binning of small customers within each category based on the SJV leading to sub-categories. As a result, **unmonitored small customers** are only modelled through the profiles of monitored small customers with a similar SJV. Two arguments speak in favour of the binning scheme: Firstly, energy consumption behaviour within a category is likely to be more comparable for customers with a similar yearly consumption. Secondly, profiles with a low SJV are scaled up to a lesser extent to model customers with a higher SJV. As explained in the previous subsection, over-scaling of spiky, low SJV profiles was conjectured in [92] to lead to the overestimation tendency of the previous MC simulation approach. Thus, the binning may allow to limit demand volatility to a more reasonable level.

- For large customers, the classification in ANDES is based on the economic activities carried out by these (according to the Netherlands Chamber of Commerce, *Kamer van Koophandel* (KvK) segmentation) and not on energy-behavioural clusters as with the small customers. Thus, it is likely that the demand characteristics of unmonitored large customers within the same KvK segment vary considerably, which punctual inspections seem to confirm. Therefore, using averaged segment profiles to model **unmonitored large customers** seems better than the alternative of randomly picking a telemetry profile of a customer from the same KvK segment, as done in [92]. This means that for large unmeasured customers, a small underestimation of demand variability is preferred over a likely large overestimation.

- The actual measurements of **monitored small and large customers** are directly used to model their demand – similarly based on the argument that a small underestimation of demand variability is preferred over a larger overestimation. The additional benefit here is that all particular characteristics of these customers are preserved by directly using their measured profiles.

### 5.1.3 Specification of the Demand Model

After the adopted demand modelling approach has been motivated in the previous subsection, it can now be specified in mathematical terms. A simplified grouping, in which each group comprises several customer categories, is adopted here in order not to obstruct the view with detail unnecessary for the core principle of the demand model. The grouping is based on the type of profile used to model the customers from the various categories which is the essential

**Table 5.1:** Major groups of customers which are distinguished in the description of the demand model. Each group contains several customer categories. The grouping is based on the type of profile which is assigned to the subsumed customer categories.

| Group | Description | Type of profile assigned and symbols used | Profile unit |
|---|---|---|---|
| 1 | Small customers, energy-behavioural classification and SJV binning | Random smart meter profile $\boldsymbol{s}_{b,\Pi} \in \mathcal{S}_b$, drawn from bin $b$ | - (norm.) |
| 2 | Mostly larger, commercial customers with telemetry | Measured profile $\boldsymbol{l}_j \in \mathcal{L}$ of customer $j$ | kW |
| 3 | Other customers, not part of group 1 and unmonitored | Average category profile $\boldsymbol{a}_c \in \mathcal{A}$ of category $c$ | - (norm.) |

here. An overview of the grouping is given in table 5.1, while a detailed break-down of all customer categories considered can be found in subsection 5.2.2.

Given is a set of smart meter profiles $\mathcal{S}$ of monitored customers of the Alliander grid. The smart meter profiles are anonymised, but the energy-behavioural category $c$ which each profile belongs to is known. The energy-behavioural categories are those of the smart meter data clustering approach outlined in subsection 3.1.1 which is used as part of the ANDES model. A binning is carried out based on the SJV $\gamma$ of all profiles within a category. The set of profiles within a bin $b$ is denoted by $\mathcal{S}_b$, thus $\mathcal{S}_b \subset \mathcal{S}$. Group 1 comprises 13 energy-behavioural categories which are split in 2-4 bins, leading to a total of 48 bins, therefore $b \in \{1, \ldots, 48\}$. A random smart meter profile drawn from bin $b$ is denoted by $\boldsymbol{s}_{b,\Pi}$, where the uniform discrete random variable $\Pi$ has the sample space $\Omega_\Pi = \{1, \ldots, n_b\}$ with $n_b$ being the number of smart meter profiles in the given bin.

Group 2 comprises several thousand individual larger, commercial customers whose power consumption is measured telemetrically. The set of telemetry profiles is denoted by $\mathcal{L}$ and individual customers within the set have the index $j$. Group 3 combines different types of profile categories totalling to 30 categories (see subsection 5.2.2 for further details). Together with the 13 energy-behavioural categories from group 1, a total of 43 customer categories is distinguished, leading to an overall category index $c \in \{1, \ldots, 43\}$. The set of average profiles for the 30 categories of group 3 is denoted by $\mathcal{A}$.

The smart meter profiles $\boldsymbol{s} \in \mathcal{S}$, telemetry profiles $\boldsymbol{l} \in \mathcal{L}$ and the average category profiles $\boldsymbol{a} \in \mathcal{A}$ are time series vectors and therefore set in bold, unless a specific time step is referred to:

$$\begin{aligned} \boldsymbol{s} &= \{s_t\} = \{s_1,\ s_2,\ \ldots,\ s_{35040}\} \\ \boldsymbol{l} &= \{l_t\} = \{l_1,\ l_2,\ \ldots,\ l_{35040}\} \\ \boldsymbol{a} &= \{a_t\} = \{a_1,\ a_2,\ \ldots,\ a_{35040}\}\ . \end{aligned} \tag{5.1}$$

Each time step of the meter profiles $\boldsymbol{s} \in \mathcal{S}$ and the average category profiles $\boldsymbol{a} \in \mathcal{A}$ is normalised with respect to the yearly total demand summed over all 35,040 quarter-hourly

time steps. Therefore,

$$\sum_{t=1}^{35{,}040} s_t = \sum_{t=1}^{35{,}040} a_t = 1 \,. \tag{5.2}$$

The SJV $\gamma$ is an *energy* quantity with the unit kWh, while the load profiles are time series of quarter-hourly average *power* values with the unit kW. Since one hour has four quarters, a factor of 1/4 or 4 arises when converting between energy and power in this case. Therefore, in order to obtain profiles in terms of power from the normalised profiles $\boldsymbol{s} \in \mathcal{S}$ and $\boldsymbol{a} \in \mathcal{A}$, which are scaled to match a given yearly energy consumption $\gamma$, the scaling factors

$$\begin{aligned} \boldsymbol{s}'_{power} &= 4 \cdot \gamma \cdot \boldsymbol{s} \\ \boldsymbol{a}'_{power} &= 4 \cdot \gamma \cdot \boldsymbol{a} \end{aligned} \tag{5.3}$$

are necessary. In the following $\boldsymbol{s}$ and $\boldsymbol{a}$ always refer to the normalised profiles.

The telemetry profiles $\boldsymbol{l} \in \mathcal{L}$ in turn are left unnormalised, as they do not need to be rescaled and can be used directly. Therefore, when properly scaled with the factor 1/4, all telemetry profiles sum to the yearly energy consumption $\gamma$ of the customer the profile belongs to:

$$\sum_{t=1}^{35{,}040} \frac{1}{4} \cdot l_t = \gamma \,. \tag{5.4}$$

It is now possible to specify the core principle of the adopted demand modelling approach. The stochastic power demand on a given network asset for all quarter-hourly time steps of a year is modelled by summing over all customers from the 3 groups summarised in table 5.1 connected to the asset:

$$\boldsymbol{D}(\boldsymbol{\Pi}) = \sum_{i=1}^{n_s} 4 \cdot \gamma_i \cdot \boldsymbol{s}_{b_i,\Pi_i} + \sum_{j=1}^{n_l} \boldsymbol{l}_j + \sum_{k=1}^{n_a} 4 \cdot \gamma_k \cdot \boldsymbol{a}_{c_k} \,, \tag{5.5}$$

where

- $\boldsymbol{D}(\boldsymbol{\Pi})$ is a resulting random annual asset demand trace or time series
$$\boldsymbol{D}(\boldsymbol{\Pi}) = \{D_t(\boldsymbol{\Pi})\} = \{D_1(\boldsymbol{\Pi}),\ D_2(\boldsymbol{\Pi}),\ \ldots,\ D_{35040}(\boldsymbol{\Pi})\} \,,$$
- $\boldsymbol{\Pi}$ is a random vector denoting a random selection of smart meter profiles,
- $\boldsymbol{s}_{b_i,\Pi_i}$ is a smart meter profile randomly drawn from bin $b_i$ of customer $i$,
- $\boldsymbol{l}_j$ is the telemetry profile of customer $j$,
- $\boldsymbol{a}_{c_k}$ is the average category profile of category $c_k$ of customer $k$,
- $\gamma_i$ and $\gamma_k$ are the yearly consumption SJV of customers $i$ and $k$, respectively and
- $n_s$, $n_l$ and $n_a$ denote the total number of customers in groups 1-3 of the modelled asset, respectively.

The random profile selection vector

$$\boldsymbol{\Pi} = (\Pi_1,\ \Pi_2,\ \ldots,\ \Pi_i,\ \ldots,\ \Pi_{n_s}) \tag{5.6}$$

has the length $n_s$ and indexes the smart meter profiles $s_{b_i,\Pi_i}$, randomly drawn from bin $b_i$ of each small unmonitored customer $i$ connected to the asset. The elements of $\boldsymbol{\Pi}$ are the uniform discrete random variables $\Pi_i$ which have the sample space $\Omega_{\Pi_i} = \{1, \ldots, n_{b_i}\}$, with $n_{b_i}$ being the number of smart meter profiles in bin $b_i$ which customer $i$ belongs to.

To sum up, while the customers from groups 2 and 3 are always assigned the same profiles, the random assignment of profiles $\boldsymbol{\Pi}$ to the customers of group 1 is the source of randomness in the asset demand model. Each evaluation of the model will therefore result (with very high probability) in a different annual realisation of asset demand. A small percentage of assets of the Alliander grid only have customers from groups 2 and 3. In these exceptional cases, the demand model looses its stochastic features and always produces the same asset demand trace.

### 5.1.4 Sample Space of the Demand Model

The explicit mathematical description of the demand model's sample space and the related considerations made in this subsection may be seen as a matter of general theoretical interest. However, more importantly for the objectives of this research, they provide a useful background for developing the Importance Sampling (IS) based methods in the next chapter 6.

The demand model allows obtaining the load demand on *individual* grid assets which means that each grid asset has its own sample space. The sample spaces of all assets can be seen as subspaces of the overall sample space of the demand model. Let $\Theta = \{\theta_1, \ldots, \theta_k, \ldots, \theta_{n_\theta}\}$ be the set of all assets of the Alliander grid, where $\theta_k$ denotes asset $k$ and $n_\theta$ the total number of assets. Formally, the overall sample space of the demand model $\Omega_{\boldsymbol{D}}$ can be defined as the union of the sample spaces of all assets:

$$\Omega_{\boldsymbol{D}} = \bigcup_{\theta_k \in \Theta} \Omega_{\theta_k} , \tag{5.7}$$

where $\Omega_{\theta_k}$ is the sample space of asset $k$ and $\Theta$ the set of all assets. In the following, however, only the sample space of a given individual asset is relevant. For greater ease of notation, the asset subscript is therefore omitted from now on and $\Omega$ refers to the sample space of an asset, which is a subspace of the demand model's sample space $\Omega_{\boldsymbol{D}}$.

For examining the sample space of an individual asset more closely, it is useful to start with the definition of the sample space concept: In general terms, a sample space is the set containing all possible outcomes of an experiment [28]. In this context, the experiment consists in evaluating the demand model $\boldsymbol{D}(\boldsymbol{\Pi})$ for a given asset and a random assignment of smart meter profiles $\boldsymbol{\Pi}$ to small unmonitored customers. Formally, the demand model can be seen as a function $\boldsymbol{D} : \Omega_{\boldsymbol{D}} \rightarrow \boldsymbol{D}(\boldsymbol{\pi})$ defined on its sample space $\Omega_{\boldsymbol{D}}$ which deterministically assigns an annual asset demand trace to each *specific* selection of smart meter profiles $\boldsymbol{\pi} \in \Omega \subset \Omega_{\boldsymbol{D}}$. Therefore, evaluating $\boldsymbol{D}(\boldsymbol{\pi})$ will always lead to the same result. Since, for a given asset, the outcome of the simulation experiment is entirely determined by the input profile selection, the sample space and the parameter space of the demand model are identical. Thus, the sample space can be defined in terms of the model inputs.

On the top level, a selection of smart meter profiles $\boldsymbol{\pi}$ results in the time series vector $\boldsymbol{D}(\boldsymbol{\pi})$ which contains the asset demand for each quarter-hour of a year. The quarter-hourly time steps of a year can be seen as a derived subspace $\Omega_t = \{t_1, t_2, \ldots, t_{35040}\}$, nested inside the top level profile selection space $\Omega_{\boldsymbol{\pi}}$. If the demand model is evaluated for a specific profile

selection and at a specific time step $t$, a single value $D_t(\boldsymbol{\pi})$ will result – whereas $D$ is not set in bold face anymore as the model returns a scalar now. Just like with $\Omega_D$, repeatedly evaluating $D_t(\boldsymbol{\pi})$ will always yield the same result.

To arrive at a definition of the overall asset sample space $\Omega$, a Cartesian product '$\times$' can be used. Given two sets, the Cartesian product yields all possible ways to combine each element of the first set with each element of the second set. In the context relevant here, all combinations of the profile selection space $\Omega_{\boldsymbol{\pi}}$ and the time step space $\Omega_t$ will result from their Cartesian product. These are the individual elements of the asset sample space $\Omega$ because are associated with a particular outcome of the random experiment in question here – evaluating the demand model. The asset sample space can thus be defined as:

$$\Omega = \Omega_{\boldsymbol{\pi}} \times \Omega_t = \{\boldsymbol{\pi}_1,\ \boldsymbol{\pi}_2,\ \ldots,\ \boldsymbol{\pi}_j,\ \ldots,\ \boldsymbol{\pi}_{n_\pi}\} \times \{t_1,\ t_2,\ \ldots,\ t_{35040}\}\,, \tag{5.8}$$

where

- $j \in \{1,\ \ldots,\ n_\pi\}$ indexes the $n_\pi$ possible ways of assigning smart meter profiles to small unmonitored customers of an asset,
- $\boldsymbol{\pi}_j = (\pi_{j,1},\ \pi_{j,2},\ \ldots,\ \pi_{j,i},\ \ldots,\ \pi_{j,n_s})$ is a vector characterising a specific profile selection,
- $i \in \{1,\ \ldots,\ n_s\}$ indexes the $n_s$ small unmonitored customers of the asset (group 3 from the previous section), therefore $\boldsymbol{\pi}_j$ contains one element for each of these customers, and
- $\pi_{j,i} \in \{1,\ \ldots,\ n_{b_i}\}$ are the elements of $\boldsymbol{\pi}_j$ which specify the selected smart meter profile from bin $b_i$ of customer $i$, whereas $n_{b_i}$ is the total amount of profiles in that bin.

The total number of elements $|\Omega|$ in the asset sample space can be obtained as a product over all 48 bins:

$$|\Omega| = |\Omega_{\boldsymbol{\pi}}| \cdot |\Omega_t| = \left(\prod_{b=1}^{48} (n_b)^{n_{s,b}}\right) \cdot 35040\,, \tag{5.9}$$

where $b$ indexes the current bin, $n_b$ is the total number of smart meter profiles in the bin and $n_{s,b}$ is the number of small unmonitored customers of the given asset which are assigned profiles from bin $b$. The important message to take away here is that, due to the power in equation (5.9), the number of elements in the asset sample space can quickly become an astronomical number. This makes the approach of summing over all sample space elements for the calculation of the risk metrics (as in equation (4.1) from the previous chapter) computationally infeasible.

## 5.2 Implementation of the Demand Model

In this section the process and the details of implementing the demand model are described. In the course of documenting the implementation, it will also become clear in which aspects the model is limited by practical constraints of data availability. This will allow to critically reflect on the model results in later chapters.

The implementation of the demand model, the methods for evaluating the risk metrics developed in chapter 6, as well as the analysis and visualisation of results in chapter 7 were carried out using the statistical computing language *R* version 3.6.1 [74]. The code was developed and run on a shared RStudio Server environment. A number of packages has been

used for various purposes: The packages *tidyverse* [96], *plyr* [95] and *reshape2* [94] for general data processing; the package *pracma* [14] for certain basic mathematical operations; the package *Rcpp* [32] to compile external *C++* code used for speeding-up matrix multiplications; the package *Brobdingnag* [43] for computations with very large numbers; the packages *binr* [50] and *dlookr* [80] for binning data; the package *extRemes* [39] for extreme value analyses; and finally *tidyverse* [96], *plotly* [85] and *ggpubr* [53] for plotting and visualisation. Furthermore, a function from the package *ggtern* [42] has been adapted for two-dimensional kernel density estimation. The packages used for very specific purposes are cited again below when the purpose in question is discussed.

### 5.2.1 Selection of Grid Assets to Be Modelled

Before proceeding with the description of the demand model implementation, the procedure for selecting subsets of grid assets for benchmarking the methods developed in chapter 6 is described. The likely underestimation of peak loads in the ANDES model is suspected to occur especially at the Medium Voltage (MV) and Low Voltage (LV) levels, where the variability and extreme features of individual customers weigh more in aggregates. This makes improving the modelling of these asset levels especially relevant. Therefore, the main focus lies on MV/LV substations, *Middenspanningsruimtes* (MSRs) in Dutch, which are the substations connecting the MV and the LV level. Additionally, it is also interesting to benchmark the methods on HV/MV substations, *Onderstations* (OSs) in Dutch, which are the substations at the top level of the Alliander grid connecting the MV level to the High Voltage (HV) level of the transmission grid operated by TenneT.

The main concern when selecting sets of MSRs and OSs was to ensure that they are well representative of the overall MSR and OS population of the Alliander grid. To accomplish this, two asset properties were used: the number of customers connected to an asset and the sum of their yearly consumption, *Standaard Jaarverbruik* (SJV). In figure 5.1 a scatter plot of the two variables is shown for the MSRs population. Therefore, each point in the plot corresponds to one MSR.

Two sets of MSRs were selected which are visualised in figure 5.1 – a set of MSRs with average characteristics and a set of MSRs with extreme characteristics in terms of the number of customers connected and/or the total yearly asset consumption. To arrive at this selection, the following steps were taken:

- Firstly, the covariance matrix $\Sigma$ of all data points shown in figure 5.1 was computed. Using the covariance matrix, a bivariate normal density $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ was set up centred on the median of each dimension.
- Secondly, a threshold was set at half the maximum value of the fitted density which separates MSRs with average (blue) and extreme (red) characteristics. Even though the MSRs with average characteristics only fill a small area of the plot, they account for 81.7% of all MSRs. Therefore, most MSRs have 'average' characteristics, whereas only about 20% are extreme in terms of customer number or yearly asset consumption.
- Thirdly, the bivariate normal density was used to assign a sampling probability to each MSR depending on its location in the space shown in figure 5.1. For the average set, $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$

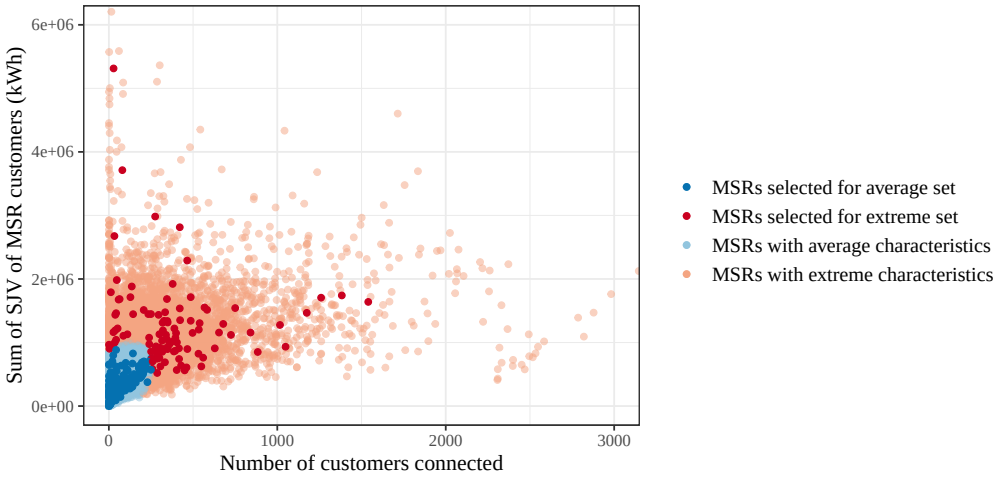


**Figure 5.1:** Characteristics of the overall MSR population and the MSRs selected for the average and extreme set. A few extreme outliers from the light red group are not shown for better visibility of the remaining data.

was used directly to sample 200 MSRs (dark blue) from the set of all MSRs with average characteristics (light blue). For the extreme set, $1 - \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$[10] was used to assign sampling probabilities to the MSRs with extreme characteristics (light red) and sample 100 MSRs (dark red). This procedure allowed to create two distinct groups which represent either MSRs with typical characteristics or MSRs with more extreme characteristics.

- Fourthly, the demand model was evaluated for each of the 300 pre-selected MSRs of both sets and the sum of the yearly asset demand was computed. In the demand model as well as in the ANDES model, all assigned profiles are scaled to match the known SJV of each customer (except for the telemetry profiles which already have the correct magnitude). Therefore, it is expected that the cumulative yearly consumption of the demand model should match the value from the results of ANDES. For 98% of the 300 pre-selected MSRs this expectation could be confirmed. However, in a few cases a bigger deviation was observed. In order to only keep MSRs where the magnitude of the results is in accordance with the ANDES results, MSRs whose cumulative yearly consumption obtained with the demand model was found to deviate more than 0.01% from the ANDES cumulative yearly consumption were filtered out. This was the case for 6 MSRs which corresponds to 2% of the 300 pre-selected MSRs.

- Lastly, the final average set was formed by randomly sampling 100 MSRs from the remaining pool of pre-selected average MSRs. The extreme set was formed analogously by randomly sampling 50 MSRs from the remaining pool of pre-selected extreme MSRs.

For the selection of OSs, the same procedure was followed until the third step leading to

**10** The density $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ was left unnormalised, 1 is therefore its maximum value and $1 - \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ inverts all probabilities.

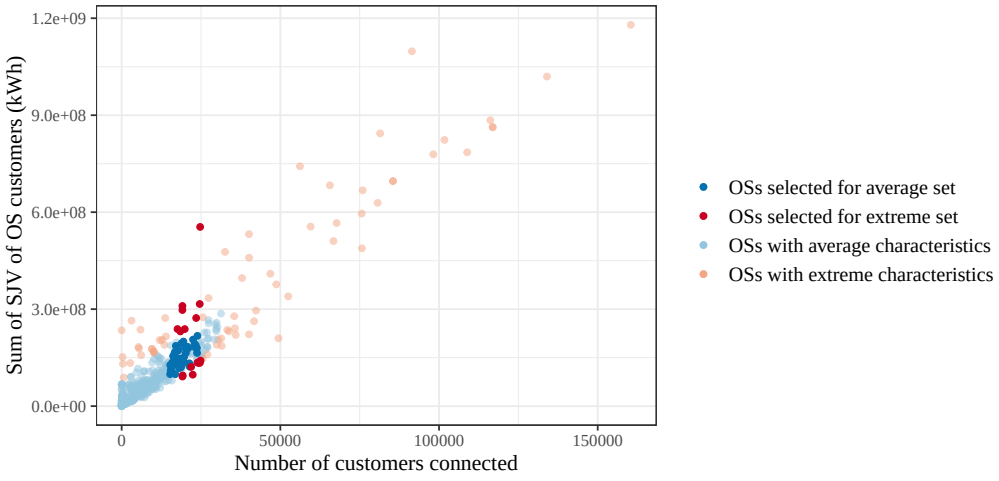


**Figure 5.2:** Characteristics of the overall OS population and the OSs selected for the average and extreme set in the window 15,000 – 25,000 connected customers.

a pre-selection of 282 OSs (the number 282 instead of 300 is due to the fact that only 82 OS were found to have extreme characteristics after the thresholding step). The fourth step could not carried out for OSs because in ANDES substation measurements are used to rescale the obtained OS load curves. This precludes a direct comparison with the results of the demand model. Due to the fact that OSs are not in the main focus of this research, as explained above, a smaller final set was selected from the 282 pre-selected OSs in the fifth step. Hereby, two considerations were made: On the one hand, it is desirable to select OSs with a large number of customers. On the other hand, computation times of the demand model increase in an approximately linear fashion with the number of customers connected to an asset. To arrive at a good balance in this trade-off, all of the pre-selected OSs who are connected to 15,000 – 25,000 customers were chosen for the two final average and extreme sets. The final selection comprising 41 average and 16 extreme OSs is visualised in figure 5.2.

### 5.2.2 Pre-processing of the Model Input Data and Implementation

In this subsection, the details of the demand model implementation and the choices made during the process are described. The core modelling principle consists in the assignment of various types of demand profiles to customers and is specified mathematically in subsection 5.1.3. Hereby, three major groups of customers are distinguished – small unmonitored customers, telemetry customer and miscellaneous unmonitored customers not part of the first group. The three groups comprise several customer categories which are broken down in detail in table 5.2. While the telemetry measurements used for group 2 customers and the average profiles used for group 3 customers were readily available, preparing the smart meter profile set used for group 1 customers involved several filtering and one binning step. Initially, a set of 3,773 anonymised and unnormalised smart meter profiles of Alliander customers was given. The steps taken and the choices made in processing the data set are described in the following:

**Table 5.2:** Detailed breakdown of the customer categories and bins distinguished within the three major groups used for the demand model. Note that 'E1 Cluster 8' and 'E2 Cluster 7' are only subdivided in two bins due to the smaller number of available smart meter profiles in these categories. For the energy-behavioural categories 'E1 Cluster 7' and 'E2 Cluster 3' no smart meter profiles are available at all, therefore they were moved to group 3 and are modelled using the respective average category profiles from ANDES. The Netherlands Chamber of Commerce, *Kamer van Koophandel* (KvK) segments are based on the Dutch standard classification of economic activities and the *Nederlandse Energiedatauitwisseling* (NEDU) categories are based on a classification of grid connection ratings.

| Group | Description | Customer categories | Bins |
|---|---|---|---|
| 1 | Small unmonitored customers, classified on the basis of energy-behavioural categories and modelled through smart meter profiles, binning based on the yearly consumption *Standaard Jaarverbruik* (SJV). | E1 Cluster 1 | 1-4 |
| | | E1 Cluster 2 | 5-8 |
| | | E1 Cluster 3 | 9-12 |
| | | E1 Cluster 4 | 13-16 |
| | | E1 Cluster 5 | 17-20 |
| | | E1 Cluster 6 | 21-24 |
| | | E1 Cluster 8 | 25-26 |
| | | E2 Cluster 1 | 27-30 |
| | | E2 Cluster 2 | 31-34 |
| | | E2 Cluster 4 | 35-38 |
| | | E2 Cluster 5 | 39-42 |
| | | E2 Cluster 6 | 43-46 |
| | | E2 Cluster 7 | 47-48 |
| 2 | Customers with telemetry measurements which are used directly, comprises mostly larger, commercial customers. | - | - |
| 3 | Unmonitored customers which are not part of group 1, modelled through average category profiles. | KvK segments 1-20<br>NEDU categories 1-8<br>E1 Cluster 7<br>E2 Cluster 3 | - |

### Step 1 – Filtering out profiles with missing values

It was noted that a small percentage of profiles contains a considerable number of missing values. In order to improve the data quality of the overall set, all profiles with missing values in more than 10 time steps were filtered out, leading to the removal of 142 profiles (3,631 profiles are left in the set).

### Step 2 – Filtering out profiles with consecutive zeros

A few profiles contain a high number of consecutive zeros which are most likely faulty measurements. All profiles containing more than one week of consecutive zeroes (therefore consecutive zeroes in more than 672 time steps) were filtered out, leading to the removal of 41 profiles (3,590 profiles are left in the set).

### Step 3 – Considerations regarding PV profiles

Potentially, there could be a correlation between smart meter data availability and the presence of Solar Photovoltaic (PV) installations, in the sense that people who own a PV system may be more likely to also own a smart meter. This could result in an over-representation of demand profiles with PV characteristics in the smart meter data set. Due to the frequently negative power flows around noon when PV generation is highest and the high simultaneity of the effect, an over-representation of PV profiles is likely to have a distorting impact on the baseload modelling.

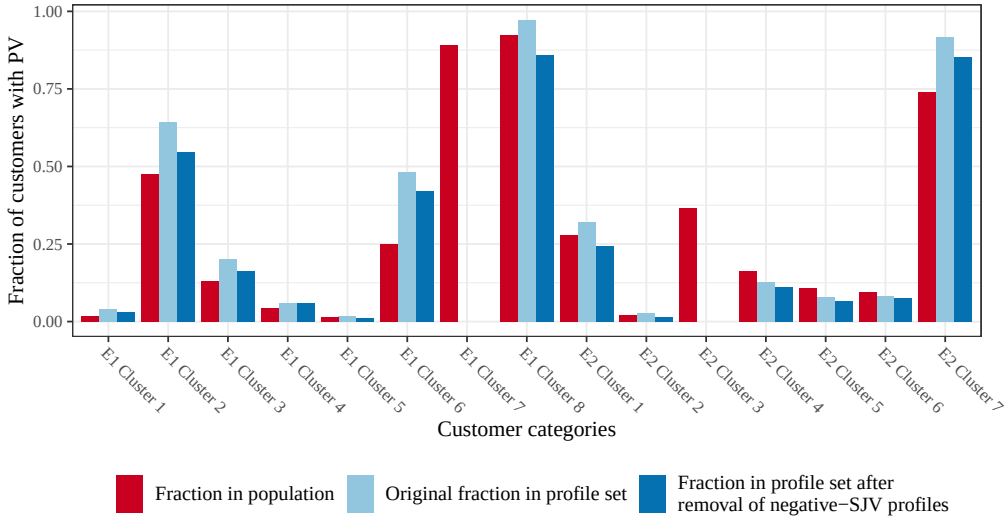


**Figure 5.3:** Comparison for each energy-behavioural customer category of the share of home-owned PV systems in the Dutch population with the fraction of PV profiles in the smart meter data set (detected by filtering for profiles with more than 350 negative values). Note that for 'E1 Cluster 7' and 'E2 Cluster 3' no smart meter profiles are available in the given set, therefore the corresponding bars are missing in the chart.

To investigate whether profiles with PV characteristics are over-represented in the smart

meter data set, using a simple decision rule firstly it was determined which profiles should be considered PV profiles: The decision rule consisted in regarding profiles with negative values in more than 350 time steps as PV profiles. Additionally, the 20 profiles with negative values in less than 350 time steps were inspected manually and six out of these were found to show PV profile characteristics (in terms of overall shape and timing of the troughs). These six special cases were also included in the set of PV profiles, the negative values in the remaining 14 profiles appeared to rather be faulty measurements. Of course this simple decision rule is somewhat problematic, as customers with a high consumption and in relation to that small PV system may never exhibit negative demand.

As a means of comparison, the fraction of customers with PV profiles in the Dutch population was used. The source of this population data used within Alliander is the website *Energieleveren* [33] which facilitates the compulsory registration of home-owned PV systems. In the data set used here, the customers registering their PV system on *Energieleveren* had already been assigned to the customer categories of group 1 (see table 5.2) using the supervised classification algorithm developed for the ANDES model. In figure 5.3, the fraction of customers owning a PV system in the population (red) are compared to the original fraction of PV profiles (light blue) detected in the smart meter profile set using the decision rule described above. The breakdown in categories shows that PV profiles indeed appear to be over-represented in some categories, especially in 'E1 Cluster 2', 'E1 Cluster 6' and 'E2 Cluster 7'.

However, it is questionable whether the two fractions – either obtained from population data or the count of PV profiles in the smart meter data set – rely on a similar enough definition of PV customers. For this reason it is unclear whether the comparison could serve as the basis for any potential correction of the PV over-representation issue. A further complicating aspect is that in the grid topology data used for ANDES and likewise the demand model only positive

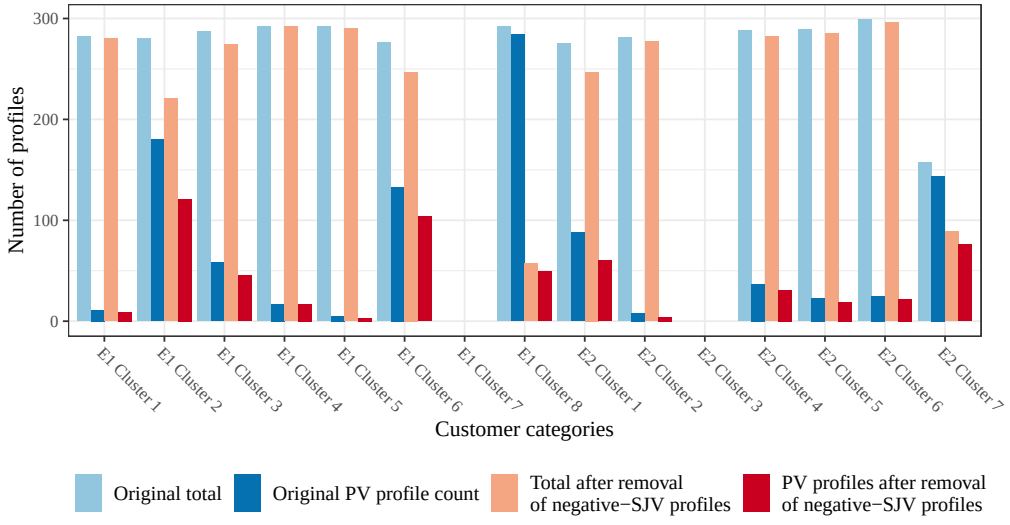


**Figure 5.4:** Comparison of the total count and the count of PV profiles before and after the removal of negative-SJV profiles per energy behavioural-customer category.

SJVs are given for the customers. The yearly energy consumption of some smart meter profiles was however found to be negative. In the demand model all profiles are scaled to match the SJVs of the customers which are modelled. This entails that smart meter profiles with a negative yearly consumption cannot be used because any transformation to make the yearly consumption positive would alter the original data significantly. Therefore, all negative-SJV profiles were removed from the smart meter data set. The fraction of PV profiles detected in the smart meter data set using the simple decision rule after the removal of negative-SJV profiles are shown in figure 5.3 in dark blue.

It can be observed in figure 5.3 that the over-representation of PV profiles is either resolved or diminished after the removal of the negative-SJV profiles. For this reason and more importantly due to the more principle concerns regarding the validity of a correction based on the comparison with the population fractions, the decision was taken to proceed with the set of smart meter profiles obtained after the removal of negative-SJV profiles. In some categories quite a large number of profiles were removed which can be observed in figure 5.4, most extremely in 'E1 Cluster 8' but also in 'E2 Cluster 7' and 'E1 Cluster 2'. Due to the lack of data for 'E1 Cluster 7' and 'E2 Cluster 3', customers from these categories are modelled through the average profiles used in ANDES (therefore the two categories are part of group 3 in table 5.2). Overall 453 negative–SJV profiles were removed, resulting in the final set of 3,137 smart meter profiles used for the demand model, obtained after the three described filtering steps.

**Step 4 – Binning of the smart meter and asset customer data**

As described in subsection 5.1.3, the customers from the energy-behavioural categories of group 1 in table 5.2 are randomly assigned smart meter profiles from the bin they are part of. The formation of the bins was carried out as the final pre-processing step of the model input data. Hereby, it was necessary to bin the small unmonitored customers to be modelled as well as the smart meter profile set used to model the former. For both, the binning variable was the annual consumption SJV.

Being a robust approach for data with a potentially large spread, quantile binning was deemed appropriate for the given application. The binning of SJV quantiles was performed using the *binr* [50] package. The critical decision to make hereby was the number of bins. In the trade-off between too much and too little variability, a target number of 50-100 profiles per bin was chosen. As figure 5.4 shows, 11 customer categories contain more than 200 profiles in total, in these cases 4 bins were formed. The remaining 2 customer categories with less than 100 profiles in total, 'E1 Cluster 8' and 'E2 Cluster 7', were subdivided in 2 bins. The number of bins determined for the smart meter profile set was then also used to bin all customers of the assets selected for modelling in subsection 5.2.1. The results of the binning for two exemplary customer categories are visualised in figure 5.5.

For the exemplary customer category shown in the top and bottom left panels of figure 5.5, the SJV range of the available smart meter data covers the SJV range of the customers to be modelled fairly well. Clearly this is not the case for the exemplary customer category shown on in the top and bottom right panels of the figure. Visualisations of this kind were inspected for all 13 customer categories subdivided in quantile bins. It was found that only for 'E1 Cluster 1' shown on the left the modelled customer and smart meter SJV ranges were in relatively good accordance. For all other customer categories the appearance was found to be similar to

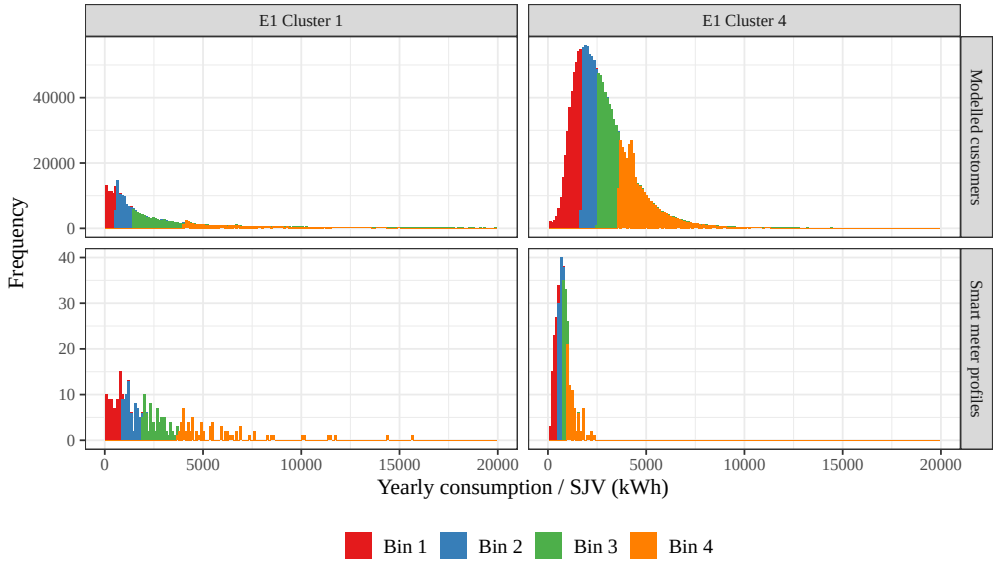


**Figure 5.5:** Results of the quantile binning according to the annual consumption of modelled customers and the smart meter profiles used to model them. Two exemplary customer categories are shown. Note the different y-axes on the top and the bottom, due to the fact that the number of customers to be modelled is much larger than the number of smart meter profiles available.

that of 'E1 Cluster 4' shown on the right (to a more or less extreme extent depending on the category). This implies that profiles from bin 3 or 4 are still scaled up significantly to match the SJV of the customers to be modelled. The over-scaling of spiky profiles was identified by Valckx et al. as a potential cause for the overestimation of peak loads which was found to occur when applying their Monte Carlo simulation approach [92]. Therefore, while the binning approach of this thesis may theoretically be able to alleviate this issue, the SJV range of the available smart meter data is likely to limit its effectiveness in practice.

### 5.2.3 Validation of the Demand Model Implementation

After the pre-processing steps described in the previous subsection, the demand model was implemented according to the principles specified in subsection 5.1.3. The primary sanity check of the demand model implementation was already mentioned in subsection 5.2.1: As the assigned profiles are scaled to match the known SJV of each customer in ANDES as well as in the demand model, the expectation is that the cumulative yearly asset consumption should be in accordance as well. For 98% of the 300 pre-selected MSRs a deviation of the demand model's yearly asset consumption of less than 0.01% from the ANDES result was found. In the six cases where bigger deviations were observed, small differences in the topological data used in the demand model and in ANDES are a possible cause. The same sanity check could not be conducted for OS because in ANDES substation measurements are used to rescale the

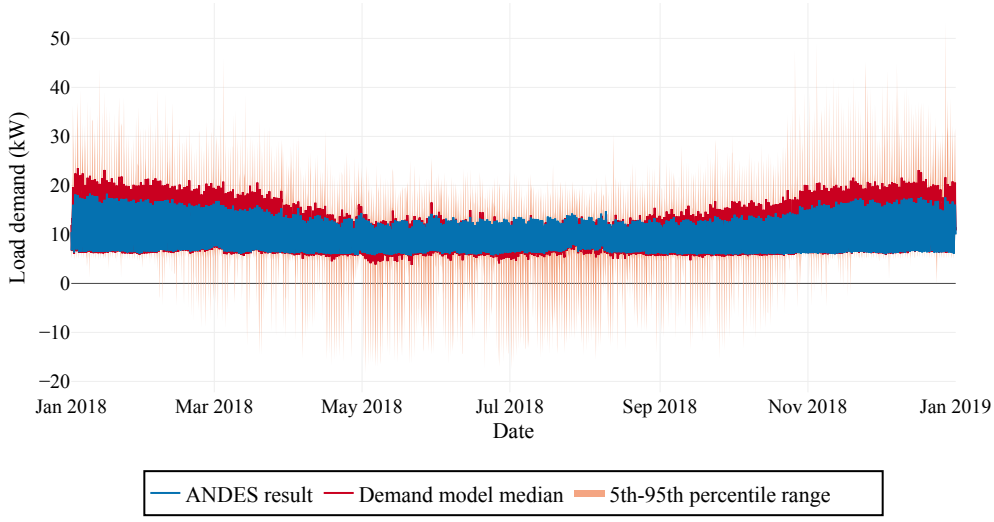


**Figure 5.6:** Comparison of the ANDES model result with the demand model result for an exemplary MSR with 41 customers. The demand model was evaluated 500 times and the shown median as well as the 5th-95th percentile range were calculated per time step.

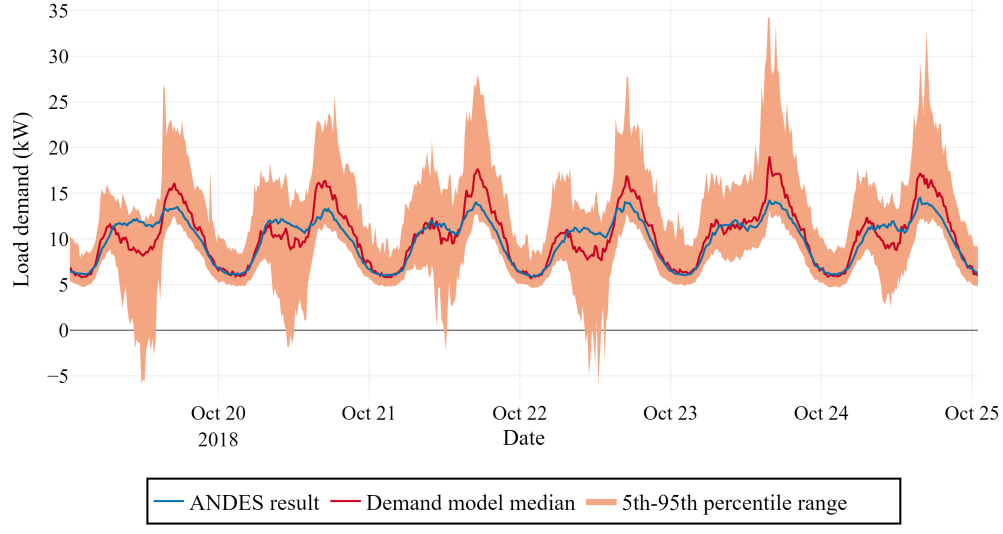


**Figure 5.7:** Detail of the comparison shown in the previous figure of the ANDES model result with the demand model result for the same MSR with 41 customers.

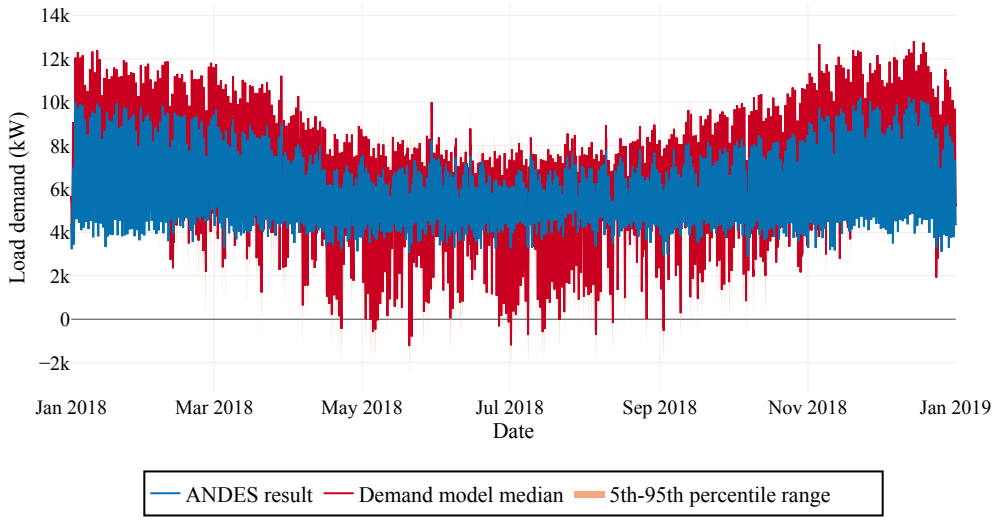


**Figure 5.8:** Comparison of the ANDES model result with the demand model result for an exemplary OS with 7,505 customers. The demand model was evaluated 500 times and the median as well as the 5th-95th percentile range were calculated per time step.

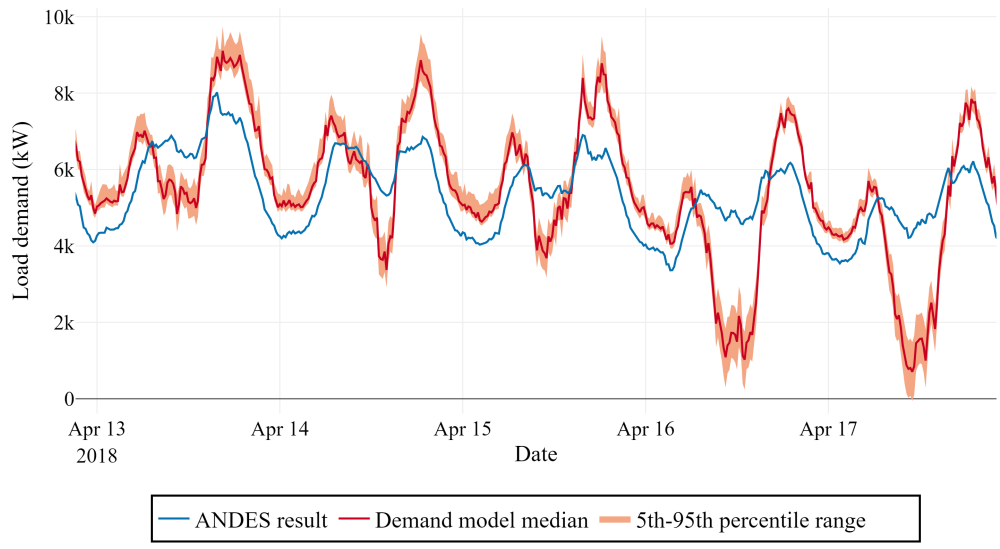


**Figure 5.9:** Detail of the comparison shown in the previous figure of the ANDES model result with the demand model result for the same OS with 7,505 customers.

obtained OS load curves. Overall, however, the very good accordance for the MSRs in 98% of the cases was seen as sufficient evidence that the model was implemented correctly.

More interesting and telling than the basic sanity check is a visual comparison of the demand model results with ANDES results for an example MSR and OS. The first comparison for an MSR with 41 customers is shown for the entire year in figure 5.6 and for a few exemplary days in figure 5.7. The figures reveal that the demand model results in a range of possible yearly realisations which show higher peaks in the winter and lower troughs in the summer. At the same time, the median is still in a comparable range to the ANDES demand curve. This suggests that the model fulfils the requirement of introducing stochasticity into the modelling of baseload demand and furthermore has desirable features in terms of demand peaks and troughs.

The second comparison for an OS with 7,505 customers is shown for the entire year in figure 5.8 and for a few exemplary days in figure 5.9. The median shows that peak demand in winter and summer is much more pronounced than in the ANDES result, therefore peak demand is very unlikely to be underestimated when evaluating the demand model in this case and likewise for the example MSR (it might however be overestimated to some extent). Furthermore, it is interesting to compare the magnitude of the variability for the MSR in figure 5.7 and the OS in figure 5.9: The variability is much smaller in the latter case which is in accordance with the expectation because for bigger assets the stochastic behaviour of individual customers is averaged out to a greater extent.

## 5.3 Risk Metrics

This section is concerned with choosing relevant risk metrics to quantify the performance and behaviour of grid assets which are modelled following the approach described in the previous sections of this chapter. Firstly, the chosen risk metrics are motivated subsection 5.3.1. Hereby, three alternative ways of specifying the risk metrics 1 and 2 are given a look to justify the chosen alternative. Secondly, the three risk metrics are specified in subsection 5.3.2.

For the definition of the risk metrics 1 and 2, impact functions are used which were introduced already in generic form in section 4.1. Generally, impact functions quantify the contribution or impact of each sampled system state to a risk metric. This definition can now be illustrated for the specific case of the demand model. An impact function

$$M(\boldsymbol{\pi}, t) : \Omega \rightarrow \mathbb{R} \tag{5.10}$$

defined on the sample space $\Omega$ of the demand model assigns a numerical outcome to every *specific* state of the demand model $(\boldsymbol{\pi}, t) \in \Omega$, where $(\boldsymbol{\pi}, t)$ denotes a specific smart meter profile selection $\boldsymbol{\pi}$ evaluated at time step $t$. The impact function may also assign the same outcome to several system states, for example all time steps of a specific profile selection could have the same impact. For a *random* state of the demand model $(\boldsymbol{\Pi}, T)$, the outcome of the impact function

$$M = M(\boldsymbol{\Pi}, T) \tag{5.11}$$

becomes a random variable itself.

### 5.3.1 Motivation and Alternatives in Specifying the Risk Metrics

To briefly reiterate, the questions which the risk metrics are require to address are:

- **Risk metric 1:** *For a given grid asset, what is the probability per time step to exceed a critical power level above the rated capacity?* This critical power level lies above the rated power capacity, it is for example 110% for OSs.
- **Risk metric 2:** *For a given grid asset, what is the expected total duration of exceeding the rated power capacity per year?* For OSs, for example, the threshold which should not be exceeded are 87.5 hours.
- **Risk metric 3:** *How high can the maximum demand (and how low can the minimum demand) on a given asset become?* There are no standardised asset level criteria for this question as for the previous two questions, because this question serves to find out how much additional capacity is necessary to prevent or limit overloads to a desired frequency.

Considering the background of Extreme Value Theory (EVT) exposed in section 4.2, the third risk metric points rather clearly to the quantiles of the Generalised Extreme Value (GEV). By indicated the load level which will not be exceeded with a prescribed probability, they allow answering the third question how high peak demand can potentially become.

For the first two risk metrics in turn, three alternative formulations are possible which are somewhat similar to each other but have different implications. Here the critical power level $d_{crit}$ of risk metric 1 is used to illustrate the alternatives. The formulation is analogous for risk metric 2 and can be obtained by replacing $d_{crit}$ with the rated power capacity of an asset $d_{cap}$. The alternatives are:

- **Alternative 1:** Randomly sampling individual states or 'snapshot' values $D_T(\boldsymbol{\Pi})$ of the demand model, leading to a definition of the risk metric $r$ of the form

$$r = \mathbb{E}[M(\boldsymbol{\Pi}, T)] = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d_{crit}}] \,, \tag{5.12}$$

  where $\mathbb{1}_A$ is the indicator function which returns a value of 1 whenever event $A$ occurs and 0 otherwise. The indicator function has the useful property $\mathbb{E}[\mathbb{1}_A] = Pr(A)$, therefore its expectation is equal to the probability of event $A$ occurring. Thus, the expectation in equation (5.12) yields the average probability of exceeding the critical capacity $d_{crit}$ at any given *time step*.

- **Alternative 2:** When randomly sampling full annual traces $\boldsymbol{D}(\boldsymbol{\Pi})$ of the demand model, a risk metric of the form

$$r = \mathbb{E}[M(\boldsymbol{\Pi})] = \mathbb{E}[\mathbb{1}_{f_{LD}(\boldsymbol{D}(\boldsymbol{\Pi})) > d_{crit}}] \,, \tag{5.13}$$

  could be defined, where $f_{LD}$ denotes the function which returns the load duration curve of a given random annual demand trace $\boldsymbol{D}(\boldsymbol{\Pi})$. In this case, the impact of a full annual trace is $M(\boldsymbol{\Pi})$ is computed and the time index is not needed. Whenever the critical capacity is exceeded in a yearly trace, the impact metric will take on the value of 1. The expectation in equation (5.13) yields therefore the average probability of exceeding the critical capacity $d_{crit}$ in any given *year*.

- **Alternative 3:** When randomly sampling full annual traces, the expectation of the random annual demand traces $\boldsymbol{D}(\boldsymbol{\Pi})$ could also directly be taken, to define a risk metric of the form

$$r = \frac{\sum_{t=1}^{35{,}040} \mathbb{1}_{\mathbb{E}[D_t(\boldsymbol{\Pi})] > d_{crit}}}{35040} \,. \tag{5.14}$$

  The expectation $\mathbb{E}[\boldsymbol{D}(\boldsymbol{\Pi})]$ results in a single trace of expected demand per time step. This is equivalent to using average annual demand profiles which are the same for each model evaluation. This means that the fraction of time steps which exceed the critical capacity $d_{crit}$ is a fixed number.

Alternative 3 relies on average profiles and corresponds closely to what is currently done in the ANDES model. It can therefore be ruled out, due to the issue of underestimating peak loads discussed in chapter 3. To decide between alternatives 1 and 2, it is useful to consider the relative merits and drawbacks: The major disadvantage of alternative 1 is that it does not convey any information on what happens in a given year. For example, if an average probability of exceeding the critical capacity corresponding to 5 hours per year is obtained, it is not clear whether 5 hours of exceeding $d_{crit}$ take place every year or if $d_{crit}$ is exceeded for 50 hours every 10 years. Alternative 2, in turn, gives the probability of exceeding $d_{crit}$ in a given year and therefore the two cases would have different probabilities. The major drawback of alternative 2 is, however, that it is blind to what happens within the year. For example, a year where $d_{crit}$ is exceeded in all time steps would weigh the same as a year where it is exceeded only in one time step.

On the practical side, a major advantage of metric 1 is that it is usually computationally much less expansive as compared to alternative 2. In conjunction with the drawback of alternative 2 of not differentiating the magnitude of the exceedance within a year, this was the decisive argument in favour of alternative 1. Therefore, alternative 1 – sampling individual states or 'snapshot' values – was chosen.

### 5.3.2 Specification of the Risk Metrics

For the remediation of congestion issues, it matters whether the congestion occurs because the energy *demanded* by customers cannot be transported to them (risk of positive overload) or the energy *generated* by customers with PV systems cannot be transported away from the customers (risk of negative overload). Therefore, separate sub-metrics targeting high and low demand values are introduced which mirror each other and are distinguished with the subscripts '+' and '−'. The risk metrics are specified and explained in the following:

- **Risk metric 1:** The probability of the stochastic power demand $D_T(\boldsymbol{\Pi})$, for a random profile selection $\boldsymbol{\Pi}$ and at a random time step $T$, to exceed a critical value $d_{crit}$ which is defined as the power rating of an asset $d_{cap}$ plus a percentage margin. Within this thesis, the percentage margin for OSs of 10% above which an overload becomes critical is used, thus yielding $d_{crit} = d_{cap} + 0.1 \cdot d_{cap}$:

$$r_{1,+} = \mathbb{E}[M_{1,+}(\boldsymbol{\Pi}, T)] = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d_{crit}}] \,, \tag{5.15}$$

$$r_{1,-} = \mathbb{E}[M_{1,-}(\boldsymbol{\Pi}, T)] = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) < -d_{crit}}] \,. \tag{5.16}$$

- **Risk metric 2:** The probability of the stochastic power demand $D_T(\boldsymbol{\Pi})$ to exceed the power rating of an asset $d_{cap}$. Formulated as below, this probability can be converted into the expected overload duration which may be a more tangible quantity. For OSs, the threshold which should not be exceeded are 87.5 hours:

$$r_{2,+} = \mathbb{E}[M_{2,+}(\boldsymbol{\Pi}, T)] = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d_{cap}}] , \tag{5.17}$$

$$r_{2,-} = \mathbb{E}[M_{2,-}(\boldsymbol{\Pi}, T)] = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) < -d_{cap}}] . \tag{5.18}$$

- **Risk metric 3:** The power value which will not be exceeded with 95% probability by the annual maxima, and equivalently the power value which will not be undercut with 95% probability by the annual minima. These values are given by the quantiles of the Generalised Extreme Value (GEV) distribution, denoted by $Q_{\xi,\sigma,\mu,+}$ for the distribution of the maxima and by $Q_{\xi,\sigma,\mu,-}$ for the distribution of the minima. The GEV distribution has the shape parameter $\xi$, the scale parameter $\sigma$ and the location parameter $\mu$. From the quantile function, also return levels $RL^k$ can be obtained through the relation

$$\hat{RL}^k = Q_{\hat{\xi},\hat{\sigma},\hat{\mu}}(p = 1 - \tfrac{1}{k}) , \tag{5.19}$$

  where $k$ denotes the return period in blocks. The return level $\hat{RL}^k$ is the magnitude which the block maxima will exceed once on average during any period of length $k$. The return level corresponding to the 95th percentile is $k = 20$. The two sub-metrics of risk metric three can thus be specified as:

$$r_{3,+} = Q_{\xi,\sigma,\mu,+}(0.95) = Q_{\xi,\sigma,\mu,+}(1 - \tfrac{1}{20}) = RL_{+}^{20} \tag{5.20}$$

$$r_{3,-} = Q_{\xi,\sigma,\mu,-}(0.95) = Q_{\xi,\sigma,\mu,-}(1 - \tfrac{1}{20}) = RL_{-}^{20} \tag{5.21}$$

## 5.4 Conclusion

In section 5.1, the methodological core of the demand model was motivated and mathematically described. Hereby, especially the binning of smart meter profiles and modelled customers according to their yearly consumption can be considered a novel aspect with respect to the related previous thesis work for Alliander in [92]. Theoretically, the binning has the potential to address the peak overestimation issue which was observed in the cited previous work. In the course of describing the implementation related aspects in section 5.2, it however became clear that the given set of smart meter profiles used in this thesis – which has a much smaller yearly consumption range compared to the customers which are to be modelled for most categories – may limit the effectiveness of the binning approach in practice. Aside from this challenging aspect, the demand model was found to fulfil the requirements specified in subsection 3.2.2 and to exhibit the desirable stochasticity in its peak demand characteristics. In section 5.3, after explaining why the specification of risk metrics 1 and 2 in terms of sampling 'snapshot' demand values was chosen, the chapter was concluded with the specification of the risk metrics.

# 6 Development of Evaluation Methods for the Risk Metrics 1 and 2

After specifying the demand model and identifying the risk metrics of interest to be obtained from its outputs in the previous chapter, the attention can now be focused on the development of methods to obtain these. The focus of this chapter lies hereby on the development of Monte Carlo methods for the efficient evaluation of risk metrics 1 and 2. As detailed in subsection 5.1.4, the sample space of the demand model comprises of two subspace: the time step space $\Omega_t$ and the space of all possible way to assign smart meter profiles to small, unmonitored customers $\Omega_{\boldsymbol{\pi}}$. A visualisation of these subspaces is shown in figure 6.1. The idea of the Importance Sampling (IS) methods developed in the section, is to give more weights to certain states of the subspaces, either alone or in combination. This is done with the aim of increasing the frequency of the overload events of interest, in order to achieve a variance reduction in the estimation of risk metrics 1 and 2. An overview of all methods developed in this section and the abbreviations used for them is given in table 6.1. The first position in the abbreviations always refers to the method for assigning demand profiles to small unmonitored customers, while the second position refers to the selection of time steps (except in the case of IS-bw).

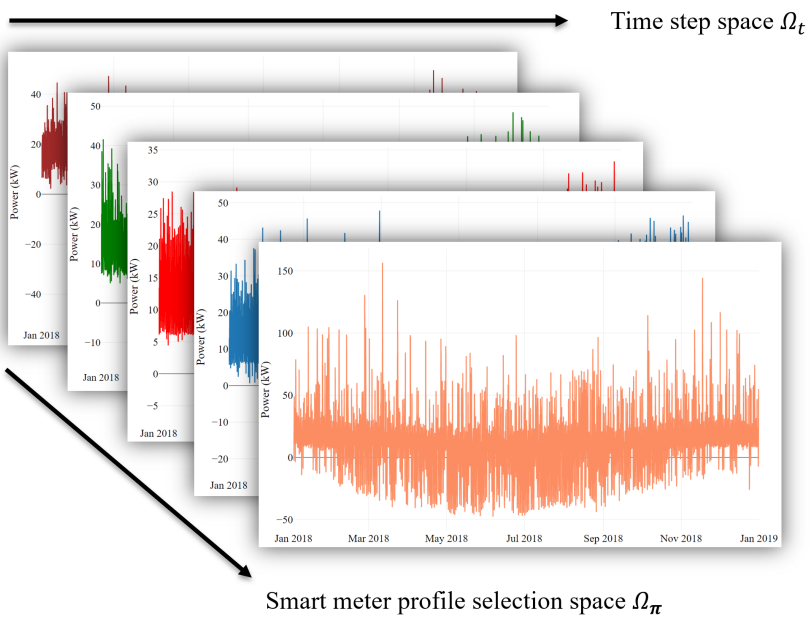


**Figure 6.1:** Visualisation of the sample space dimensions of the demand model over which importance sampling can be performed.

**Table 6.1:** Overview of the Monte Carlo methods developed in this section and the abbreviations used.

| Method Abbreviation | Description |
|---|---|
| MC-full | Random sampling of profiles to obtain entire annual traces. |
| MC-MC | Random sampling of profiles and time steps. |
| MC-IS | Random sampling of profiles and importance sampling of time steps. |
| IS-MC | Importance sampling of profiles and random sampling of time steps. |
| IS-IS | Importance sampling of profiles and importance sampling of time steps. |
| IS-bw | Importance sampling of profiles based on bin weights (bw) and random sampling of time steps. More consistent would be the abbreviation IS-MC-bw, but for brevity IS-bw is used. |

## 6.1 Reference Method – Sampling Full Annual Traces

The Advanced Net DEcision Support (ANDES) model works with yearly traces of 35,040 quarter-hourly load demand values. Being the application within Alliander which this research is geared to, the reference method here is the sampling of full annual traces of the same format to allow comparisons of direct practical relevance. The risk metrics 1 and 2 from subsection 5.3.2 in the previous chapter are, however, specified in the form of expectations over 'snapshot' random demand values $D_T(\boldsymbol{\Pi})$ from the demand model for a random selection of profiles $\boldsymbol{\Pi}$ evaluated at a random time step $T$. These risk metrics verify whether the snapshot demand values exceed certain thresholds $d$. In some sense this is the opposite of sampling full annual traces and therefore the question might arise how one relates to the other.

To reason about this matter, a generic risk metric

$$r = \mathbb{E}[M(\boldsymbol{\Pi}, T)] = \mathbb{E}[\,\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d}] \tag{6.1}$$

is used, which has the same structure as and is thus representative of risk metrics 1 and 2 (see subsection 5.3.2). Based on the impact function $M(\boldsymbol{\pi}, t)$ for snapshots, a new impact function $H(\boldsymbol{\pi})$ for entire or partial annual traces can be defined. For now, entire annual traces are considered:

$$H(\boldsymbol{\pi}) = \frac{1}{35,040} \sum_{t=1}^{35,040} M(\boldsymbol{\pi}, t) = \frac{1}{35,040} \sum_{t=1}^{35,040} \mathbb{1}_{D_t(\boldsymbol{\pi}) > d} \ . \tag{6.2}$$

Both impact functions allow obtaining risk metric $r$ by calculating the respective expectation,

$$r = \mathbb{E}[M(\boldsymbol{\Pi}, T)] = \mathbb{E}[H(\boldsymbol{\Pi})] \ , \tag{6.3}$$

because when enumerating all states in the sample space it does not matter whether this is done by means of random snapshots over both subspaces or by considering annual blocks of states. As described in section 4.1, expectations of functions can be estimated with the conventional Monte Carlo (MC) estimator. Therefore, if $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ is an independent and identically distributed (i.i.d.) sample of profile selections, the MC estimator based on sampling

full annual traces takes the form

$$\hat{R}_{MC-full} = \frac{1}{n}\sum_{j=1}^{n} H(\boldsymbol{\Pi}_j) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{35,040}\sum_{t=1}^{35,040} \mathbb{1}_{D_t(\boldsymbol{\Pi}_j) > d} \ . \tag{6.4}$$

If snapshot values were to be used directly, the estimator would be

$$\hat{R}_{snapshot} = \frac{1}{n}\sum_{j=1}^{n} M(\boldsymbol{\Pi}_j, T_j) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{D_{T_j}(\boldsymbol{\Pi}_j) > d} \ . \tag{6.5}$$

The important difference between these two cases is that while each snapshot value can be considered as an independent sample, this is not true for each time step of a trace. The reason for this is the correlation structure of time steps within traces. For example, it is possible that entire traces are especially low or high relative to other traces. In the first case, sampling an exceptionally high value would make it more likely to sample another exceptionally high value from the same trace compared to the case of sampling from a different random trace. Therefore, the coefficient of variation or relative error of the MC-full estimator should be estimated according to

$$\hat{RE}_{\hat{R}_{MC-full}} = \frac{S_H}{\sqrt{n}\cdot \hat{R}_{MC-full}} \ , \tag{6.6}$$

where $n$ denotes the number of *traces* sampled and $S_H$ is the sample standard deviation of the impact values $H(\boldsymbol{\Pi}_j)$, obtained by taking the square root of the sample variance

$$S_H^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(H(\boldsymbol{\Pi}_j) - \hat{R}_{MC-full}\right)^2 . \tag{6.7}$$

The error estimation works analogously for all following methods and is therefore only shown here.

## 6.2 Conventional Monte Carlo

Neither the computation of snapshot values, nor of full annual traces is particularly efficient computationally. On the one hand, sampling a selection of profiles takes a constant computational effort which has to be invested for each newly generated snapshot value. When sampling more than one time step per trace, the computational cost of drawing profiles and of assembling the traces can be redistributed over the sampled time steps. On the other hand, when using all time steps of a trace as in the MC-full method, all 35,040 time step values have to be handled entailing a much higher computational cost than handling one time step only. The idea of the conventional MC method of this subsection, is to find a reasonable middle-ground between these two extremes by sampling partial annual traces.

To this end, a vector of the time steps sampled per trace $\boldsymbol{\theta}$ is introduced as an additional variable in the impact function $H$. The vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_t, \ldots, \theta_m\}$, with $\theta_t \in \{1, \ldots, 35040\}$, contains a specific sample of time steps. The extended trace impact function can then be

defined as follows:

$$H(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{1}{m} \sum_{t=1}^{m} M(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{1}{m} \sum_{t=1}^{m} \mathbb{1}_{D_{\theta_t}(\boldsymbol{\pi}) > d} \ . \tag{6.8}$$

Let $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ be an i.i.d. sample of profile selections. In the conventional Monte Carlo method a new random sample of time steps $\boldsymbol{\Theta} = \{\Theta_1, \ldots, \Theta_m\}$ is drawn per trace. Therefore, random sampling over profile selections and traces is performed and the corresponding estimator takes the form:

$$\hat{R}_{MC-MC} = \frac{1}{n} \sum_{j=1}^{n} H(\boldsymbol{\Pi}_j, \boldsymbol{\Theta}) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{t=1}^{m} \mathbb{1}_{D_{\Theta_t}(\boldsymbol{\Pi}_j) > d} \ . \tag{6.9}$$

The MC-MC approach is likely to have a higher efficiency in obtaining estimates than the MC-full method in many cases, because only a trace and not a time step can be seen as an independent sample and the MC-MC method is faster in computing additional traces. This comes at the cost of missing some overload moments, however in contrast to the snapshot estimator still much more information is obtained per trace.

## 6.3 Time Step Importance Sampling

The motivation for performing time step IS links directly to the considerations regarding the MC-MC estimator at the end of the previous subsection: It is desirable to extract as much information as possible per sampled time step in order to lower the number of time steps sampled per trace. When sampling time steps randomly, it is possible that some overload events will be missed when sampling too few time steps. Under the assumption that overload moments occur with a higher chance at particular times during the year, a targeted sampling scheme, which gives more importance to some time steps, could lead to a higher efficiency of estimating the trace impact metric $H$.

The fundamental idea of the time step IS approach developed in this section is visualised in figure 6.2. The one-dimensional space of time steps is firstly transformed into the two-dimensional space shown in the figure. Inspiration for this transform came from [57], where a similar transformation is used for a singular value decomposition analysis. The line of reasoning for considering this two-dimensional space is that it allows aligning the times of the day with each other for the different seasons of the year. As it is likely that demand extremes occur during similar times of the day and close to each other in the season of the year, the two-dimensional space can be used to identify a suitable IS distribution.

For doing so, firstly conventional Monte Carlo sampling is carried out to observe where overload moments are located in the two-dimensional space (visualised as red dots in the example of figure 6.2). After a certain number of overload events has been collected and located, a kernel density estimation is carried out to extrapolate around the observed events. The kernel density estimation also has the desired effect, that regions where overload events are more densely situated will lead to larger values of the resulting density. A function from the *R* package *ggtern* [42] has been adapted for carrying out the two-dimensional kernel density estimation. The bandwidth of the kernel density estimation was tuned manually until a good interpolation between the overload events was observed. Two exemplary results of the

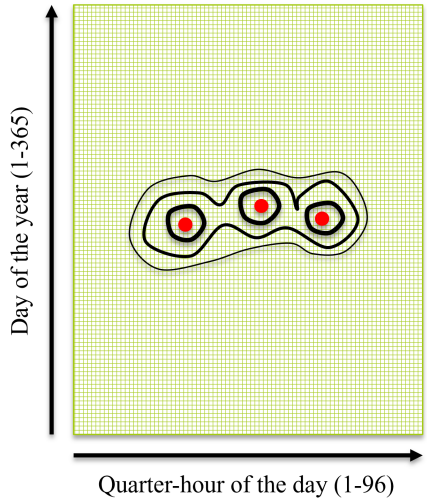


**Figure 6.2:** Visualisation of the two-dimensional time step space which is employed to estimate the IS distribution. The red dots are exemplary overload moments, while the black contour lines represent the result of the kernel density estimation with a Gaussian kernel.

kernel density estimation procedure are shown in figure 6.3. On the left, moments of positive overload and the kernel density estimated for these are shown, while on the right moments of negative overloads and the respective density are depicted. It is striking how different the patterns are. However, it corresponds well to the expectation that moments of high positive demand occur on winter evenings, while demand minima happen during the summer earlier in the day due to Solar Photovoltaic (PV) generation.

The resulting two-dimensional normalised density estimate is then transformed back to obtain a density value $\hat{g}(t)$ for each of the 35,040 time steps of the year. A fixed percentage of the maximum value of the density in the range 1-5% is used as the floor value $\hat{g}_{floor}$ of the density which cannot be undercut. Thus, whenever $\hat{g}(t) < \hat{g}_{floor}$, the original value $\hat{g}(t)$ will be replaced by $\hat{g}_{floor}$. This density $\hat{g}$ is now used as the IS distribution. As the original sampling density $f$ is uniform, it does not need a time dependence. For a time step $t$ the importance weights then take the form

$$W(t) = \frac{f}{\hat{g}(t)} = \frac{1/35040}{\hat{g}(t)} \,. \tag{6.10}$$

The trace impact function $H$ is adjusted to take the density $\hat{g}$ as a parameter. Let $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ be an i.i.d. sample of profile selections. To carry out time step IS a sample of time steps $\boldsymbol{\Theta} = \{\Theta_1, \ldots, \Theta_m\}$ is drawn from density $\hat{g}$ for each trace. The estimator for random profile selection in combination with time step importance sampling takes then the form:

$$\hat{R}_{MC-IS} = \frac{1}{n}\sum_{j=1}^{n} H(\boldsymbol{\Pi}_j, \boldsymbol{\Theta}; \hat{g}) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\sum_{t=1}^{m}\left(\mathbb{1}_{D_{\Theta_t}(\boldsymbol{\Pi}_j) > d}\right)\frac{f}{\hat{g}(\Theta_t)} \,. \tag{6.11}$$

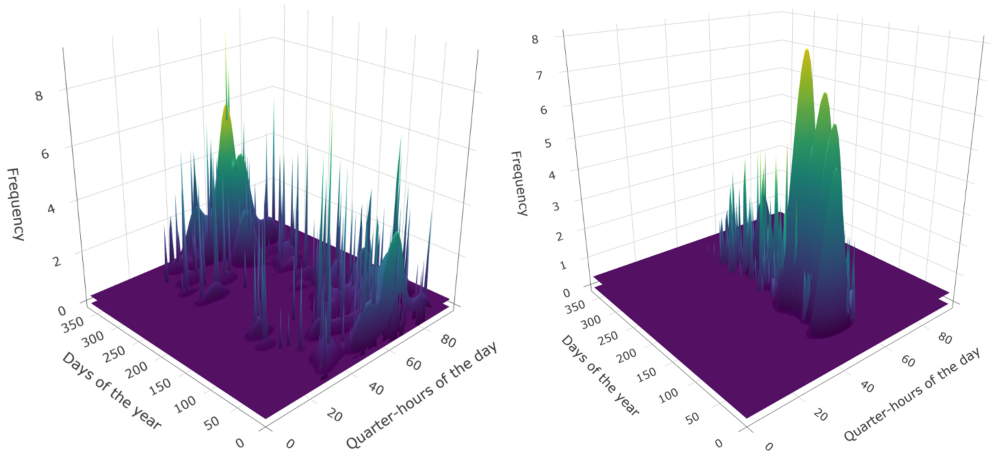


**Figure 6.3:** Two exemplary results of the kernel density estimation procedure in the employed two-dimensional space for a MV/LV substation are shown. The spikes show the location and frequency of overload moments, while the smooth surfaces visualise the resulting kernel density estimates (the densities are scaled for the purpose of visualisation and have, therefore, arbitrary units). On the left, moments of positive overloads can be observed, while on the right moments of negative overload are shown.

### 6.3.1 Description of the Sequential Time Step Importance Sampling Algorithm

A sequential algorithm has been develop which update the time step importance distribution in an initial optimisation stage. The algorithm relies on importance weights of the form,

$$W_k(\Theta_t) = \frac{f}{\hat{g}_k(\Theta_t)} = \frac{1/35040}{\hat{g}_k(\Theta_t)}\,, \tag{6.12}$$

where $k$ denotes the k-th step in the optimisation stage and $W_k(\Theta_t)$ is the corresponding weight. The steps of the algorithm are as follows:

1. Set $\hat{g}_0 = f$ to start with conventional MC sampling. Set the iteration counter k = 1. Set the initial optimisation threshold $d_{optim} = 0.5 \cdot d_{cap}$.

2. Generate a random sample of time steps $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{n_{optim}}$ which follows $\hat{g}_k(\cdot)$ for each of the $n_{optim}$ traces of the optimisation iteration. Also generate a random sample of profile selections $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_{n_{optim}}$. Using the samples, evaluate whether $\boldsymbol{D}_{\boldsymbol{\Theta}_j}(\boldsymbol{\Pi}(\boldsymbol{X}_j)) > d_{optim}$. Whenever that occurs, store the corresponding demand value in the vector $\boldsymbol{D}_{res,optim}$. Store the location in the two-dimensional space of all overload events $m_{event}$ and the corresponding IS weights $W_k$.

3. If the number of overloads events $m_{event} < m_{kde}$, repeat step 2. Otherwise carry out a weighted kernel density estimation using the stored IS weights for each overload event, respectively. Use the resulting estimated density $\hat{g}_{kde}$ as the IS density: $\hat{g}_k = \hat{g}_{kde}$.

4. Set the threshold $d_{optim} = Q_{D_{res,optim}}(1-\lambda)$ of demand values where $\lambda$ is a parameter indicating the 'ease' of raising the threshold. Set counter k = k + 1. If $d_{optim} > d_{cap}$ set k = K (final iteration) and proceed with step 5, otherwise go back to step 2.

5. Generate a random sample of time steps $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{n_{optim}}$ which follows $\hat{g}_K(\boldsymbol{\Theta}_j)$ for $n$ production run traces. Also generate a random sample of profile selections $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_{n_{optim}}$. Obtain the estimate by evaluating

$$\hat{R}_{MC-IS} = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\sum_{t=1}^{m}\left(\mathbb{1}_{D_{\Theta_t}(\boldsymbol{\Pi}_j) > d_{cap}}\right)\frac{f}{\hat{g}_K(\Theta_t)} \,. \tag{6.13}$$

6. Verify the convergence in blocks of 50 traces, calculate therefore the relative error $RE_{\hat{R}_{MC-IS}}$ of the estimate. If $RE_{\hat{R}_{MC-IS}} < 0.1$, terminate the algorithm. Otherwise go back to step 5.

## 6.4 Profile Selection Importance Sampling

The reasoning of why performing importance sampling on the profile selection space may lead to a variance reduction, relies on two key ideas: Firstly, it was observed in example cases and is generally plausible that very spiky profiles are crucially involved in causing overloads. Therefore, sampling profiles which deviate considerably from the average behaviour with higher probability, is likely to lead to more overload events. Secondly, when studying an example asset with 29 customers, it was observed that overloads occur frequently due to spiky-profiles being assigned to very few specific customers. In some cases, this may have to do with the over-scaling of smart meter profiles with a relatively small yearly consumption which occurs due to the limited amount of available smart meter profile (see subsection 5.2.2). Therefore, it is also more generally plausible that a small number of customers is mainly responsible for overload events – at least when it comes to assets with an overall relatively small number of customers. Whether this holds also for larger assets with a few hundred or thousand of customers is not in itself apparent and subject of further investigation in the next chapter.

The indications that only a limited number of customers may be mainly responsible for overload events are good news, because assigning the most spiky profiles to all customers is problematic. The reason why the latter approach is problematic is that assigning certain profiles with a much higher frequency to all customers can easily lead to very extreme importance sampling weights. For example, if the top 10% of the most spiky profiles are assigned to all 29 customers of the example asset with probability 0.8 instead of 0.1, a large number of importance weights would be of the order $10^{-27}$. Such extreme importance weights are likely to lead to estimates which are orders of magnitude to small initially and only converge to the correct order for very large sample sizes. Thus, it is desirable to assign the most spiky profiles with higher probability only to those customers which are mainly responsible for causing overload events, while randomly assigning all kinds of profiles to the remaining customers.

The question then arises, how to identify which customers are crucial in causing overload events, preferably in an automated way. More abstractly, the problem consists in automatically finding a suitable IS distribution for a given asset. For accomplishing this, the Cross-Entropy

(CE) method is an appropriate choice as it allows to obtain near-optimal IS distributions in a self-tuning manner. In the general description of the Cross-Entropy (CE) method in subsection 4.1.3, it became clear that the method optimises the parameters of an IS sampling distribution from a given family of *parametric* distributions. Therefore, it is necessary to firstly parametrise the problem at hand such that a parametric distribution can be used as the IS distribution.

### 6.4.1 Parametrising the Demand Model

The parametrisation chosen here is based on the idea to divide the smart meter profiles in each of the 48 bins used in the demand model in two sets – a set of very spiky profiles and a set of all other profiles with more average characteristics. This requires identifying which of the profiles in a bin are most spiky. By visual inspection of several examples, it was found that the following deviation metrics are well suited to accomplish the task: Let $\boldsymbol{s}_b = \{s_{b,t}\}$ be a smart meter profile from a given bin $b$ and $\tilde{\boldsymbol{s}}_b = \{\tilde{s}_{b,t}\}$ the median profile of that bin, obtained by computing the median of all profiles in the bin per time step. For greater ease of notation, the bin indices $b$ are dropped below. The metrics $\Delta_+$ and $\Delta_-$ have been calculated for each profile in each bin by evaluating

$$\Delta_+(\boldsymbol{s}) = \sum_{t=1}^{35{,}040} h(t)\,, \quad \text{where } h(t) = \begin{cases} (s_t - \tilde{s}_t)^2 & s_t > \tilde{s}_t \\ 0 & \text{otherwise}\,, \end{cases} \tag{6.14}$$

and

$$\Delta_-(\boldsymbol{s}) = \sum_{t=1}^{35{,}040} h(t)\,, \quad \text{where } h(t) = \begin{cases} (s_t - \tilde{s}_t)^2 & s_t < \tilde{s}_t \\ 0 & \text{otherwise}\,. \end{cases} \tag{6.15}$$

In order to obtain two categories on the basis of the median deviation metrics, their quantiles were used. The corresponding parameter of the algorithm is termed $q_{success}$, because it determines where to draw the line between profiles considered as spiky (success) and other profiles (failure). For each bin and separately for positive and negative overloads two sets were formed: A profile $\boldsymbol{s}$ was assigned to the 'success set' of spiky profiles for positive overloads whenever $\Delta_+(\boldsymbol{s}) \geq Q_{\Delta_+}(p = q_{success})$, where $Q_{\Delta_+}(p)$ denotes the quantile function of the median deviation metric which returns the estimated sample $p$-quantile for the discrete set of $\Delta_+$ values. For a given bin $b$, the two sets formed in this way are denoted by $\mathcal{S}_{b,success}$ and $\mathcal{S}_{b,failure}$. Analogously, the decision rule for negative overloads was to assign a profile $\boldsymbol{s}$ to the 'success set' whenever $\Delta_-(\boldsymbol{s}) \geq Q_{\Delta_-}(p = q_{success})$.

The Bernoulli distribution describes experiments with two possible outcomes, success or failure, which occur with probability $u$ and $1-u$, respectively. Using the Bernoulli distribution, the random assignment of profiles from a bin to small, unmonitored customers in the demand model can be parametrised in two steps: Taken the example of a single customer $i$, in the first step a Bernoulli trial is carried out with the success probability $u_i$ of the customer in question. The outcome of the Bernoulli trail is either success or failure. In the case of a success, in the second step a profile will be randomly chosen with uniform probability from the success set of profiles $\mathcal{S}_{b_i,success}$ of the bin $b_i$ which the customer belongs to. In case of a failure, a profile from $\mathcal{S}_{b_i,failure}$ will be randomly sampled with uniform probability.

In the demand model, as described in subsection 5.1.3, the random assignment of profiles with uniform probability in one step is the usual procedure. The described two step procedure will now lead to the same outcomes in terms of the resulting random profile selections $\boldsymbol{\Pi}$, under the condition that the success probability $u_i$ of customer $i$ corresponds exactly to the ratio

$$u_i = \frac{|\mathcal{S}_{b_i,success}|}{|\mathcal{S}_{b_i,success}| + |\mathcal{S}_{b_i,failure}|}\,, \tag{6.16}$$

where $|\cdot|$ denotes the number of elements in the sets, respectively. An example can illustrate why this is the case: Assuming a bin which contains 100 profiles, the probability of each profile of being chosen would be 0.01. For a quantile threshold $q_{success} = 0.9$, the success set would contain 10 profiles. The probability of a profile from the success set of being chosen in the two step procedure is then $0.1 \cdot 0.1 = 0.01$, the probability of a profile from the failure set of being chosen $0.9 \cdot 1/90 = 0.01$. Thus, for all profiles in the bin, the probabilities of being chosen remain unchanged.

The overall sampling distribution $f$ of a modelled asset, from which random profile selections $\boldsymbol{\Pi}$ are drawn, can therefore be parametrised as a chain of $n_s$ Bernoulli trials

$$f(\boldsymbol{x};\boldsymbol{u}) = \prod_{i=1}^{n_s} (1-u_i)^{x_i} \cdot (u_i)^{1-x_i}\,, \tag{6.17}$$

where $n_s$ is the number of unmonitored customer modelled with smart meter profiles whose success probabilities $u_i$ are determined according to equation (6.16). The outcome of each Bernoulli trail is collected in the vector $\boldsymbol{x} = (x_1, \ldots, x_i, \ldots, x_{n_s})$ with $x_i \in \{0, 1\}$. Given the vector of success probabilities $\boldsymbol{u} = (u_1, \ldots, u_i, \ldots, u_{n_s})$ with $u_i \in [0, 1]$, the value of the sampling distribution $f$ for a particular success-failure sequence can then be determined by evaluating equation (6.17).

### 6.4.2 Application of Importance Sampling and the Cross-Entropy Method

To evaluate the demand model in the described parametrised form, firstly, a random success-failure sequence $\boldsymbol{X}$ is sampled according to the success probabilities collected in vector $\boldsymbol{u}$. Secondly, per customer $i$ a profile is sampled either from $\mathcal{S}_{b_i,success}$ or form $\mathcal{S}_{b_i,failure}$, depending on whether $X_i$ was determined to be a success or a failure in the first step. Framing the demand model in this way allows to perform importance sampling by using a different vector of success probabilities $\boldsymbol{v}$, where the probabilities of sampling spiky profiles from the success set can be changed with respect to the original probabilities $\boldsymbol{u}$. For example, customer $i$ will be assigned one of the most spiky profiles with higher probability if $v_i > u_i$.

To carry out the importance sampling, firstly, again a random success-failure sequence $\boldsymbol{X}$ is sampled, now however according to the success probabilities collected in vector $\boldsymbol{v}$. The second step remains the same and profiles are still uniformly sampled from the success or failures set. The difference to the evaluating the demand model without IS consists, therefore, in sampling $\boldsymbol{X}$ according to the changed probabilities $\boldsymbol{v}$ in the first step, such that for certain customers more successes occur resulting in the profiles from the success sets of various bins to appear more frequently in the aggregate demand traces. To correct this purposefully introduced bias,

importance weights of the following form are used:

$$W(\boldsymbol{x};\boldsymbol{u},\boldsymbol{v}) = \frac{f(\boldsymbol{x};\boldsymbol{u})}{g(\boldsymbol{x};\boldsymbol{v})} = \frac{\prod_{i=1}^{n_s}(1-u_i)^{x_i}\cdot(u_i)^{1-x_i}}{\prod_{i=1}^{n_s}(1-v_i)^{x_i}\cdot(v_i)^{1-x_i}}\,, \tag{6.18}$$

where $g(\boldsymbol{x};\boldsymbol{v})$ is the IS distribution. Let $T_1, \ldots, T_n$ be an i.i.d. sample of time steps and $X_1, \ldots, X_n$ and i.i.d. sample of success-failure sequences from $g$. The IS estimator for this specific case follows then directly from the general form in equation (4.11):

Let $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ be an i.i.d. sample of profile selections. In the conventional Monte Carlo method a new random sample of time steps $\boldsymbol{\Theta} = \{\Theta_1, \ldots, \Theta_m\}$ is drawn per trace. Use H equation above.

$$\hat{R}_{IS-MC} = \frac{1}{n}\sum_{j=1}^{n} H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j)\cdot W(\boldsymbol{X}_j;\boldsymbol{u},\boldsymbol{v})\,, \tag{6.19}$$

where $n$ is the number of independent samples and $H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j)$ measures the impact of the random profile selection $\boldsymbol{\Pi}(\boldsymbol{X}_j)$ for a random sample of time steps $\Theta_j$ of the $j$-th trace. A subtle point to note here is that a specific success-failure sequence $\boldsymbol{x}$ is not *deterministically* linked to the resulting specific profile selection, because of the second step in evaluating the demand model which consists in uniformly sampling profiles from the success or failure sets based on the values of $\boldsymbol{x}$. Therefore, $\boldsymbol{\Pi}(\boldsymbol{x})$ is still a random vector. This point does not have any further implications for the following, but is useful to be aware of for the sake of clarity.

It is not clear a priori, what a suitable choice for the parameter vector $\boldsymbol{v}$ of $g(\boldsymbol{x};\boldsymbol{v})$ for a given asset could be. To automatically find a parameter vector $\boldsymbol{v}$ such that the 'distance'[11] to the theoretically optimal, unknown IS distribution $g^*$ is minimised, the CE method can be employed. If the distribution whose parameters are being CE optimised belongs to an exponential family, an analytical solution to the optimisation problem can be obtained [79]. This is the case here, since the Bernoulli distribution belongs to an exponential family. Furthermore, as detailed in the general description of the CE method in subsection 4.1.3, the solution to the optimisation problem can be estimated using a sampling-based formula. The method combines therefore optimisation and estimation.

Before deriving the sampling-based formula for the problem at hand, it is useful to recapitulate what was discussed at the end of subsection 4.1.3: When using the CE method in the context of rare events, finding the near-optimal IS distribution is often a rare event estimation problem itself. This is due to the fact that the optimal IS distribution $g^*$ is usually very different from the original sampling distribution in these cases. Using conventional MC to estimate the solution of the parameter optimisation problem would then require large sample sizes and defeat the purpose of why the CE method is being used in the first place. This difficulty can be circumvented by using a sequence of intermediate target regions which cover a gradually shrinking portion of the sample space. These intermediate regions allow estimating IS distributions which gradually approach the near-optimal IS distribution.

In each iteration of this sequential procedure, the parameter vector $\boldsymbol{v}$ of the biasing distribution $g(\boldsymbol{x};\boldsymbol{v})$ is newly estimated. The formula used for this purpose can be obtained by solving equation (4.24) which was derived in subsection 4.1.3. The equation is restated in the

**11** To be precise, the Kullback-Leibler divergence is minimised which is not distance metric in the rigorous mathematical sense.

following in the appropriate form for the problem at hand and such that it incorporates the idea of sequential updating:

$$\frac{1}{n}\sum_{j=1}^{n} H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j) \cdot W(\boldsymbol{X}_j; \boldsymbol{u}, \hat{\boldsymbol{v}}_{k-1}) \cdot \nabla_{\hat{\boldsymbol{v}}_k} \, ln\big(g(\boldsymbol{X}_j; \hat{\boldsymbol{v}}_k)\big) = 0\,, \tag{6.20}$$

where $n$ is the number of i.i.d. samples, $k$ is the current iteration of the sequential procedure, $\hat{\boldsymbol{v}}_{k-1}$ the estimated parameter vector of the previous iteration and $\nabla_{\hat{\boldsymbol{v}}_k}$ denotes the gradient with respect to parameter vector $\hat{\boldsymbol{v}}_k$ of the current iteration to be estimated. By taking the gradient $\nabla_{\hat{\boldsymbol{v}}_k}$ and resolving the system of equations for $\hat{\boldsymbol{v}}_k$, the sampling-based updating formula for the success probability $\hat{v}_{k,i}$ of customer $i$ can be obtained:

$$\hat{v}_{k,i} = \frac{\sum_{j=1}^{n} H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j) \cdot W(\boldsymbol{X}_j; \boldsymbol{u}, \hat{\boldsymbol{v}}_{k-1}) \cdot X_{j,i}}{\sum_{j=1}^{n} H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j) \cdot W(\boldsymbol{X}_j; \boldsymbol{u}, \hat{\boldsymbol{v}}_{k-1})}\,. \tag{6.21}$$

### 6.4.3 Description of the Cross-Entropy Algorithm

Based on the idea of sequential updating of the parameters $\boldsymbol{v}$ of the biasing distribution $g(\boldsymbol{x}; \boldsymbol{v})$ exposed above, the steps of the CE algorithm are as follows:

1. Set $\hat{\boldsymbol{v}}_0 = \boldsymbol{u}$ to start with conventional MC sampling. Set the iteration counter k = 1. Set the initial optimisation threshold $d_{optim} = 0.5 \cdot d_{cap}$.

2. Generate a random sample $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_{n_{optim}}$ that follows $g(\boldsymbol{X_i}; \hat{\boldsymbol{v}}_{k-1})$. Also generate a random sample of time steps $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{n_{optim}}$ for each of the $n_{optim}$ traces of the CE optimisation iteration. Obtain the maximum for each trace and store it in the vector $\boldsymbol{D}_{max,optim} = max(\boldsymbol{D}_{\boldsymbol{\Theta}_j}(\boldsymbol{\Pi}(\boldsymbol{X}_j)))$.

3. Set the optimisation capacity threshold $d_{seq} = Q_{\boldsymbol{D}_{max,optim}}(1-\rho)$.

4. Evaluate equation (6.21) with $d_{optim}$ as the capacity threshold to obtain the new updated success probability $\hat{v}'_{k,i}$ for each customer $i$. Hereby, use smoothing parameter $\alpha \in [0,1]$ for the updating such that $\hat{v}_{k,i} = \alpha \cdot \hat{v}'_{k,i} + (1-\alpha)\, \hat{v}_{k-1,i}$ and bound all success probabilities between $1 - q_{success}$ and 0.9.

5. Set counter k = k + 1. If $d_{seq} > d_{cap}$ set k = K (final iteration) and proceed with step 6, otherwise go back to step 2.

6. Generate a random sample $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ that follows $g(\boldsymbol{X_i}; \hat{\boldsymbol{v}}_K)$. Also generate a random sample of time steps $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_n$. Evaluate for all production traces $n$ computed so far

$$\hat{R}_{IS-MC} = \frac{1}{n}\sum_{j=1}^{n} H(\boldsymbol{\Pi}(\boldsymbol{X}_j), \boldsymbol{\Theta}_j) \cdot W(\boldsymbol{X}_j; \boldsymbol{u}, \hat{\boldsymbol{v}}_K)\,. \tag{6.22}$$

7. Verify the convergence in blocks of 50 traces, calculate therefore the relative error $RE_{\hat{R}_{IS-MC}}$ of the estimate. If $RE_{\hat{R}_{IS-MC}} < 0.1$, terminate the algorithm. Otherwise go back to step 6.

## 6.5 Combining Time Step and Profile Selection Importance Sampling

To combine time step and profile selection importance sampling, the optimisation stage of the profile selection IS algorithm is used. During the CE optimisation stage, also overload moments are stored, which allow to carry out a kernel density estimation after the CE optimisation stage is terminated. In the production stage, IS of time steps and profiles is carried out simultaneously according to the relations given above, when the methods were discussed separately.

## 6.6 Generalisation of Profile Selection Customer Weights to Bin Weights

In order to derive a general IS distribution for a wider range of assets from asset specific distributions, the results of evaluating the profile selection IS algorithm IS-MC on 150 MV/LV substations were used (see chapter 7 for further details). More specifically, the obtained customer success probabilities were given a look mapped to the 48 bins of the demand model which the customers are part of. Figure 6.4 allows to evaluate whether certain bins show distinctly higher success probabilities. While much noise is visible in the figure, it also appears that in some bins high success probabilities occur more often than on average.

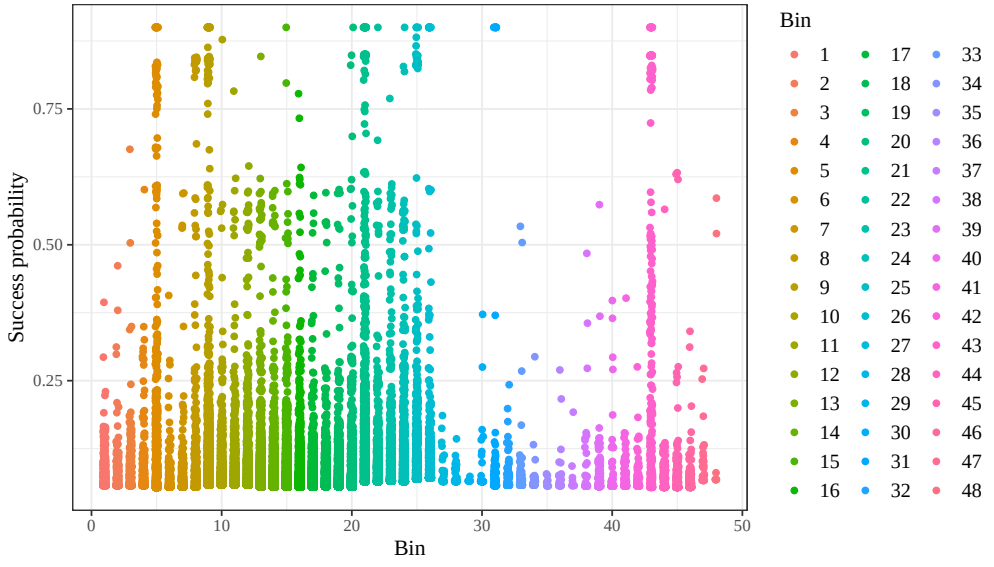


**Figure 6.4:** Customer success probabilities obtained from testing profile importance sampling on 150 MV/LV substations.

Taking this promising initial finding as a starting point, all assets with more than 80 customers were filtered out. It was observed then that the tendency of certain bins showing

higher probabilities became even clearer. Subsequently, the mean of all success probabilities was computed to obtain bin success probabilities. Finally, a thresholding was performed such that only bins with average success probabilities higher than 0.15 were assigned these higher probabilities. The initial probabilities were assigned to the remaining bins. The bin probabilities, which translate to bin weights when carrying out IS, obtained with the described procedure are employed directly in the IS-bw method. Therefore, when initialising the method for an asset, it is checked which bins the customers belong to. Then, they are assigned the respective bin success probabilities and importance sampling is performed from the first trace onwards.

# 7 Sampling Performance and Demand Model Evaluation Results

This chapter presents the results of evaluating the performance and suitability of the Monte Carlo sampling and extreme value inference methods described in the previous chapter. These methods were used to obtain the risk metrics of interest, which quantify certain properties of the demand model described in chapter 5. The accuracy as well as the efficiency of the investigated Monte Carlo (MC) sampling methods in obtaining risk metrics 1 and 2 – which are both probabilities of asset demand exceeding high thresholds – are assessed and discussed in section 7.1. The results of applying Extreme Value Theory (EVT) to infer distributions of annual maxima and minima from the model outputs and obtain risk metric 3 are presented in section 7.2.

Two kinds of grid assets were used to evaluate the various methods: the MV/LV substations, *Middenspanningsruimtes* (MSRs), and HV/MC substations, *Onderstations* (OSs), selected in subsection 5.2.1. Hereby, the major focus lied on the former because the Advanced Net DEcision Support (ANDES) model is presumed to underestimate peak demands especially on the lower voltage levels were demand variability is larger. Some results were collected on the full set of the 150 selected MSRs, comprising 100 MSRs with average and 50 MSRs with extreme characteristics in terms of the number of customers connected and their total yearly consumption. However, due to limited computational resources, other results were not collected on the full set of MSRs and similarly for OSs only a small subset was considered. A detailed breakdown of the number of assets and methods used to obtain the different results presented and discussed in this chapter is given in table 7.1.

## 7.1 Accuracy and Efficiency of the Investigated Monte Carlo Sampling Methods – Risk Metrics 1 and 2

Before diving into the details of the parameter tuning as well as the accuracy and efficiency results of the investigated Monte Carlo methods in this section, two pre-considerations may be helpful: Firstly, a brief recapitulation of the precise definitions of the first two risk metrics and, secondly, an illustration of how their magnitude may be interpreted. Based on the latter considerations, a relevant range of the risk metric values can be identified which is a useful reference frame for discussing the merits and drawbacks of the investigated methods in the context of distribution capacity planning.

All risk metrics distinguish the risk of positive and negative overload in separate sub-metrics. The definitions of first two risk are as follows: **Risk metric 1** is the probability of stochastic power demand $D_T(\boldsymbol{\Pi})$, for a random profile selection $\boldsymbol{\Pi}$ and at a random time step $T$, to exceed a critical value $d_{crit}$ which is defined as the power rating of an asset $d_{cap}$ plus a percentage margin. For the positive sub-metric this can be formulated in mathematical terms as the expectation

$$r_{1,+} = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d_{crit}}] \,, \tag{7.1}$$

**Table 7.1:** Overview of the investigated sets of assets and the methods used to obtain the results presented in the various sections of this chapter. Note that in the table legend the abbreviations used for the Monte Carlo methods are explained.

| Section | Content | Assets investigated | Methods investigated |
|---|---|---|---|
| Subsection 7.1.1 | Parameter tuning | 10 MSRs | MC-full, MC-MC, MC-IS, IS-MC, IS-IS |
| Subsection 7.1.2 | Accuracy of the MC sampling methods | 150 MSRs | MC-full, MC-MC, MC-IS, IS-MC, IS-IS |
| | | 117 MSRs | IS-bw |
| Subsection 7.1.3 | Efficiency of the MC sampling methods | 150 MSRs | MC-full, MC-MC, MC-IS, IS-MC, IS-IS |
| | | 117 MSRs | IS-bw |
| Subsection 7.1.4 | Results of the MC sampling methods for substations | 4 OSs | MC-full, MC-MC, MC-IS, IS-MC, IS-IS |
| Section 7.2 | Inference of extreme demand values | 150 MSRs | Block maxima method from extreme value theory |

The abbreviations used throughout this chapter for the various investigated Monte Carlo methods are explained below. The first position in the abbreviation always refers to the method for assigning demand profiles to small unmonitored customers, while the second position refers to the selection of time steps (except in the case of IS-bw):

- **MC-full:** Random sampling of profiles to obtain entire annual traces.
- **MC-MC:** Random sampling of profiles and time steps.
- **MC-IS:** Random sampling of profiles and importance sampling of time steps.
- **IS-MC:** Importance sampling of profiles and random sampling of time steps.
- **IS-IS:** Importance sampling of profiles and importance sampling of time steps.
- **IS-bw:** Importance sampling of profiles based on bin weights (bw) and random sampling of time steps. More consistent would be the abbreviation IS-MC-bw, but for brevity IS-bw is used.

while the negative sub-metric is defined analogously for the demand undercutting the critical power value, $D_T(\boldsymbol{\Pi}) < -d_{crit}$. **Risk metric 2** is the probability of the stochastic power demand $D_T(\boldsymbol{\Pi})$ to exceed the power rating of an asset $d_{cap}$. Therefore, its positive sub-metric can be specified as

$$r_{2,+} = \mathbb{E}[\mathbb{1}_{D_T(\boldsymbol{\Pi}) > d_{cap}}] , \tag{7.2}$$

and the negative sub-metric in analogous manner for the demand undercutting the power rating, $D_T(\boldsymbol{\Pi}) < -d_{cap}$. Therefore, risk metric 2 is the more fundamental quantity as it directly relates to the physical rating of an asset. At the same time, it is also the more conservative measure which is exceeded more easily. For these reasons, the positive and negative variants of metric 2 were respectively used as the target quantities in the algorithms which sequentially optimise the Importance Sampling (IS) distribution (MC-IS, IS-MC, IS-IS). Due to the structural

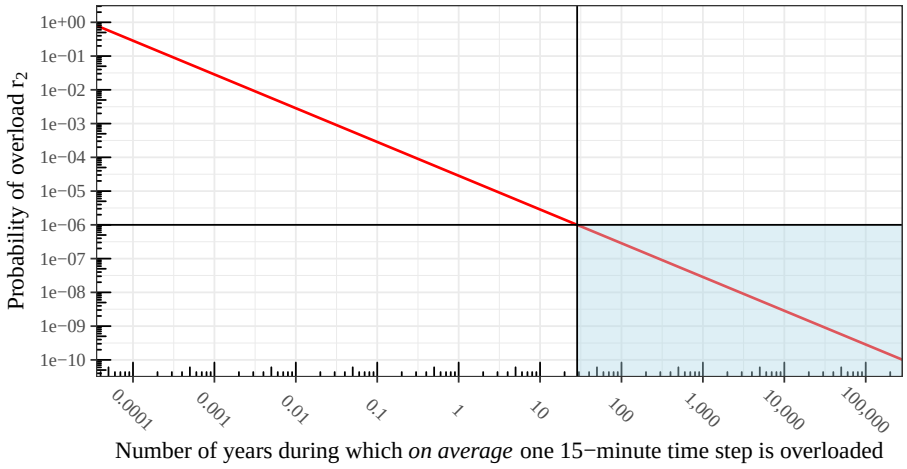


**Figure 7.1:** Visualisation of how per 15-minute overload probability magnitudes relate to the number of years during which one time step is overloaded on average. Highlighting the average character of metrics $r_1$ and $r_2$ is important, as they do not convey any information on how overloads are distributed over different years. The area shaded in blue is the range of probabilities not relevant for distribution system planning.

similarity to metric 1 this does not mean any loss in generality. Metric 1 was evaluated simultaneously to metric 2 and is often reported as well in this chapter.

Due to the specification of the above risk metrics to detect 'snapshot' samples which exceed a threshold, their output values are per 15-minute *average* overload probabilities. The average character is emphasised here, because the metrics do not convey any information on how regularly or irregularly the overloads occur (e.g. one overloaded time step per year would result in the same average probability as ten overloaded time steps every ten years). To make the per 15-minute probabilities more tangible, their magnitude is related to the number of years during which *on average* one quarter-hourly time step is overloaded in figure 7.1. From the visualisation in the figure, it becomes clear that probabilities of the order $10^{-6}$ and smaller are not very relevant for distribution capacity planning, as one overloaded time step on average in 28.5 years (which is the rounded value corresponding to $r_2 = 10^{-6}$) or more is tolerable on the distribution level. The range of years and probabilities which are considered as not relevant following this reasoning are shaded in blue in the figure. These considerations provide the relevant contextual background for interpreting the results of the Monte Carlo methods in this section.

### 7.1.1 Parameter Tuning

The fundamental idea of the time step and profile selection importance sampling algorithms described in sections 6.3–6.5 of the previous chapter, is to automate the process of identifying a suitable biasing distribution for any given asset. Despite this intention, certain parameters of the algorithms still have to be set manually which were also introduced in the previous chapter.

**Table 7.2:** Overview and description of all relevant parameters of the developed sampling algorithms which rely on Importance Sampling (IS) and the Cross-Entropy (CE) method.

| | Parameter | Description | Default value | Tuned |
|---|---|---|---|---|
| Common parameters of all methods | $m$ | Number of time steps sampled per trace | 5000 | yes |
| | $n_{prod}$ | Number of traces in each block for which convergence is tracked in the production stage | 50 | no |
| | $n_{max}$ | Maximum number of traces which is computed | 20,000 | no |
| | $n_{max,zero}$ | Number of traces after which the simulation is stopped if all estimates are zero | 10,000 | no |
| Parameters of time step IS | $n_{seq}$ | Number of traces used during the sequential updatingof the IS distribution stage | 200 | no |
| | $m_{kde}$ | Minimum number of time steps to be used for the kernel density estimation | 250 | yes |
| | $\lambda$ | 'Ease' of raising the threshold in the sequential stage | 0.1 | yes |
| | $g_{floor}$ | Percentage of the maximum value of the IS distribution which is set as the floor value | 0.03 | yes |
| Parameters of profile selection IS | $n_{optim}$ | Number of traces used during the CE optimisation of the IS distribution | 500 | yes |
| | $\rho$ | 'Ease' of raising the threshold in the CE optimisation stage | 0.05 | yes |
| | $\alpha$ | Amount of smoothing when updating the IS distribution | 0.6 | yes |
| | $q_{success}$ | Quantile to separate spiky ('success') profiles from less-spiky profiles | 0.9 | yes |

As it is not clear beforehand which values are optimal, simulation experiments with varying parameter values have been conducted to observe the effect of parameter changes. An overview and a description of all relevant parameters of the developed sampling algorithms is provided in table 7.2. Not all of the parameters listed in the table have been tuned: The common parameters of all algorithms $n_{prod}$, $n_{max}$ and $n_{max,zero}$ are concerned with the amount of samples taken and do not qualitatively influence the performance of the algorithms. Similarly, the parameter $n_{seq}$ is not influential in terms of how well the time step IS algorithm performs, because in case the minimum number of time steps above the current threshold $m_{kde}$ is not reached, another block of $n_{seq}$ traces is evaluated without discarding the time steps collected so far.

Tuning various parameters in combination can quickly become computationally very expansive, because the number of combinations increases exponentially with the number of parameters considered. To circumvent the issue, only the effects of varying one parameter at a time, while keeping all other parameters constant, were studied. To this end, sensible ranges for the parameters were identified, either based on the experience gained in trial runs or based on typical values suggested in the literature. The latter approach could be taken for the parameters $\rho$ and $\alpha$ of the Cross-Entropy (CE) method based on [26, 27]. The ranges were subsequently discretised to obtain a small number of parameter values for parameter tuning experiments. The discretisation was either done evenly, or using smaller step sizes in those parts of the parameter ranges where the sensitivity was presumed to be higher compared to other parts. The default values for each parameter shown in table 7.2 were then chosen such that they respectively lie in the middle of the investigated parameter ranges.

Furthermore, in order to limit the computational cost, the parameter tuning experiments were only carried out on a small number of assets, namely on a subset of ten MSRs out of the total set of 150 MSRs with average or extreme characteristics. Evidently, when using such a small subset it is desirable to make it as representative as possible. For accomplishing this, the

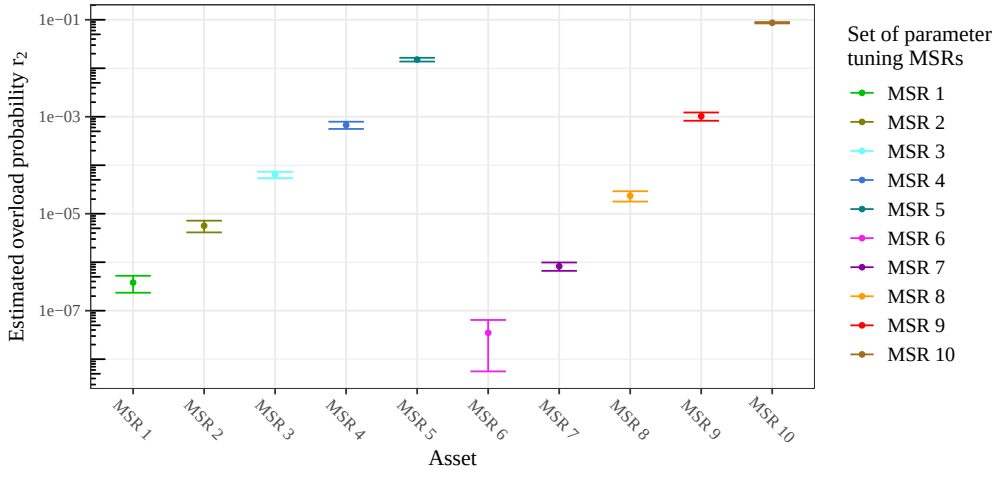


**Figure 7.2:** Selection of ten MSRs for the parameter tuning experiments which cover the overload probability range of all orders of magnitude. The five MSRs to the left are representative of the $r_{2,+}$ range, while the five MSRs to the right are representative of the $r_{2,-}$ range.

principle consideration was that the efficiency of Monte Carlo methods critically depends on the order of magnitude of the estimated quantity. For this reason it is plausible that the effects of parameter changes are similar for MSRs with similar overload probabilities $r_2$ – and vice versa. Therefore, choosing MSRs with overload probabilities of different orders of magnitude is likely to result in a more representative set than a randomly chosen subset. Furthermore, it was suspected that there may be differences in the effects of parameter changes when estimating positive versus negative overload probabilities ($r_{2,+}$ versus $r_{2,-}$).

Based on these considerations, the overload probability ranges of $r_{2,+}$ and $r_{2,-}$ observed in an initial trial run for the total set of 150 MSRs were respectively divided into 5 equally spaced bins on a logarithmic scale. Additionally, assets requiring more than 100 s to reach a relative error below 10% were filtered out. Subsequently, one MSR was randomly drawn from each bin resulting in five MSRs representative of the $r_{2,+}$ range and five MSRs representative of the $r_{2,-}$ range. The estimates with 95% confidence intervals of the ten parameter tuning MSRs are shown in figure 7.2. It can be gathered from the figure that the estimates span various orders of magnitude and are approximately equally spaced on a logarithmic scale, as desired. Furthermore, it is noteworthy that MSRs 1, 6 and 7 have overload probabilities below the relevant range. The parameter tuning results for these MSRs are shown in the following, however they are not taken into account when deciding on which parameter values are most appropriate.

Three sets of parameter tuning experiments were run: Firstly, a set of experiments to investigate the parameters $m_{kde}$, $\lambda$ and $g_{floor}$ of the time step IS algorithm. Secondly, a set of experiments to investigate the parameters $n_{optim}$, $\rho$, $\alpha$ and $q_{success}$ of the profile IS algorithm. It was not deemed necessary to investigate the parameters from the first and the second set of experiments again for the algorithm combining time and profile IS. Thirdly, a set of experiments was conducted to investigate the number of time steps $m$ sampled per annual trace for all methods. The parameter $m$ was given this special attention because of its potential to directly influence computation time.

Varying the parameters of the **time step IS** algorithm did not result in any noticeable overall trends in the number of traces necessary until reaching a relative error $\leq 10\%$. Due to the lack of trends and for brevity in this chapter, a detailed visualisation of the results is shown in figure A.1 of appendix A. These findings may be seen as an indication that the time step IS algorithms are quite robust, at least for the investigated assets and parameter ranges. Overall, therefore, no clear indication was found to change any of the parameters away from the default values shown in table 7.2.

The results of varying the parameters of the **profile selection IS** algorithm are shown in figure 7.3. Again, the number of traces until reaching the target relative error of 10% are considered. For interpreting the findings, it is important to note the general tendency that assets with lower overload probability require more traces for convergence. For example, the MSRs 1, 6 and 7 (in green, pink and purple in figure 7.3) with overload probabilities below the relevance threshold of $10^{-6}$ stand out quite clearly. In contrast to the results of the time step IS, here clearer trends are visible for two out of four parameters: For $n_{optim}$ there seems to be a trade-off in the sense that for assets with $r_2 > 10^{-6}$ lower values of $n_{optim}$ lead to a faster convergence, while for the assets with $r_2 < 10^{-6}$ higher values are favourable. This is sensible because a larger block of traces $n_{optim}$ per optimisation step is likely to lead to a more optimal IS distribution for rare event probabilities. Considering that the asset with $r_2 < 10^{-6}$

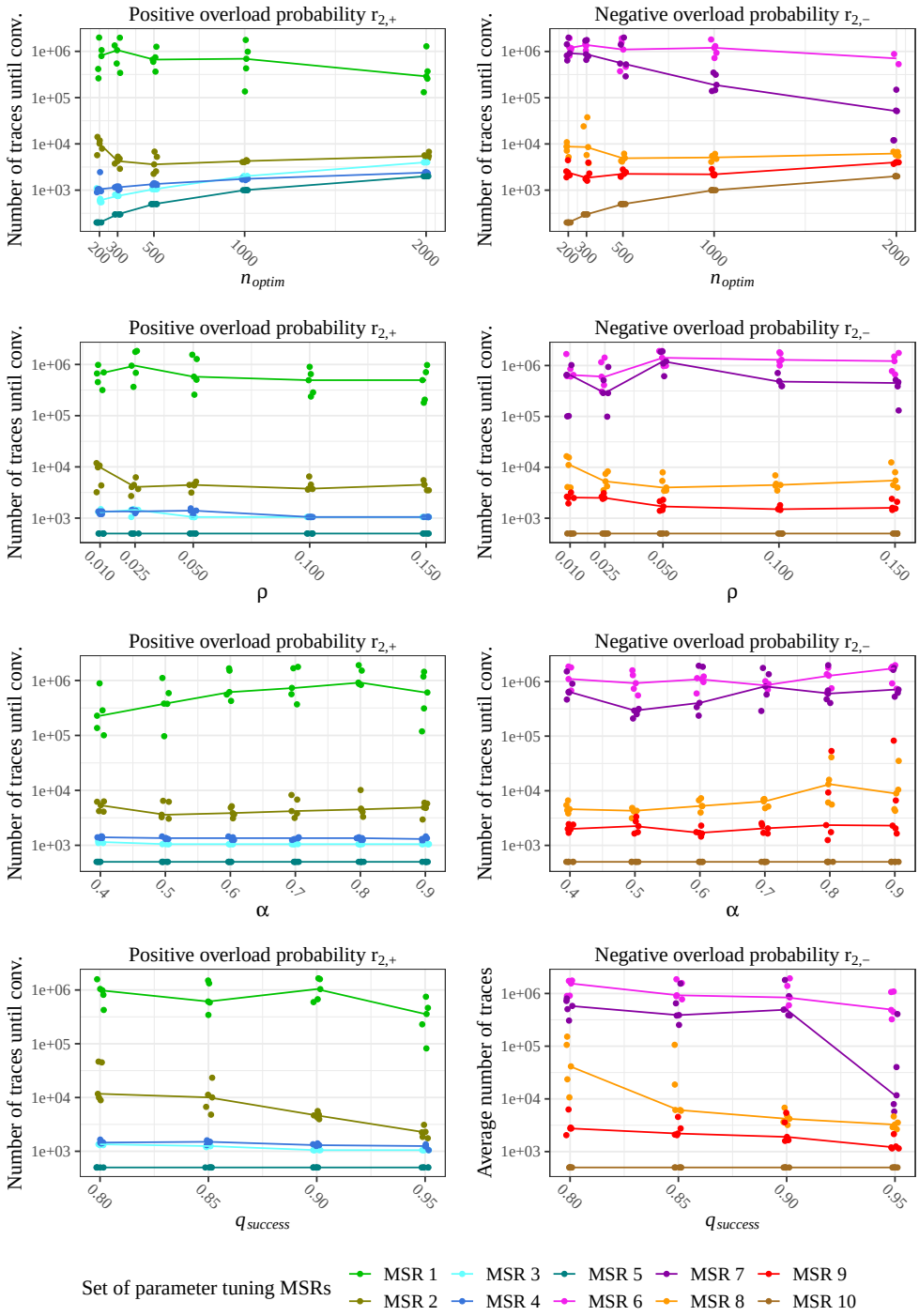


**Figure 7.3:** Parameter variation results for the profile selection IS algorithm. The number of traces needed to reach a relative error ≤ 10% is shown for each tested parameter value and replicate (points). A small amount of random noise has been added to the x-coordinate of the points to avoid overplotting. The medians of each set of five replicates are connected to highlight potential trends (lines).

are less relevant, the default value of $n_{optim} = 500$ is maintained since it constitutes a good compromise in the described trade-off. No clear trends are visible for the parameter $\rho$, apart from not setting it too low, therefore the default value of $\rho = 0.05$ is maintained as well. For the parameter $\alpha$ the situation is similar, except for the observation that values above 0.8 appear to result in a large spread for MSRs 8 and 9 which may indicate a less robust behaviour for high values. Thus, the default value of $\alpha = 0.6$ appears to be a reasonable choice already. For $q_{success}$, a relatively clear downward trend is visible for almost all MSRs. This observation can be well explained, because higher values of $q_{success}$ mean that the set of profiles given more weight in the IS shrinks and only the most spiky profiles remain – which is likely to lead to a better variance reduction. Therefore, a clear indication that it is reasonable to raise the value of $q_{success}$ from 0.9 to 0.95 was found in this case.

The effect of varying the **number of time steps *m*** sampled per annual trace was investigated for all methods (except for bin weight IS of profiles, IS-bw). Here, the computation time until reaching a relative error $\leq 10\%$ was deemed to be the better measure than the number of traces. The reason for not looking at the number of traces is that sampling more time steps per trace is likely to result in less traces needed for reaching the target relative error. However, this does not necessarily mean that also computation time is lower, because sampling more time steps per trace is computationally more expensive – thus there could be a trade-off between decreasing the number of traces needed for convergence and limiting the computation time per trace.

The results of the corresponding simulation experiments can be observed in figure 7.4. Large fluctuations are visible for MSRs 1, 6 and 7, especially when profile selection IS is involved which make it hard to identify clear trends. However, as these MSRs have overload probabilities below the relevant range, their behaviour does not necessarily need to be taken into account when choosing a suitable value for $m$. For the MSRs with relatively large overload probabilities and therefore rather small computation times, an increasing tendency in the computation times with the number of time steps sampled per trace is visible, as expected. On the other hand, MSR 2, with an overload probability at the lower end of $10^{-5}$, indicates that sampling only 500 or 1,000 time steps may lead to slightly higher computation times for assets with $r_2$ values of this order compared to sampling 2,000 time steps and above. The likely cause is that when sampling less than 2,000 time steps, some overload events will be missed requiring more traces to be computed until convergence occurs. Fortunately, for MSR 2 it does not appear that sampling more than 2,000 time steps decreases computation time further. Thus, a value of $m = 2{,}000$ time steps seems to be a reasonable compromise which allows assets with $r_2 > 10^{-5}$ to be computed faster than with the default value of $m = 5{,}000$.

To sum up, only for two cases indications where found that it is reasonable to change parameters away from their default values: Firstly, for $q_{success}$ a value of 0.95 requires less traces for convergence than the default of 0.9 and therefore $q_{success}$ is adjusted accordingly. Secondly, the number of time steps sampled per annual trace $m$ is decreased from 5,000 to 2,000. For all other parameters the default values are retained.

### 7.1.2 Accuracy of the Investigated Monte Carlo Methods

The accuracy of the developed MC sampling methods was assessed on the set of 150 MSRs, comprising of 100 MSRs with average and 50 MSRs with extreme characteristics, for all methods

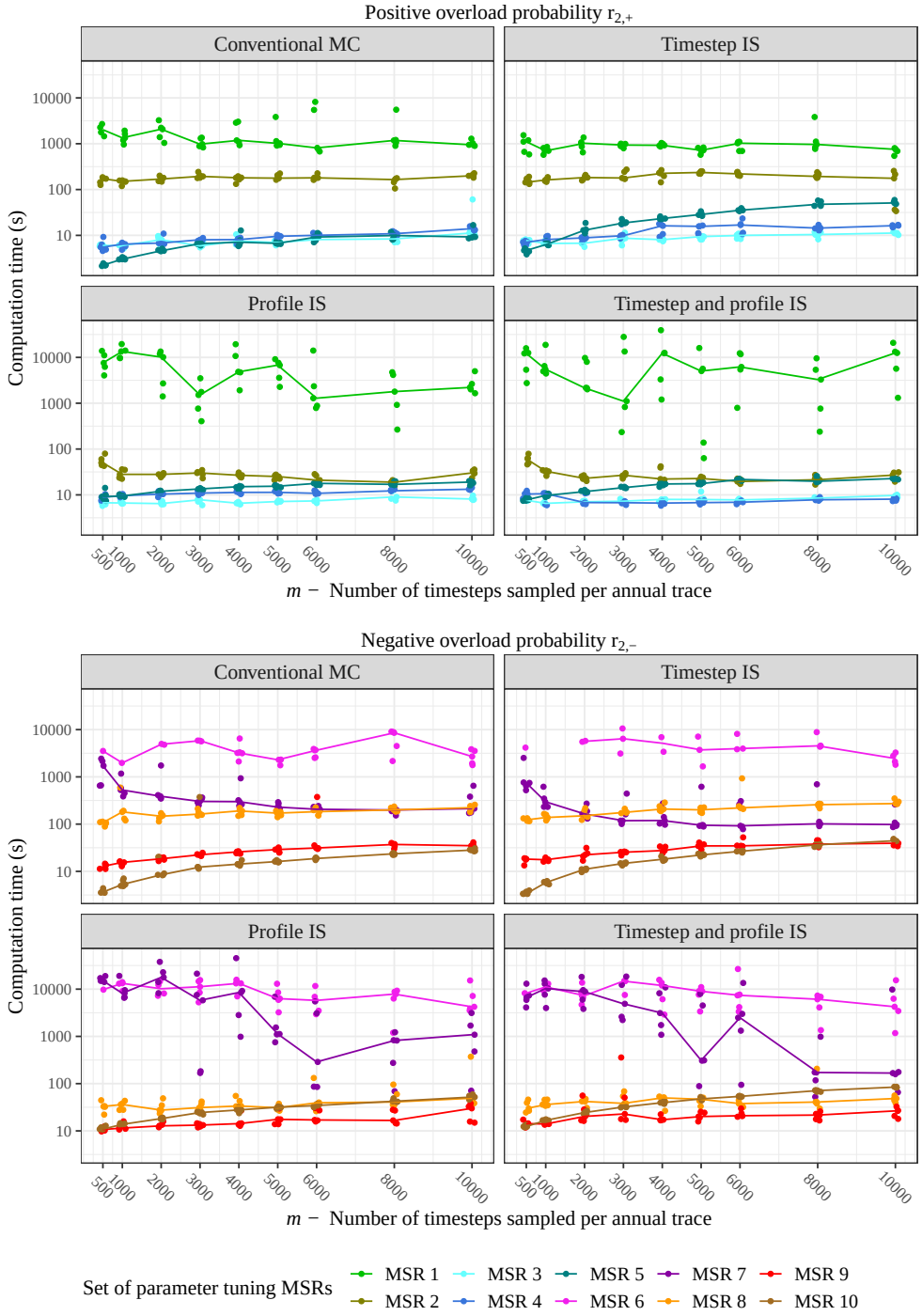


**Figure 7.4:** Results of varying the number of time steps $m$ sampled per annual trace. The computation time until reaching a relative error $\leq 10\%$ is shown for each tested parameter value and replicate (points). A small amount of random noise has been added to the x-coordinate of the points to avoid overplotting. The medians of each set of five replicates are connected to highlight potential trends (lines).

except for bin weight IS (IS-bw). The IS-bw method is a special case because it required the weights resulting from the profile selection importance sampling method (IS-MC) as an input for generalising customer specific weights to bin weights, as described in section 6.6. Therefore, it was run after the simulation experiments for all other methods and, due to computational limitations, it was chosen to test the method only on those 117 out of the 150 MSRs with nonzero values of metrics 1 and 2. To further limit the computational effort, only five replicates where produced for the IS-bw method instead of the nine replicates considered for all other methods.

Assessing the accuracy of the sampling methods is meant here in the sense of verifying whether the estimates produced by them are of a similar magnitude. Theoretically, the MC estimator and the IS estimator are both unbiased, as detailed in section 4.1. Thus, the theory predicts that the estimates produced by all investigated methods converge to the same population value for infinite sample sizes. However, when visually inspecting the nonzero estimates obtained for 117 out of the 150 MSRs, it became apparent that the estimates for the methods involving profile importance sampling (IS-MC, IS-IS and IS-bw) were not always in accordance with the estimates produced by the remaining methods (MC-full, MC-MC and MC-IS). The estimates of all methods and independent replicates obtained for two exemplary MSRs are shown in figure 7.5. While the estimates for MSRs 18 on the left are in good accordance for both variants of risk metrics 1 and 2, for MSRs 117 on the right the IS-MC and IS-IS estimates for $r_{1,-}$ and $r_{2,-}$ are orders of magnitude smaller than the other estimates. Interestingly, the bin weight method IS-bw does not appear to suffer from the issue in this case, even though it also involves profile selection importance sampling. However, it was noted that for other assets the issue arose as well with bin weight importance sampling.

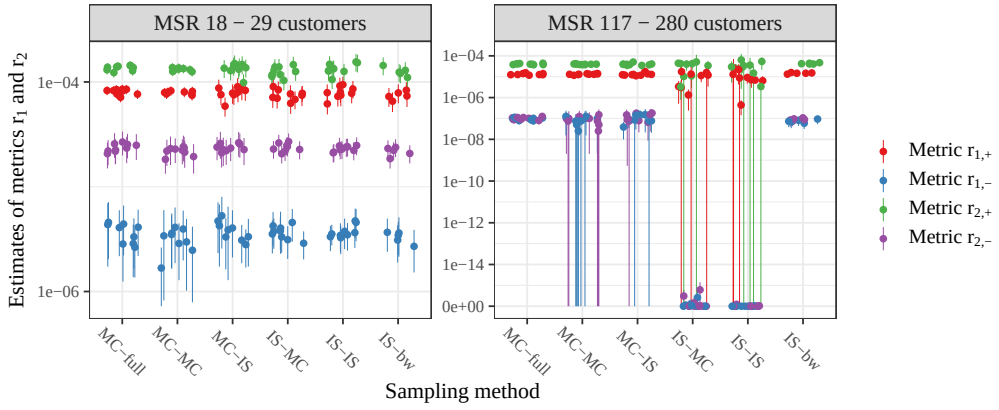


**Figure 7.5:** Comparison of the risk metric 1 and 2 estimates obtained with different sampling methods for two exemplary MSRs. Point estimates and 95% confidence intervals of all independent replicates produced are shown on a pseudo-logarithmic scale which becomes linear as zero is approached. Confidence intervals are clipped at zero, as probabilities cannot be negative.

As the results of Monte Carlo methods are inherently stochastic, visual comparisons between groups of estimates do not always allow to clearly conclude whether there is a bias or not. A more objective and theoretically sound approach for comparisons between groups of stochastic results are hypothesis tests for equal means. Since in this case the variances of

two groups of replicates, as well as their sizes may be different (the IS-bw method has fewer replicates), Welch's $t$-test is the appropriate choice.[12] The test was used to systematically evaluate how often the magnitude of estimates produced by different methods deviates by testing the hypothesis of equal means between groups of replicates. Random sampling of full annual traces (MC-full) was used as the reference method to which all other methods were compared, since it directly uses the demand model outputs without any further assumptions.

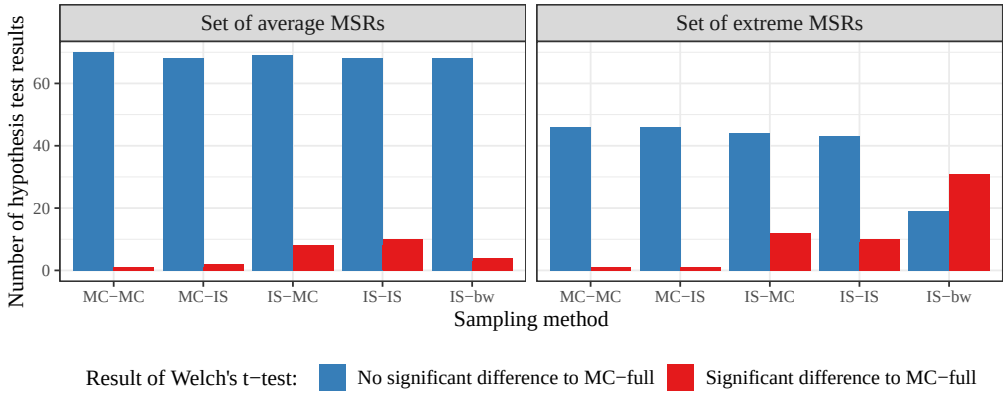


**Figure 7.6:** Visualisation of the hypothesis testing results with the significance level $\alpha = 5\%$ for the MSRs with average and the MSRs with extreme characteristics. Note that the number of assets in the set of extreme MSRs is lower compared to the set of average MSRs, therefore also the overall count of hypothesis tests results shown on the right is lower.

The results of testing for each asset whether the replicates of the MC-full method have a significantly different mean from the replicates of the five other methods are shown in figure 7.6. The common significance level of $\alpha = 5\%$ was adopted and the Holm method to adjust for multiple testing was used when carrying out the hypothesis tests with the *R* package *rstatix* [54]. Figure 7.6 shows that mainly estimates of the profile selection IS methods differ significantly from the MC-full estimates. Furthermore, much more estimates of these methods differ significantly from the MC-full estimates in the extreme set than in the average set.

Considering the high number of 280 customers of MSR 117 in figure 7.5 and that a customer number above average is common for MSRs in the extreme set, inaccurate estimates might occur more often for assets with a high number of customers. To verify whether this is the case, the following ratio was calculated for the risk metrics $r_{2,+}$ and $r_{2,-}$ of each asset with nonzero estimates:

$$log_{10}\left(\frac{r_{method}}{\bar{r}_{MC-full}}\right), \tag{7.3}$$

where $\bar{r}_{MC-full}$ is the mean of all MC-full replicates of a given asset and risk metric and $r_{method}$ stands for individual replicates of the estimates obtained with the remaining methods. This

12 Welch's $t$-test assumes normally distributed samples. Due to the central limit theorem, the estimates produced by the MC and the IS estimator follow a normal distribution for sufficiently large sample size, determined in this case by the number of traces computed. As this number is at least 200 and frequently more than 1,000, the assumption of normality appears reasonable to make in this case.

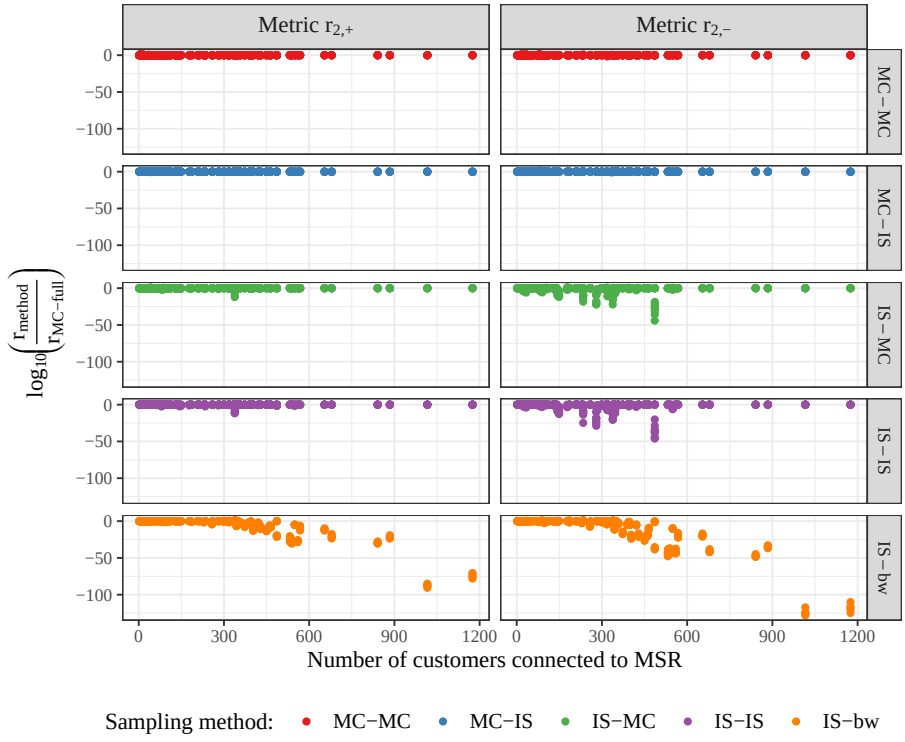


**Figure 7.7:** Order of magnitude comparison of the estimates with respect to the reference method MC-full, plotted against the number of MSR customers. Negative values indicate how many orders of magnitudes an estimate is smaller than the mean reference estimate.

ratio is plotted against the number of customers connected to a given MSR for all sampling methods in figure 7.7. It becomes clear that the issue arises only for the profile importance sampling methods IS-MC, IS-IS and IS-bw. Furthermore, it can be gathered from the figure that the issue is much more pronounced for negative overloads. A clear correlation of the estimate deviations with the number of customers is visible for the IS-bw method above 250-300 customers. Below 250 customers, estimates remaining fairly accurate, while deviations from the MC-full estimate increase up to more than 100 orders of magnitude for the two MSRs with the most customers. It is noteworthy, that the low accuracy was not observed for these MSRs with the IS-MC and the IS-IS methods. Manual inspection of the cases has revealed that $r_{2,+}$ and $r_{2,-}$ are both greater 0.01 and therefore very large in comparison to other overload probabilities. This leads to the IS-MC and IS-IS methods converging in the first iteration of the optimisation stage, before any importance sampling occurs. The IS-bw method, in contrast, uses bin weights to perform importance sampling from the first trace onwards. Overall, these findings indicate that estimates can become very inaccurate when using profile selection IS for assets with a large number of customers for which overloads are rather rare events.

The likely explanation for why estimates become inaccurate has to do with the way the importance weights are determined in the profile selection IS algorithms developed in the

previous chapter. The challenging aspect hereby is that whenever more than a moderately large number, i.e. more than 15-20, customers are assigned high probabilities of picking spiky profiles, the resulting importance weights are frequently very small. To illustrate this, MSR 117 from figure 7.5 is used as an example: A detailed look at the results of the first replicate of the IS-MC method for $r_{2,-}$ revealed that at the end of the CE optimisation stage, 25 out of the total 280 customers received a success probability of larger than 0.6 (the probability of picking a profile from the set of spiky profiles is called success probability). For the sake of this illustration, a situation is assumed where 25 customers have a success probability of 0.6 and the remaining 255 customer retain their initial success probabilities of 0.5.[13] Therefore, the situation can be described using binomial distributions, since success probabilities are the same for large blocks of customers. The importance weights can then be calculated by evaluating

$$W(x) = \frac{f(x)}{g(x)} = \frac{h(x; n = 25, p = 0.05) \cdot h(x; n = 255, p = 0.05)}{h(x; n = 25, p = 0.6) \cdot h(x; n = 255, p = 0.05)} , \tag{7.4}$$

where $f$ is the original sampling distribution, $g$ the biased sampling distribution and $h(x; n, p)$ denotes the probability mass function of the binomial distribution with the number of successes $x$, number of trails $n$ (here the number of customers) and success probability $p$. The original sampling distribution $f$ in the numerator is split in two binomial distributions to highlight that the terms for the 255 customers, whose success probabilities remain unchanged, cancel out. With a probability of approximately 8%, sampling a number of $x = 18$ successes from $g$ is not an unlikely event. If in this case an overload event occurs, the trace importance weight is

$$W(x) = \frac{f(x)}{g(x)} = \frac{h(x = 18; n = 25, p = 0.05)}{h(x = 18; n = 25, p = 0.6)} \approx \frac{1.3 \cdot 10^{-18}}{0.080} \approx 1.6 \cdot 10^{-17} , \tag{7.5}$$

and therefore very small. In itself, small importance weights are not an issue because they mean that certain regions in the sample space are targeted with high probability. If the states in these regions are the only states which lead to overload events, then variance reduction will work very well. A problem arises, however, if also other states in the sample space farther away from to the highly targeted regions lead to overload events. Continuing with the illustrative example, if with a small probability overload events can also occur for a number of successes $x = 4$, then the corresponding importance weight is

$$W(x) = \frac{f(x)}{g(x)} = \frac{h(x = 4; n = 25, p = 0.05)}{h(x = 4; n = 25, p = 0.6)} \approx \frac{0.027}{7.2 \cdot 10^{-6}} \approx 3.7 \cdot 10^{3} . \tag{7.6}$$

Observing such large importance weights is very rare because the corresponding states are only sampled with a probability of $7.2 \cdot 10^{-6}$ from $g$ and furthermore, overload events only occur infrequently when fewer spiky profiles are sampled.

Two conclusions can be drawn from this illustrative example: Firstly, even though estimates may be dragged down due to very low weights and appear strongly biased initially, in the limit of large sample sizes they will nevertheless converge to the correct quantity due to the infrequent occurrence of very large weights. Therefore, the inaccuracy of estimates observed

**13** The initial success probabilities are only approximately 0.5, as explained in the previous chapter, however for the sake of the illustration probabilities of exactly 0.5 are assumed.

for certain cases here is not in conflict with the theory which predicts unbiased importance sampling estimates. Secondly, the example shows that care should be taken to prevent success probabilities from becoming too extreme – because otherwise the initial estimates are of a wrong order of magnitude and the goal of variance reduction is missed, because for convergence eventually much larger sample sizes might be needed compared to conventional MC.

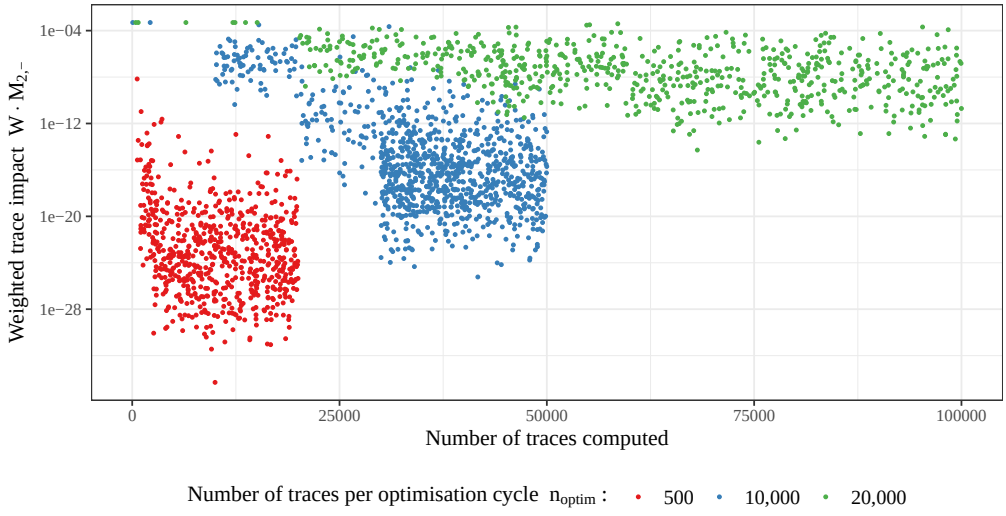


**Figure 7.8:** Comparison of the weighted trace impact values $W \cdot M_{2,-}$ of MSR 117 for various numbers of traces used per iteration in the optimisation stage of the IS-MC method. Only nonzero weighted trace impacts are shown, corresponding to traces for which overload events occurred.

The idea of deploying the Cross-Entropy (CE) method in this case, was to prevent the described scenario by setting targeted, but not too extreme importance weights. However, for accomplishing this, a sufficiently large amount of samples in the optimisation stage of the profile IS algorithms IS-MC and IS-IS is necessary for the CE method to work well. Due to this, the suspicion arose that $n_{optim} = 500$ may be too small for reliably estimating a larger number of parameters – in the case of MSR 117 with 280 customers, 280 parameters need to be estimated. For the example of MSR 117, it was investigated whether increasing the number of traces $n_{optim}$ used for each iteration in the optimisation stage alleviates the accuracy issue. The results of these trials are visualised in terms of the nonzero weighted trace impact values in figure 7.8. As shown in equation (4.11), the estimates are calculated by taking the average of all zero and nonzero trace impacts. Therefore, a high number of low trace impact values will drag down the overall estimate. It can be observed, that with an increasing number of traces used in the CE optimisation stage, the trace impact values are indeed of larger orders of magnitude. For $n_{optim} = 20,000$ the estimate was found to be still slightly to small, yet in the correct order of magnitude.

The problem for the IS-bw method is likely of a very similar nature: Due to the fact that customer success probabilities are determined based on their bin membership, for larger assets many customers may be part of a given bin. If many customers are part of bins with large bin success probabilities, a large number of high customer success probabilities is likely. For

example, the MSR shown in the bottom right of the right IS-bw panel in figure 7.7 has a total of 1,175 customers. A look at the results of one replicate of the IS-bw method for $r_{2,-}$ revealed that 495 customers were assigned success probabilities larger than 0.48. Therefore, it is clear what caused the deviation of more than 100 orders of magnitude in this case. Furthermore, the clearer correlation of the accuracy with the number of customers for the IS-bw method compared to the other profile selection IS methods (IS-MC and IS-IS) is also sensible in this light, because the number of customers in bins with high weights should on average increase proportionally with the overall number of customers.

### 7.1.3 Efficiency of the Investigated Monte Carlo Methods

Before comparing the efficiency of the investigated Monte Carlo methods in terms of computational speed, it is useful to consider the overview of the nonzero $r_2$ and converged estimates in figure 7.9. Note that the higher nonzero fractions of the IS-bw method are due to the fact that the method was only evaluated for the 117 MSRs which resulted in nonzero estimates for the other methods. It is notable that for approximately 2/3 of the 100 average MSRs nonzero overload probabilities were found, while the share is with approximately 80-90 % even higher in the set of MSRs with extreme characteristics. Furthermore, it can be observed that only a part of the estimates converged within the maximum number of 20,000 traces per simulation run, where convergence is defined as reaching a relative error $\leq 10\%$.

To benchmark the computational efficiency of the various methods, their computation

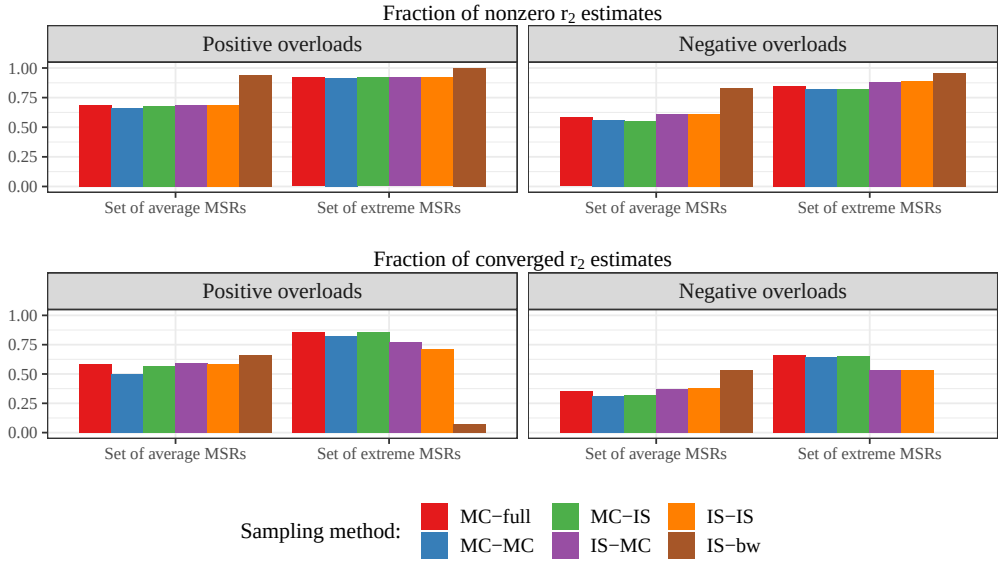


**Figure 7.9:** Overview of the nonzero $r_2$ estimates and the $r_2$ estimates which reached a relative error $\leq 10$ within a simulation run of 20,000 traces. Note that the higher nonzero fractions of the IS-bw method are due to the fact the method was only evaluated for the 117 MSRs which produced nonzero estimates for the other methods.

times were measured. On a shared access server, such as the one used for carrying out the simulations, time measurements are not always reliable because computation times may vary depending on the current load of the server. In order to evaluate how reliable the time measurements are, a generic calculation with a computation time of approximately one minute was implemented and timed before the simulation of each estimate. Of course, this only gives a snapshot impression of how busy the server currently is, but it can at least give a rough indication. It was found that the computation time of the generic calculation varied with a standard deviation of 48.7 seconds around a mean of 57.5 seconds. Due to the presence of a few extreme outliers, the standard deviation is relatively high compared to the mean. However, a look at the histogram in figure 7.10 reveals that the computation times varied in most cases approximately within a 15 second range. This is seen as an indication that at least time measurements averaged over several replicates can be seen as reasonably reliable.

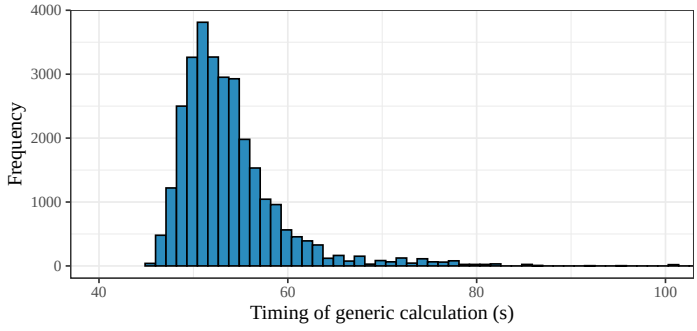


**Figure 7.10:** Histogram showing the timing results of the generic calculation to evaluate the current load of the shared access server. A small number of extreme outliers is omitted in the figure.

Another important issue which makes the direct comparison of computation times challenging is the fact that only a part of the estimates reached the pre-defined convergence target of a relative error $\leq 10\%$, as shown in figure 7.9. If it is assumed that the estimates which do not converge are on a steady convergence path close enough to the asymptotic limit when hitting the maximum number of 20,000 traces, an approximation can be made to estimate the total computation time necessary for convergence to a relative error $\leq 10\%$. Being close to the asymptotic limit is important because then the central limit theorem is valid and the relation for the relative error or coefficient of variation given in equation (4.8) holds. From the expression for the relative error, the following approximation for the computation time $t_{RE=0.1}$ necessary to reach a relative error of 10% can be obtained:

$$t_{RE=0.1} = \left(\frac{SE_{\hat{r}}}{RE \cdot \hat{r}}\right)^2 \cdot t_{n_{max}} = \left(\frac{SE_{\hat{r}}}{0.1 \cdot \hat{r}}\right)^2 \cdot t_{n_{max}} \tag{7.7}$$

where $\hat{r}$ denotes the estimate, $SE_{\hat{r}}$ its standard error and $t_{n_{max}}$ the actually measured computation time until reaching the maximum number of trace $n_{max} = 20{,}000$. Therefore, the relation rescales the measured computation time based on the desired target relative error and the estimate as well as its standard error, obtained when reaching $n_{max}$.

A comparison of the average speed-up of all methods with respect to the reference method

MC-full is shown in figure 7.11. To compute the average speed-up, the average computation time over all MC-full replicates $\bar{t}_{MC-full}$ was calculated per asset and metric. Then the ratio $\bar{t}_{MC-full}/t_{method}$ was formed individually for each replicate, metric and asset. Subsequently, the averages of the individual speed-up ratios were formed for each method. The methods were bench-marked only for those cases where they produce accurate estimates, because this indicates their full potential for possible specific use cases (e.g. assets with a number of customers up to 80). Therefore, all estimates whose accuracy was found to differ significantly from the estimates of the reference method MC-full (in the hypothesis tests of the previous subsection) were removed before calculating the average speed-ups. The average speed-ups without removing the inaccurate estimates are shown in figure A.2 of appendix A. Furthermore, also zero estimates have been excluded from the average speed-up computation, because their computation times depend on how stopping criteria are set (in this case $n_{zero,max}$ was set to 10,000) and do not allow conclusions on the estimation efficiency.

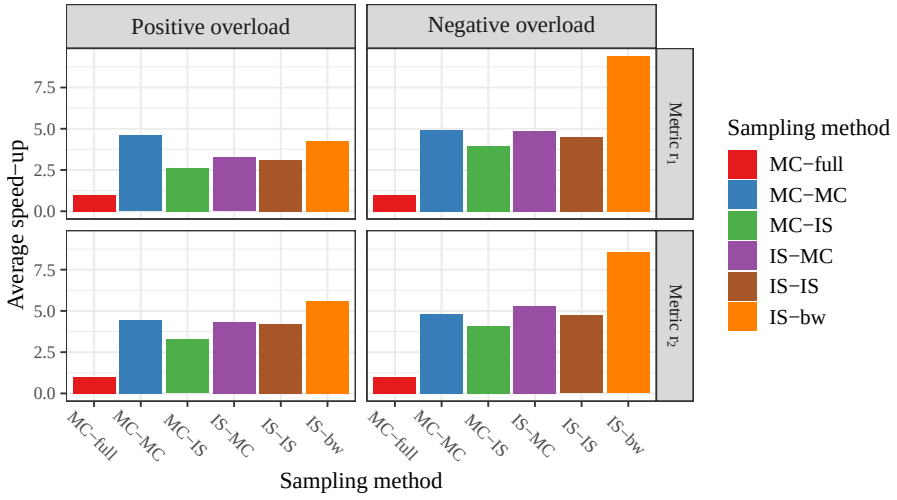


**Figure 7.11:** Comparison of the average speed-up $\bar{t}_{MC-full}/t_{method}$ for all methods with respect to the reference method MC-full. Only nonzero estimates and estimates whose accuracy was not found to differ significantly from the estimates of the reference method were used to compute the average speed-ups.

Figure 7.11 displays, that for negative overloads the IS-bw method clearly outperforms all other methods in terms of computational speed. This indicates that the chosen bin weights are able to achieve a considerable variance reduction in case of negative overloads. In contrast to other profile selection IS methods, due to the pre-specified bin weights importance sampling begins with the first trace in the IS-bw method. The computational cost of the optimisation stage which is part of the IS-MC and IS-IS profile importance sampling methods is saved. Also for positive overloads and metric $r_2$, the IS-bw method still outperforms all other methods. However, in light of the discussion in the previous section, the big caveat of bin weight importance sampling in the current form is that it only produces reliably accurate estimates for assets with less than 50-100 customers (depending on the desired safety margin the upper customer bound can be set more or less conservatively).

The solid performance of conventional Monte Carlo sampling of partial annual traces

(MC-MC) across all cases shown in figure 7.11 with respect to sampling entire annual traces (MC-full) is also a promising result, as it is easy to implement and its accuracy does not depend on the number of asset customers. Importance sampling of time steps does not appear to bring about speed-advantages compared to conventional MC sampling, neither alone (MC-IS) nor in combination with profile importance sampling (IS-IS).

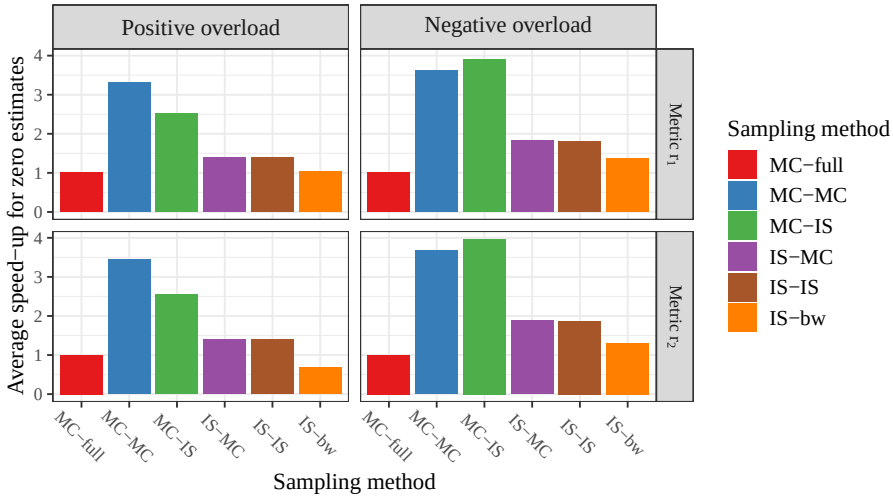


**Figure 7.12:** Comparison of the average speed-up $\bar{t}_{MC-full}/t_{method}$ of estimates with zero probability for all methods with respect to the reference method MC-full.

It is also interesting to consider average computation times in the case of MSRs which did not show any overload events, resulting in zero overload probabilities. As mentioned above, the computation times in these cases depend mainly on the stopping criteria which can be set more or less restrictively. Here, a simulation run was stopped after $n_{zero,max} = 10{,}000$ traces. Comparing average speed-ups for zero estimates, shown in figure 7.12, is telling in the sense that the relative overhead of the various methods can be compared. Clearly, and as expected, conventional Monte Carlo sampling (MC-MC) has the smallest overhead. Also, the time step importance sampling method MC-IS performs well.

Average speed-up comparisons across many assets with different properties hide much of the underlying detail. Generally, one of the most crucial parameters which influences the required computational effort of Monte Carlo methods is the magnitude of the quantity being estimated. Therefore, all estimates of metrics $r_{2,+}$ and $r_{2,-}$ have been binned according to their order of magnitude, and average speed-ups for each bin and method have been calculated. As for figure 7.11, only nonzero estimates and estimates which do not suffer from the accuracy issue are considered, and the same approach is used to calculate the average speed-ups shown in figure 7.13. Furthermore, the figure shows the number of traces (number of independent samples) necessary for convergence of each method in relation to MC-full, termed trace ratio. This ratio is computed as $\bar{n}_{MC-full}/n_{method}$, where $\bar{n}_{MC-full}$ is the average number of traces per asset and metric for the baseline method and $n_{method}$ is the number of traces for an individual replicate for all methods. Averages over these ratios calculated for individual replicates were formed per method, bin and metric, analogously to the procedure for the speed-up ratios. In

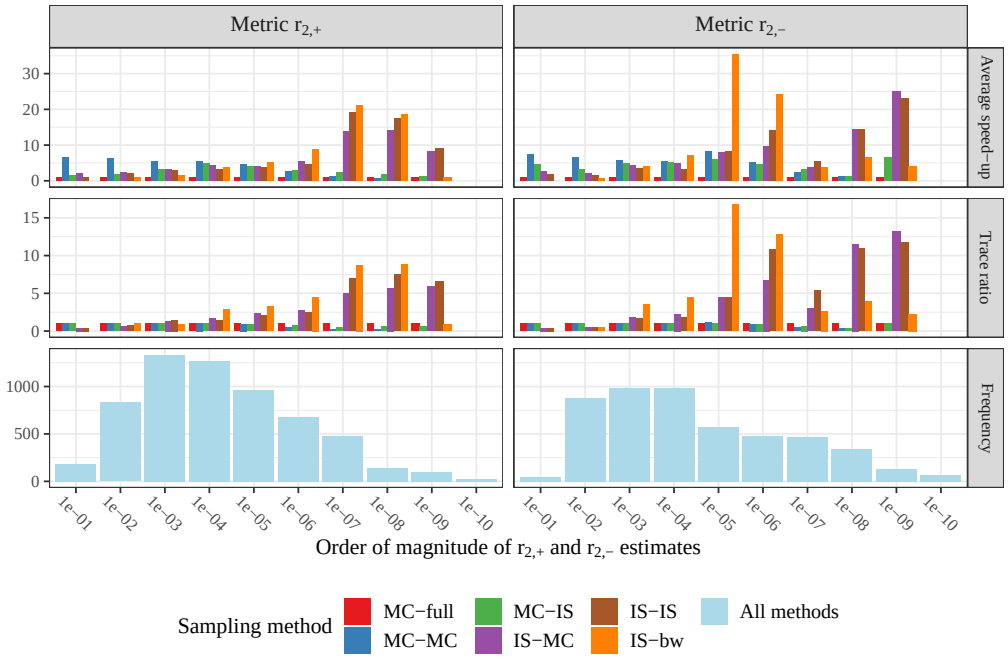


**Figure 7.13:** Comparison of the average speed-ups $\bar{t}_{MC-full}/t_{method}$ (top panels) and average trace ratios $\bar{n}_{MC-full}/n_{method}$ (middle panels) of estimates per bin for all methods with respect to the reference method MC-full. The bottom panels show histograms indicating how many estimates fall in which order of magnitude.

the bottom panel of the figure, the absolute frequency of estimates in all bins is shown, to indicate how the overload probability magnitudes are distributed over the shown range.

Three interesting conclusions can be drawn from figure 7.13: Firstly, for estimates in the order $10^{-5}$ and smaller, impressive speed-ups of more than 10 times are achieved for several bins. This is in good accordance with the theory referred to in section 2.3 and section 4.1, which describes the usefulness of importance sampling especially for the estimation of rare events. The highest speed-up of more than 30 times was found for metric $r_{2,-}$ in the $10^{-5}$ bin with the IS-bw method. Also the other profile importance sampling methods (IS-MC and IS-IS) show a strong performance for several bins on the right hand side. It is noteworthy that for the bins $10^{-8}$ and $10^{-9}$ IS-MC and IS-IS outperform the IS-bw method. The reason for this could be that a IS distribution optimised individually for each assets with the CE method is likely to be more suitable than the generalised IS distribution based on bin weights used in the IS-bw method.[14] For these particular two bins, this effect may outweigh the advantage of the IS-bw method of skipping the optimisation stage, which clearly materialises for several other bins.

Secondly, smaller advantages in terms of a lower number of traces required for conver-

**14** To avoid confusion: The bins of the IS-bw method are the 48 bins used in the demand model and are not related to the bins shown in figure figure 7.13.

gence is also visible for all profile IS methods in the bins $10^{-3}$ and $10^{-4}$. This indicates that the variance reductions works well also for estimates in these orders of magnitude. However, due to the higher overhead of the profile IS methods, this advantage does not materialise in the computation times (expect for bin $10^{-4}$ of metric $r_{2,-}$ with the IS-bw method). Thirdly, the bottom panel in figure 7.13 shows that many of the 150 MSRs have overload probabilities $r_2 > 10^{-6}$. As $10^{-6}$ was identified as the threshold above which overload probabilities become relevant from a distribution capacity planning perspective in the introduction of this section, this can be seen as a concerning finding regarding the available headroom of MV/LV substations. However, this finding is of course a result of how the demand model was specified and its validity strongly depends on the validity of the model. Therefore, one should be cautious in drawing general conclusion on this basis. More important for the main concern of this subsection is another implication of the observation that many MSRs have overload probabilities $r_2 > 10^{-6}$: It explains why the high speed-ups for the smaller orders of magnitude of the profile IS methods do not appears as pronounced in the average speed-ups shown in figure 7.11.

### 7.1.4 Performance of the Investigated Monte Carlo Methods for Substations

Due to limited computational resources, the Monte Carlo methods were only bench-marked on a small subset of HV/MV substations, *Onderstation* (OS) in Dutch, selected for modelling in subsection 5.2.1. This subset was identified after doing a trial run with conventional MC sampling and for a maximum of 500 traces on all 57 originally selected OSs. Nonzero overload probabilities were found for 13 out of the 57 OSs, whereas only four OSs did not converge to a relative error of $\leq 10\%$ before the maximum number of 500 traces was reached. These four OSs are the interesting cases for further investigation because only here the IS-based methods can potentially unfold their advantages. Furthermore, in the trail run no nonzero probabilities of negative overloads where found and therefore $r_{1,-}$ and $r_{2,-}$ were not considered further. A main run with $n_{max} = 2{,}500$ was then conducted for $r_{1,+}$ and $r_{2,+}$ for all methods (except for IS-bw which was not considered due to the accuracy issues discussed in subsection 7.1.2).

The comparison of the estimates obtained with the different methods in figure 7.14 reveals that almost all $r_{1,+}$ estimates are zero. Therefore, only the $r_{2,+}$ estimates are considered for bench-marking the computational efficiency of the various sampling methods in the following. It can also be gathered from the figure that the estimates obtained with the profile importance sampling methods IS-MC and IS-IS are in a much smaller order of magnitude than the estimates of the other methods. This is hardly surprising in light of the findings and discussion in subsection 7.1.2, because with at least 18,000 customers all of the considered OSs are very large assets and the 500 traces used in the CE optimisation stage are far from sufficient for correctly estimating suitable success probabilities for each customer.

The average speed-ups shown in figure 7.15 were computed analogously to how speed-ups were computed for MSRs in the previous subsection, using $\bar{t}_{MC-full}/t_{method}$ and similarly for the trace ratio $\bar{n}_{MC-full}/n_{method}$. In light of the results for the MSRs in the previous subsection, perhaps the most striking observation that can be made in the figure is the average speed-up of more than 3.5 times of the time step importance sampling method (MC-IS) which clearly outperforms all other methods for the considered four OSs. The reason why time step IS

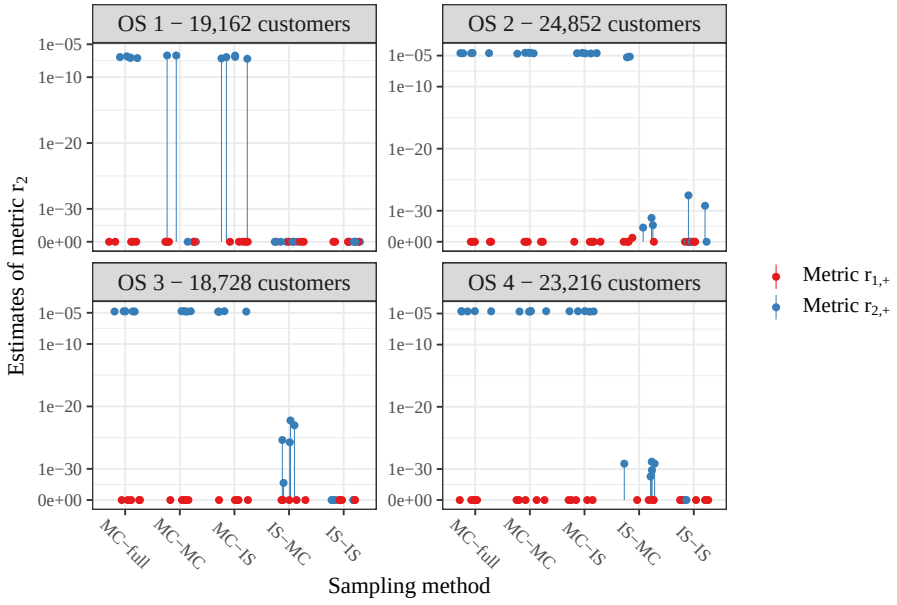


**Figure 7.14:** Comparison of the risk metric $r_{1,+}$ and $r_{2,+}$ estimates obtained with different sampling methods for the four investigated OSs. Point estimates and 95% confidence intervals of all independent replicates produced are shown on a pseudo-logarithmic scale which becomes linear as zero is approached. Confidence intervals are clipped at zero, as probabilities cannot be negative.

works so well here might have to do with the fact that large assets with many customers show stochastic variability to a smaller extent than assets with few customers. The effect can be observed quite clearly when comparing the 5th-95th percentile ranges of figure 5.7 and figure 5.9 in subsection 5.2.3, where the MSR with 41 customers shows much more variability than the OS with 7,505 customers. The lower degree of stochasticity in the aggregate behaviour of large OSs could result in overload events occurring more regularly at the same time of the year. A smaller subset of relevant time steps can be targeted more effectively by the time step IS algorithm which could explain its significantly better performance than all other methods.

While it is not as surprising that the profile selection IS methods IS-MC and IS-IS do not perform well here due to the large number of asset customers, it is more surprising that the conventional MC method of sampling 2,000 time steps per annual trace only leads to a small speed advantage compared to sampling 35,040 time steps. A look at the right panel in figure 7.15 suggests why this is the case: The ratio $\bar{n}_{MC-full}/n_{MC-MC}$ is much smaller than one which means that the MC-MC method requires much more traces for convergence than the MC-full method in the four considered cases on average. The likely reason for this is that many relevant time steps might be missed when randomly sampling partial annual traces. The effect results in the advantage of a lower computation time per partial annual trace to almost disappear, as a higher number of traces is required for convergence. This implies that for OSs with a large number of customers the optimal number of time steps sampled per trace is possibly higher than the value of $m = 2{,}000$ used here.

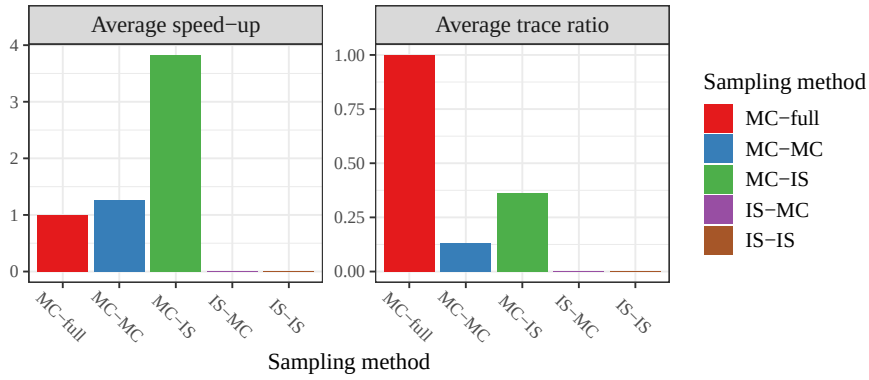


**Figure 7.15:** Comparison of the average speed-up $\bar{t}_{MC-full}/t_{method}$ (left panel) and average trace ratio $\bar{n}_{MC-full}/n_{method}$ (right panel) of estimates for all methods with respect to the reference method MC-full.

## 7.2 Inference of Demand Maxima and Minima – Risk Metric 3

Before proceeding with the presentation of the extreme value inference results, a brief reiteration of the definition of risk metric 3 may be useful: The metric is defined as the power value which will not be exceeded with 95% probability by the annual maxima, and equivalently the power value which will not be undercut with 95% probability by the annual minima. These values are given by the quantiles of the Generalised Extreme Value (GEV) distribution which can also be framed in terms of return levels – here the demand level which will only be exceeded once in 20 years. The results presented in this section have been obtained using the *R* package *extRemes* [39]. Estimates of the parameters of the GEV have been obtained by maximum likelihood estimation and confidence intervals were computed with the normal approximation.

The results of fitting the GEV distribution to the demand maxima series obtained from entire annual traces are shown for two exemplary MSRs in figure 7.16 and figure 7.17. It can be observed for MSR 23 shown in figure 7.16, that the distribution of maxima produced by the demand model is in good accordance with the theory for all diagnostic plots. This is in contrast to MSR 18 shown in figure 7.17, where the demand maxima deviate substantially from what the theory predicts above a certain magnitude. In the plot of the return level against the return period in the bottom right panel of the figure, two distinct steps and an upper plateau for the demand maxima magnitude are visible. In the density plot of the bottom left panel, the same phenomenon manifests as to clearly discernible bulges.

This phenomenon can be seen as a result of how the demand model was constructed: It relies on a fixed and finite set of smart meter profiles and is therefore a discrete model. In the example of figure 7.17 and similarly for other inspected examples, this discreteness shows quite distinctly. Potentially two profiles or two specific combinations of profiles are mainly responsible for the highest demand maxima in this example, leading to the two steps visible in

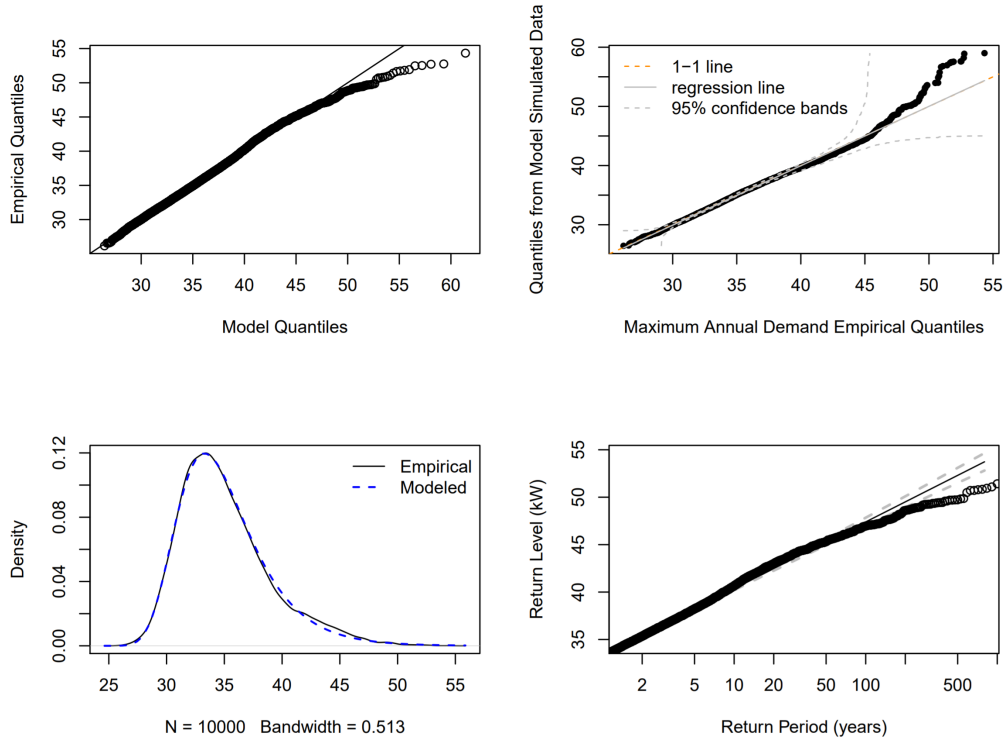


**Figure 7.16:** Diagnostic plots of the GEV distribution fit for the exemplary MSR 23 obtained with the *extRemes* package [39].

the figure, while other profiles cause the small variations around the plateau values which can be observed.

Extreme Value Theory (EVT), in turn, implicitly assumes the process generating the extreme values under scrutiny to be continuous and extrapolates beyond the available observations under that assumption. This assumption is certainly sensible, as many processes in nature and society are continuous. Also the real-world power demand behaviour of customers is continuous – it varies around characteristic patterns which will never repeat in exactly the same way. In the demand model, however, this continuous range is approximated by a discrete set of profiles. In some cases this approximation works very well, as can be observed in figure 7.16, while in other cases discrete characteristics become apparent. All this suggests that it would likely be wrong to conclude from the discrepancies between theoretical predictions and the simulated demand extreme values that the GEV distribution is not a suitable model for the given context. Rather, EVT assumes a more realistic model and extrapolates the given observations on that basis. Thus, the inference results obtained from the demand model can still give practically relevant indications.

In view of the overarching objective to obtain the risk metrics in the computationally most efficient way possible, it would be desirable to integrate the evaluation of risk metric 3 with the evaluation of risk metrics 1 and 2. The series of maxima used for the exemplary GEV

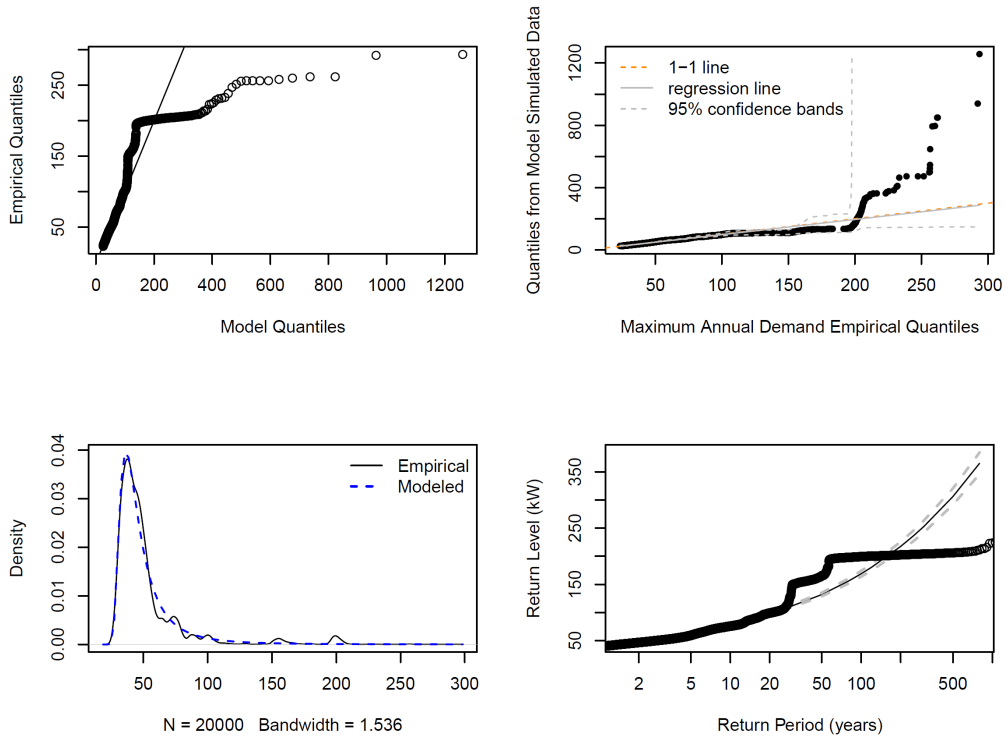


**Figure 7.17:** Diagnostic plots of the GEV distribution fit for the exemplary MSR 18 obtained with the *extRemes* package [39].

distribution fits in figure 7.16 and figure 7.17 were obtained from entire annual traces sampled for the respective assets. The crucial question is then, whether fitting the GEV distribution to the series of demand maxima and minima obtained from the other Monte Carlo sampling methods leads to approximately the same resulting 20-year return levels for metric 3.

To investigate this question, the estimates for $r_{3,+}$ and $r_{3,-}$ obtained for the replicates of each method (except for IS-bw) are compared in figure 7.18. While it appears as if the minima return levels of MSR 29 could be approximately in the same range for MC-full, MC-MC and MC-IS, this is certainly not the case for MSR 33. The visible tendency of smaller return levels in absolute terms for the MC-MC and MC-IS methods and larger return levels for the profile IS methods compared to the reference method MC-full, was also observed more generally when inspecting other examples. Due to the purposefully introduced bias of sampling spiky profiles with higher probability in the profile IS methods, it stands to reason that here the frequency of high demand maxima increases leading to generally higher return level estimates. It may be possible to make use of the importance weights to correct for the introduced bias, as done in IS. However, the theoretical foundations for carrying out such a correction could be new territory – no previous research was found on this particular matter – and are outside the scope of this thesis. Also the tendency for return levels of smaller absolute magnitude for MC-MC and MC-IS with respect to the reference method is explainable, considering that it is

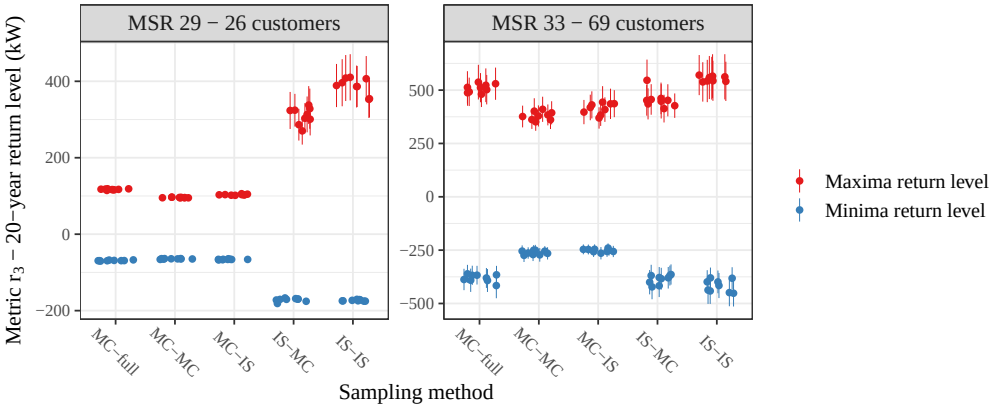


**Figure 7.18:** Comparison of the risk metric 3 estimates obtained based on the results of the different sampling methods for two exemplary MSRs. Point estimates and 95% (normal approximation) confidence intervals of all independent replicates produced obtained with the *extRemes* package [39] are shown. Note that in some cases the confidence intervals are too small to be visible.

likely to miss some high values when sampling a subset of time steps per annual trace. If this happens, smaller extreme values in absolute terms and consequently return levels result.

To systematically assess how comparable the $r_3$ estimates based on the different sampling methods are in relation to the reference method MC-full, Welch's $t$-test was employed in analogous manner as in subsection 7.1.2. The number of hypothesis test showing non-significant and significant differences with respect to the confidence level $\alpha = 0.05$ are shown in figure 7.19. It can be observed that for the vast majority of tests a significant difference was found. This suggests clearly that the 20-year return levels obtained from the results of MC-full and all other sampling methods are not in good accordance with each other. Therefore, sampling full annual traces appears to be required for a reliable estimation of extreme value return levels in the given context.

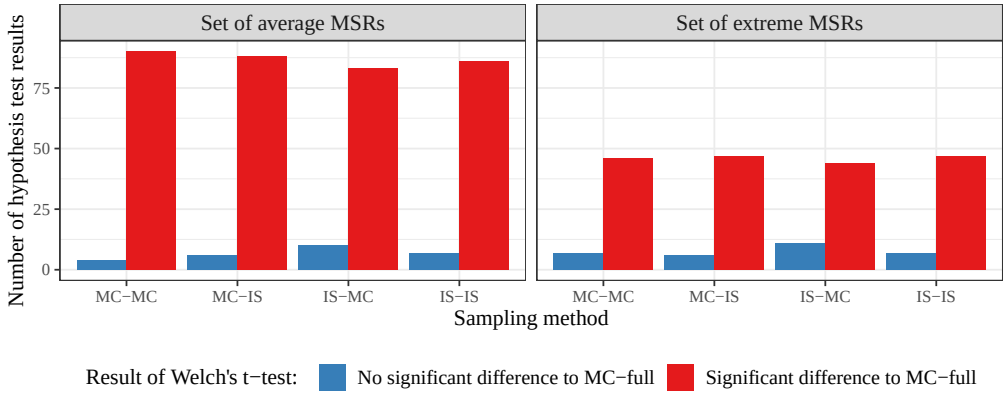


**Figure 7.19:** Visualisation of the hypothesis testing results with the significance level $\alpha = 5\%$ for the MSRs with average and the MSRs with extreme characteristics. Note that the number of assets in the set of extreme MSRs is lower compared to the set of average MSRs, therefore also the overall count of hypothesis tests results shown on the right is lower.

# 8 Conclusion and Outlook

The research in this thesis focused on two main objectives: Firstly, the creation of a probabilistic power demand model which reflects the volatile nature of real customer demand in an adequate manner. Secondly, the development and evaluation of methods to estimate certain model output quantities of interest in a computationally efficient manner. These two overarching objectives are motivated by the ambition of the Distribution System Operator (DSO) Alliander to explore possibilities of improving its predictive modelling capabilities of power demand – with the ultimate goal of preparing the distribution network operated by Alliander for the ongoing energy transition in an optimal way. The specific modelling application which may benefit from this research is the Advanced Net DEcision Support (ANDES) model. The model's purpose is the provision of detailed predictions on which long-term capacity planning decision can be based.

Even though the ANDES model represents a major improvement with respect to traditional capacity planning methods, it is likely that demand peaks and troughs are underestimated in many cases due to the use of average category profiles for the modelling of unmonitored customers. Monte Carlo simulation-based probabilistic demand models, such as the one implemented within this thesis, have the potential to reflect the natural variability of customer demand in a more realistic manner. However, the major drawback of the approach is the high computational cost which the simulation of many random realisations of customer demand entails.

With the goal of estimating model output quantities of interest in an efficient manner, variance reduction techniques have been investigated. The sample space of the developed demand model has two fundamental subspaces: The space comprising all possible ways in which measured smart meter profiles can be randomly assigned to unmonitored customers and the space comprising the 35,040 quarter-hourly time steps of a yearly demand profile. Importance Sampling (IS) of the states in both subspaces has been investigated separately and in combination. The art of devising effective importance sampling strategies consists in choosing a suitable importance distribution by exploiting available knowledge on the structure of the problem. In order to automate this process, several methods have been developed and evaluated which aim to sequentially approach the optimal IS distribution for a given grid asset as far as possible. Furthermore, an approach of generalising the importance distributions optimised for individual assets to all assets has been studied.

## 8.1 Research Questions

At the end of this work, the research questions posed in the introduction are taken up again to present the answers obtained and discuss where limitations were encountered:

**1. Which approaches exist to model distribution network power demand on the basis of limited customer monitoring data?**

In the literature study in chapter 2, three major groups of demand modelling approaches relying on limited smart meter measurement data as input were identified: Firstly, approaches in which appliance demand profiles are simulated and aggregated to obtain household and eventually residential sector electricity demand. Secondly, approaches which employ characteristic average household demand profiles for the modelling of unmonitored customers. Usually, the characteristic profiles are assigned to the modelled customers based on socio-economic or other household attributes. The ANDES model in its current version falls into this group of approaches. Thirdly, Monte Carlo simulation approaches were reviewed which, in this specific context, consist frequently in the repeated and random assignment of measured electricity consumption or generation profiles from a large pool to unmeasured customers.

**2. Which methods are able to improve the computational efficiency of estimating model output quantities of interest?**

The model output quantities of interest in this case are the risk metrics specified in chapter 5: The probability of demand exceeding a critical value above the rated asset capacity (1), the probability of exceeding the rated asset capacity (2) and the return levels of demand maxima and minima (3). Among the most common variance reduction methods reviewed in chapter 2, IS was identified as particularly suitable to estimate risk metrics 1 and 2 due to their rare event character. For the task of automatically determining a suitable IS distribution, the Cross-Entropy (CE) method was seen as promising. For risk metric 3, Extreme Value Theory (EVT), and more specifically the block maxima method, is the most relevant approach.

**3. What assumptions underlie the current ANDES model? Which of them are reasonable in light of research question (1) and which requirements for the implementation of the demand model can be derived from the discussion?**

The principle issue identified in chapter 3 regarding these matters is the modelling of unmonitored customers using average category profiles which are smoother and show less of the desirable stochastic variability of real customer demand. This is likely to lead to the underestimation of demand extremes. Furthermore, certain aspects of how the future development of technologies is modelled may be problematic. However, this is not within the scope of this thesis and the focus was placed on the first issue. To address the issue, the principle requirement regarding the baseload demand modelling derived from the discussion was to introduce more stochastic variability – however, while striking a reasonable balance between introducing too little and too much variability, to avoid under- or overestimating network impacts.

With this aim in mind, a binning scheme of smart meter profiles and modelled customers according to their yearly consumption was integrated in the demand model. The idea here was to confine variability to a reasonable extent by randomly assigning profiles to the modelled customers within similar yearly consumption ranges. However, a practical limitation was encountered when implementing the idea in chapter 5: It became clear that the given set of smart meter profiles used in this thesis has a much smaller yearly consumption range compared to the customers which are to be modelled for most of the customer categories.

This may limit the effectiveness of the binning approach in practice, as it can occur that spiky profiles with a relatively small yearly consumption are scaled to a probably unrealistic extend. Aside from this challenging aspect, the demand model was found to fulfil the requirements specified in subsection 3.2.2.

**4. How do the most promising methods from research question (2) perform in the given context?**

As introduced in the beginning of this chapter, self-tuning IS methods giving different weight to time steps and/or customer load profiles have been developed. The purpose of the self-tuning methods is to find the most suitable IS distribution per grid asset. In order to make the optimisation stage superfluous, additionally an approach to generalise the per customer probabilities of choosing more spiky profiles to per bin probabilities (where the bins are the 48 bins of the demand model referred to above) has been implemented. Next to these methods, also conventional Monte Carlo (MC) sampling of time steps and profiles was implemented. Sampling entire annual traces from the demand model (corresponding to conventional MC sampling of profiles only) served as the reference method for the evaluation of the accuracy and efficiency of the various approaches.

The essential finding from evaluating all considered methods was that their performance in terms of efficiency and accuracy depends mainly on two variables – the order of magnitude of the estimated quantity and the number of customer connected to an asset in question:

- For small assets with less than 50-80 customers and an estimated overload probability of the order $10^{-5}$ or smaller, all profile IS methods and especially the generalised bin probability IS showed the strongest performance with average speed-ups of 5-30 times with respect to the reference method of sampling full annual traces. Furthermore, the estimates were found to be accurate.

- For assets with more than 80 customers and small overload probabilities, the profile IS methods were found to frequently produce estimates of a much too small order of magnitude. Conventional MC sampling and time step IS, in turn, were found to produce reliable estimates regardless of the number of customers.

- For assets of all sizes with an estimated overload probability of the order $10^{-4}$ and larger, conventional MC sampling showed the best performance with speed-ups above 5 times.

- For the few investigated very large assets with more than 18,000 customers, time step IS was found to outperform all other methods with speed-ups of more than 3.5 times.

Thus, the bottom line here is that conventional MC sampling was found to perform robustly in all circumstances, while IS demonstrated its potential to significantly increase the estimation efficiency of rare event probabilities in certain cases. This is in good accordance with what is theoretically expected (as described in section 2.3). An interesting result was that using generalised per bin probabilities for the profile IS resulted in the highest observed speed-ups. This suggest that it is principally possible to use the same distribution for many assets and to avoid the computational cost of the optimisation stage. A perhaps more unexpected finding was the accuracy issue observed for profile IS in the case of large assets. Its occurrence is most probably related to the usage of too extreme importance weights, as detailed in subsection 7.1.2,

leading to the conclusion that caution should be taken to prevent extreme weights. This shows what the main challenge in finding a generalised profile IS distribution is. Finally, it also became apparent that most of the MV/LV substations have overload probabilities $> 10^{-6}$. This implies that rare event probabilities occur less frequently than initially expected and that, therefore, conventional MC sampling is a good choice in many cases.

The major limitation in the evaluation of the developed sampling methods were scarce computational resources. For example, it would have been desirable to tune the parameters of the developed methods also for much larger assets and in multiple combinations. For the four very large HV/MV substations investigated, it was observed that time step IS performed best which was never the case for smaller assets. To verify the generality of this finding, it would have been useful to test on a bigger set of HV/MV substations, however, this was challenging to do because computation times are significantly higher here than for smaller assets. A limitation in evaluating the generalised bin probability IS approach was that the same assets were used for training and testing. Testing the approach on different assets would have required testing the other sampling methods on the different assets as well to allow for speed-up comparisons. Due to limited time and computational resources, this was not possible. Nevertheless, if generalisation of asset importance distributions were not possible at all, the strong performance of the bin weight IS method would probably not have been observed.

**5. How compatible are the MC sampling-based methods developed for the first two model output quantities of interest and extreme value inference for the third quantity in the given context?**

Risk metric 3 – the 20-year return levels of demand maxima and minima – was obtained by fitting the Generalised Extreme Value (GEV) distribution to the extreme values obtained from the yearly traces of the demand model. To investigate the compatibility of the Monte Carlo methods with extreme value inference, it was investigated whether similar 20-year return levels result based on the reference method compared to all other sampling methods. This was not found to be the case, sampling full annual traces appears to be required for a reliable estimation of maximum and minimum demand return levels. Therefore, obtaining risk metric 3, while also estimating risk metrics 1 and 2 with one of the computationally more efficient sampling methods, is not possible. On the upside, extreme value inference works already well for relatively small sample sizes from 200 upwards (see section 4.2) and the separate evaluation of risk metric 3 would, therefore, require much less samples than a typical run of the Monte Carlo methods to obtain the other risk metrics.

## 8.2 Avenues for Future Research

The following ideas and suggestions for future research have emerged while carrying out the research of this thesis:

- During the implementation of the demand model, it became clear that it would be worthwhile to further investigate how customers with Solar Photovoltaic (PV) installations are modelled. In the newest version of ANDES (which was released after the demand model had already been built for this thesis), PV customers and other customers are distinguished in separate categories. Incorporating the new customer categories into the demand model

of this thesis has the potential of improving its prediction accuracy. Furthermore, it would also be interesting to use a set of smart meter profiles with a wider yearly consumption range as input for the demand model and to observe to what extent the model's properties change. If the yearly consumption range of smart meter profiles used corresponded more closely to the yearly consumption range of the customers to be modelled, the over-scaling of spiky profiles with a low yearly consumption could be prevented.

- It was found that there is not a single best-performing Monte Carlo method for all types of assets. Rather, a strong performance dependency on the number of asset customers and the magnitude of the estimate became apparent. This suggests that it could be promising to implement a flexible sampling algorithm which starts with conventional MC sampling and decides then, based on the magnitude of the estimate obtained so far and other asset parameters, whether to proceed with conventional MC or IS, if a rare event probability is encountered. Such a flexible algorithm could also start with sampling 200 entire annual traces, which can be used for extreme value inference, before switching to the sampling of partial annual traces. Overall, this would allow evaluating all risk metrics together and reaping the benefits of all methods at once.

- At the level of the sampling algorithms themselves, there are also several interesting possibilities for future research: It was observed that not all bins of smart meter profiles contain the same amount of very spike profiles. Therefore, it could be worthwhile to set the threshold which decides whether a profile is considered spiky or not spiky in a more flexible manner. If this is possible, the advantage could be a more targeted profile IS. Continuing along these lines, it could also be effective to split the profiles in each bin in more than two categories (e.g. average, spiky and very spiky) and optimise the probabilities of drawing profiles from these category with the CE method.

- In the literature on the CE method an extension of the common algorithm is proposed, termed Fully Automated Cross-Entropy (FACE) algorithm [27]. In the FACE algorithm, the sample size used per iteration in the optimisation stage is adapted in a flexible manner. The idea here is to update the CE optimised parameters always based on a fixed number of the best performing samples, and to continue collecting samples in the optimisation stage until this fixed number is reached. Potentially, the FACE algorithm could prevent weights from being set in a too extreme manner which was identified as the likely cause of the accuracy issue.

- It would be interesting to further investigate possibilities of generalising asset-centred IS distributions to wider classes of assets. Clearly, hereby strategies to prevent too extreme IS distributions would be crucial. For example, the usage of high bin weights could be restricted to a fixed maximum number of customers, potentially those with the highest yearly consumption as they are likely to have the biggest impact on the aggregate asset demand.

- An application within Alliander, related to the ANDES model, are the grid losses calculations. In their current version, the grid losses calculations rely as well on the average customer category profiles used within ANDES. Therefore, they are prone to the same peak demand underestimation issues as ANDES. Investigating a suitable sampling strategy for the grid

losses application would be an interesting topic for future research, because estimating the power losses over a given period of time is not a rare event estimation task. For this reason, using a different variance reduction technique than importance sampling would likely be more appropriate. For example, a stratified sampling approach could lead to an effective variance reduction here, since for estimating the expected grid losses over a period of time all parts of their distribution matter. By taking samples from all strata, stratified sampling ensures more representative samples of the overall distribution than random sampling. A refined stratified sampling algorithm with a self-adapting sample size is proposed in [82], which constitutes a promising starting point for potential future research on this topic.

# Bibliography


[1] Joana M Abreu, Francisco Câmara Pereira and Paulo Ferrão. **Using pattern recognition to identify habitual behavior in residential electricity consumption**. *Energy and buildings* 49 (2012), 479–487 (see page 12).

[2] Jamshid Aghaei, Mohammad Amin Akbari, Alireza Roosta and Amir Baharvandi. **Multiobjective generation expansion planning considering power system adequacy**. *Electric Power Systems Research* 102 (2013), 8–19 (see page 6).

[3] R.N. Allan, Roy Billinton, A.M. Breipohl and C.H. Grigg. **Bibliography on the application of probability methods in power system reliability evaluation**. *Power Systems, IEEE Transactions on* 9 (Mar. 1999), 51–57 (see page 8).

[4] C Henggeler Antunes, A Gomes Martins and Isabel Sofia Brito. **A multiple objective mixed integer linear programming model for power generation expansion planning**. *Energy* 29:4 (2004), 613–627 (see page 6).

[5] Søren Asmussen and Peter W Glynn. **Stochastic simulation: algorithms and analysis**. Vol. 57. Springer Science & Business Media, 2007 (see page 33).

[6] DB Belzer and MA Kellogg. **Incorporating sources of uncertainty in forecasting peak power loads-a Monte Carlo analysis using the extreme value distribution**. *IEEE transactions on power systems* 8:2 (1993), 730–737 (see page 19).

[7] R Billinton and L Goel. **Overall adequacy assessment of an electric power system**. In: *IEE Proceedings C (Generation, Transmission and Distribution)*. Vol. 139. 1. IET. 1992, 57–63 (see page 9).

[8] Roy Billinton and Ronald Norman Allan. **Reliability evaluation of engineering systems: concepts and techniques**. Springer, 1992 (see pages 14, 15).

[9] Roy Billinton and Ronald N. Allen. **Reliability assessment of large electric power systems**. Kluwer Academic Publishers, Norwell, MA, 1988 (see pages 7–9).

[10] Roy Billinton and Yi Gao. **Multistate wind energy conversion system models for adequacy assessment of generating systems incorporating wind energy**. *IEEE Transactions on Energy Conversion* 23:1 (2008), 163–170 (see page 8).

[11] Roy Billinton and Dange Huang. **Effects of load forecast uncertainty on bulk electric system reliability evaluation**. *IEEE Transactions on Power Systems* 23:2 (2008), 418–425 (see page 9).

[12] Roy Billinton and A Jonnavithula. **Composite system adequacy assessment using sequential Monte Carlo simulation with variance reduction techniques**. *IEE Proceedings-Generation, Transmission and Distribution* 144:1 (1997), 1–6 (see page 16).

[13] Gino Biondini. 'An introduction to rare event simulation and importance sampling'. In: *Handbook of Statistics*. Vol. 33. Elsevier, 2015, 29–68 (see pages 15, 17, 35–37).

[14] Hans W. Borchers. *pracma: Practical Numerical Math Functions*. R package version 2.2.9. 2019. URL: https://CRAN.R-project.org/package=pracma (see page 50).

[15] Jilin Cai, Qingshan Xu, Minjian Cao and Bin Yang. **A novel importance sampling method of power system reliability assessment considering multi-state units and correlation between wind speed and load**. *International Journal of Electrical Power & Energy Systems* 109 (2019), 217–226 (see pages 17, 18).

[16] S Caires. **A comparative simulation study of the annual maxima and the peaks-over-threshold methods**. *Deltares report 1200264-002 for Rijkswaterstaat, Waterdienst* (2009) (see page 39).

[17] CBS. **Onderzoek Verplaatsingen in Nederland (OViN) 2015**. Tech. rep. Centraal Bureau voor de Statistiek, 2016 (see page 24).

[18] CBS. **Standaard Bedrijfsindeling 2008 - update 2019**. Tech. rep. Centraal Bureau voor de Statistiek, 2019. URL: https://www.kvk.nl/overzicht-standaard-bedrijfsindeling/ (visited on 18/03/2020) (see page 22).

[19] Can Chen, Wenchuan Wu, Boming Zhang and Chanan Singh. **An analytical adequacy evaluation method for distribution networks considering protection strategies and distributed generators**. *IEEE Transactions on Power Delivery* 30:3 (2015), 1392–1400 (see page 9).

[20] Haoling Chen and Tongtiegang Zhao. **Modeling power loss during blackouts in China using non-stationary generalized extreme value distribution**. *Energy* 195 (2020), 117044 (see pages 19, 20).

[21] Quan Chen and Lamine Mili. **Composite power system vulnerability evaluation to cascading failures using importance sampling and antithetic variates**. *IEEE transactions on power systems* 28:3 (2013), 2321–2330 (see page 16).

[22] Lonnie Chrisman. *Latin hypercube vs. Monte Carlo sampling*. 2014. URL: https://lumina.com/latin-hypercube-vs-monte-carlo-sampling/ (visited on 28/07/2020) (see page 16).

[23] World Energy Council and Oliver Wyman. **World Energy Trilemma Report 2018**. Tech. rep. World Energy Council, 2018. URL: https://www.worldenergy.org/assets/downloads/World-Energy-Trilemma-Index-2018.pdf (visited on 22/07/2020) (see page 6).

[24] SHF Cunha, MVF Pereira, LMVG Pinto and GC Oliveira. **Composite generation and transmission reliability evaluation in large hydroelectric systems**. *IEEE transactions on power apparatus and systems*: 10 (1985), 2657–2663 (see page 17).

[25] AM Leite Da Silva, J Endrenyi and L Wang. **Integrated treatment of adequacy and security in bulk power system reliability evaluations**. *IEEE Transactions on Power Systems* 8:1 (1993), 275–285 (see page 9).

[26] Armando M Leite Da Silva, Reinaldo AG Fernandez and Chanan Singh. **Generating capacity reliability evaluation based on Monte Carlo simulation and cross-entropy methods**. *IEEE Transactions on Power Systems* 25:1 (2010), 129–137 (see pages 17, 81).

[27] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor and Reuven Y Rubinstein. **A tutorial on the cross-entropy method**. *Annals of operations research* 134:1 (2005), 19–67 (see pages 81, 107).

[28] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä and Ludolf Erwin Meester. **A Modern Introduction to Probability and Statistics: Understanding why and how**. Springer Science & Business Media, 2005 (see pages 33, 48).

[29] Ahmad Salehi Dobakhshari and Mahmud Fotuhi-Firuzabad. **A reliability model of large wind farms for power system adequacy studies**. *IEEE Transactions on Energy Conversion* 24:3 (2009), 792–801 (see page 8).

[30] Arnaud Doucet, Nando De Freitas and Neil Gordon. 'An introduction to sequential Monte Carlo methods'. In: *Sequential Monte Carlo methods in practice*. Springer, 2001, 3–14 (see page 18).

[31] K Dragoon and V Dvortsov. **Z-method for power system resource adequacy applications**. *IEEE Transactions on Power Systems* 21:2 (2006), 982–988 (see page 8).

[32] Dirk Eddelbuettel and Romain François. **Rcpp: Seamless R and C++ Integration**. *Journal of Statistical Software* 40:8 (2011), 1–18. DOI: 10.18637/jss.v040.i08. URL: http://www.jstatsoft.org/v40/i08/ (see page 50).

[33] *Energieleveren. Registration of home-owned PV systems (no page title available)*. 2020. URL: https://www.energieleveren.nl/ceres/index.html (visited on 17/08/2020) (see page 55).

[34] Energieonderzoek Centrum Nederland. *FLEXNET*. 2020. URL: https://www.ecn.nl/nl/flexnet/index.html (visited on 17/03/2020) (see page 24).

[35] Ana Ferreira and Laurens De Haan. **On the block maxima method in extreme value theory: PWM estimators**. *The Annals of Statistics* 43:1 (2015), 276–298 (see page 39).

[36] Len L Garver. **Reserve planning using outage probabilities and load uncertainties**. *IEEE Transactions on Power Apparatus and Systems*: 4 (1970), 514–521 (see page 8).

[37] Sebastian Geyer, Iason Papaioannou and Daniel Straub. **Cross entropy-based importance sampling using Gaussian densities revisited**. *Structural Safety* 76 (2019), 15–27 (see pages 36, 37).

[38] Georgios Giasemidis, Stephen Haben, Tamsin Lee, Colin Singleton and Peter Grindrod. **A genetic algorithm approach for modelling low voltage network demands**. *Applied Energy* 203 (2017), 463–473 (see pages 11–14).

[39] Eric Gilleland and Richard W. Katz. **extRemes 2.0: An Extreme Value Analysis Package in R**. *Journal of Statistical Software* 72:8 (2016), 1–39. DOI: 10.18637/jss.v072.i08 (see pages 50, 98–101).

[40] Manfred Gilli and Evis Këllezi. **An application of extreme value theory for measuring financial risk**. *Computational Economics* 27:2-3 (2006), 207–228 (see pages 38–40).

[41] S. Haben, M. Rowe, D. V. Greetham, P. Grindrod, W. Holderbaum, B. Potter and C. Singleton. **Mathematical solutions for electricity networks in a low carbon future**. In: *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*. 2013, 1–4 (see page 13).

[42] Nicholas E. Hamilton and Michael Ferry. **ggtern: Ternary Diagrams Using ggplot2**. *Journal of Statistical Software, Code Snippets* 87:3 (2018), 1–17. DOI: 10.18637/jss.v087.c03 (see pages 50, 67).

[43] R. K. S. Hankin. **Very large numbers in R: Introducing package Brobdingnag**. *R News* 7 (3 2007) (see page 50).

[44] Y. G. Hegazy, M. M. A. Salama and A. Y. Chikhani. **Adequacy assessment of distributed generation systems using Monte Carlo simulation**. *IEEE Transactions on Power Systems* 18:1 (2003), 48–52 (see page 9).

[45] Adriaan P Hilbers, David J Brayshaw and Axel Gandy. **Quantifying demand and weather uncertainty in power system models using the m out of n bootstrap**. *arXiv preprint arXiv:1912.10326v2* (2020) (see page 28).

[46] Gerwin Hoogsteen, Albert Molderink, Johann L Hurink, Gerard JM Smit, Friso Schuring and Ben Kootstra Liandon. **Impact of peak electricity demand in distribution grids: A stress test**. In: *2015 IEEE Eindhoven PowerTech*. IEEE. 2015, 1–6 (see page 11).

[47] Dange Huang and Roy Billinton. **Effects of wind power on bulk system adequacy evaluation using the well-being analysis framework**. *IEEE transactions on power systems* 24:3 (2009), 1232–1240 (see page 9).

[48] ASN Huda and Rastko Živanović. **Improving distribution system reliability calculation efficiency using multilevel Monte Carlo method**. *International Transactions on Electrical Energy Systems* 27:7 (2017), e2333 (see page 16).

[49] IRENA. **Innovation landscape brief: Future role of distribution system operators**. Tech. rep. International Renewable Energy Agency, Abu Dhabi, 2019. URL: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Feb/IRENA_Landscape_Future_DSOs_2019.PDF?la=en&hash=EDEBEDD537DE4ED1D716F4342F2D55D890EA5B9A (visited on 20/07/2020) (see page 1).

[50] Sergei Izrailev. *binr: Cut Numeric Values into Evenly Distributed Groups*. R package version 1.1. 2015. URL: https://CRAN.R-project.org/package=binr (see pages 50, 56).

[51] Maria Jacob, Cláudia Neves and Danica Vukadinović Greetham. **Forecasting and Assessing Risk of Individual Electricity Peaks**. Springer, 2020 (see pages 19, 38, 40).

[52] Wojciech Jarosz. **Efficient Monte Carlo Methods for Light Transport in Scattering Media**. PhD thesis. UC San Diego, Sept. 2008 (see page 16).

[53] Alboukadel Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.2.4. 2019. URL: https://CRAN.R-project.org/package=ggpubr (see page 50).

[54] Alboukadel Kassambara. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.6.0. 2020. URL: https://CRAN.R-project.org/package=rstatix (see page 87).

[55] Suresh K Khator and Lawrence C Leung. **Power distribution planning: A review of models and issues**. *IEEE Transactions on Power Systems* 12:3 (1997), 1151–1159 (see page 7).

[56] Walid El-Khattam, Kankar Bhattacharya, Yasser Hegazy and MMA Salama. **Optimal investment planning for distributed generation in a competitive electricity market**. *IEEE Transactions on power systems* 19:3 (2004), 1674–1684 (see page 7).

[57] Abdolrahman Khoshrou, Andre B Dorsman and Eric J Pauwels. **SVD-based Visualisation and Approximation for Time Series Data in Smart Energy Systems**. In: *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE. 2017, 1–6 (see page 67).

[58] Sudhir Kumar and Roy Billinton. **Adequacy equivalents in composite power system evaluation**. *IEEE transactions on power systems* 3:3 (1988), 1167–1173 (see page 9).

[59] Yun Li and Ben Jones. **The Use of Extreme Value Theory for Forecasting Long-Term Substation Maximum Electricity Demand**. *IEEE Transactions on Power Systems* 35:1 (2020), 128–139 (see page 20).

[60] David JC MacKay and David JC Mac Kay. **Information theory, inference and learning algorithms**. Cambridge university press, 2003 (see page 36).

[61] Michael D McKay, Richard J Beckman and William J Conover. **A comparison of three methods for selecting values of input variables in the analysis of output from a computer code**. *Technometrics* 42:1 (2000), 55–61 (see page 16).

[62] Fintan McLoughlin, Aidan Duffy and Michael Conlon. **A clustering approach to domestic electricity load profile characterisation using smart metering data**. *Applied energy* 141 (2015), 190–199 (see page 12).

[63] Fintan McLoughlin, Aidan Duffy and Michael Conlon. **Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study**. *Energy and buildings* 48 (2012), 240–248 (see page 12).

[64] Fintan McLoughlin, Aidan Duffy and Michael Conlon. **Evaluation of time series techniques to characterise domestic electricity demand**. *Energy* 50 (2013), 120–130 (see page 12).

[65] Dougal HO McQueen, Patrick R Hyland and Simon J Watson. **Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks**. *IEEE Transactions on power systems* 19:3 (2004), 1685–1689 (see page 10).

[66] Janine Morley and Michael Hazas. **The significance of difference: Understanding variation in household energy consumption**. In: eceee. 2011 (see page 13).

[67] Matteo Muratori, Matthew C Roberts, Ramteen Sioshansi, Vincenzo Marano and Giorgio Rizzoni. **A highly resolved modeling technique to simulate residential power demand**. *Applied Energy* 107 (2013), 465–473 (see page 12).

[68] Alejandro Navarro-Espinosa, Luis F Ochoa and Dan Randles. **Assessing the benefits of meshed operation of LV feeders with low carbon technologies**. In: *ISGT 2014*. IEEE. 2014, 1–5 (see page 13).

[69] Alejandro Navarro, Luis F Ochoa and Dan Randles. **Monte Carlo-based assessment of PV impacts on real UK low voltage networks**. In: *2013 IEEE Power & Energy Society General Meeting*. IEEE. 2013, 1–5 (see page 13).

[70] Myriam Neaimeh, Robin Wardle, Andrew M Jenkins, Jialiang Yi, Graeme Hill, Padraig F Lyons, Yvonne Hübner, Phil T Blythe and Phil C Taylor. **A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts**. *Applied Energy* 157 (2015), 688–698 (see page 13).

[71] M Nijhuis, M Gibescu and JFG Cobben. **Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data**. *Energy and Buildings* 112 (2016), 121–129 (see page 12).

[72] JH Pickels and IH Russell. **Importance sampling for power system security assessment**. In: *1991 Third International Conference on Probabilistic Methods Applied to Electric Power Systems*. IET. 1991, 47–52 (see page 17).

[73] PV Output. *Source of PV Generation Data in the Utrecht Region (no page title available)*. 2020. URL: https://www.pvoutput.org (visited on 17/03/2020) (see page 24).

[74] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: https://www.R-project.org/ (see page 49).

[75] Rüdiger Rackwitz. **Reliability analysis—a review and some perspectives**. *Structural Safety* 23:4 (2001), 365–395 (see page 15).

[76] N Ragi Manoj. **First Order Reliability Method: Concepts and Application**. MA thesis. 2016 (see page 15).

[77] Teemu Räsänen, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas and Mikko Kolehmainen. **Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data**. *Applied Energy* 87:11 (2010), 3538–3545 (see page 12).

[78] Ian Richardson, Murray Thomson, David Infield and Conor Clifford. **Domestic electricity use: A high-resolution energy demand model**. *Energy and buildings* 42:10 (2010), 1878–1887 (see page 12).

[79] Reuven Y Rubinstein and Dirk P Kroese. **Simulation and the Monte Carlo method**. Vol. 10. John Wiley & Sons, 2016 (see pages 15–18, 31, 37, 73).

[80] Choonghyun Ryu. *dlookr: Tools for Data Diagnosis, Exploration, Transformation*. R package version 0.3.13. 2020. URL: https://CRAN.R-project.org/package=dlookr (see page 50).

[81] Peter van de Sande, Matthijs Danes and Tobias Dekker. **ANDES: grid capacity planning using a bottom-up, profile-based load forecasting approach**. *CIRED-Open Access Proceedings Journal* 2017:1 (2017), 2097–2100 (see pages 1, 10, 11, 23–25, 28).

[82] Michael D Shields, Kirubel Teferra, Adam Hapij and Raymond P Daddazio. **Refined stratified sampling for efficient Monte Carlo based uncertainty quantification**. *Reliability Engineering & System Safety* 142 (2015), 310–325 (see page 108).

[83] Zhen Shu and Panida Jirutitijaroen. **Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources**. *IEEE Transactions on Power Systems* 26:4 (2011), 2066–2073 (see page 16).

[84] Zhen Shu, Panida Jirutitijaroen and Bordin Bordeerath. **Reliability evaluation of composite power systems using sequential simulation with Latin Hypercube Sampling**. In: *2014 Power Systems Computation Conference*. IEEE. 2014, 1–6 (see page 16).

[85] Carson Sievert. *plotly for R*. 2018. URL: https://plotly-r.com (see page 50).

[86] Sebastián Solari, Marta Egüen, María José Polo and Miguel A Losada. **Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value**. *Water Resources Research* 53:4 (2017), 2833–2849 (see page 39).

[87] Lukas G Swan and V Ismet Ugursal. **Modeling of end-use energy consumption in the residential sector: A review of modeling techniques**. *Renewable and sustainable energy reviews* 13:8 (2009), 1819–1835 (see pages 11–13).

[88] J. Tan and L. Wang. **Adequacy Assessment of Power Distribution Network With Large Fleets of PHEVs Considering Condition-Dependent Transformer Faults**. *IEEE Transactions on Smart Grid* 8:2 (2017), 598–608 (see page 9).

[89] Hatice Tekiner, David W Coit and Frank A Felder. **Multi-period multi-objective electricity generation expansion planning problem with Monte-Carlo simulation**. *Electric Power Systems Research* 80:12 (2010), 1394–1405 (see page 6).

[90] Simon Tindemans and Goran Strbac. **Accelerating System Adequacy Assessment using the Multilevel Monte Carlo Approach**. *arXiv preprint arXiv:1910.13013v2* (2019) (see pages 16–18, 31, 32).

[91] Egill Tomasson and Lennart Söder. **Improved importance sampling for reliability evaluation of composite power systems**. *IEEE Transactions on Power Systems* 32:3 (2016), 2426–2434 (see page 17).

[92] Tom Valckx, Eric Cator and Jacco Heres. **Stochastic effects of customers on peak loads in a power distribution grid**. MA thesis. 2019 (see pages 2, 11, 43–45, 57, 63).

[93] Vereniging Nederlandse EnergieDataUitwisseling. *Verbruiksprofielen*. 2020. URL: https://www.nedu.nl/documenten/verbruiksprofielen/ (visited on 18/03/2020) (see page 22).

[94] Hadley Wickham. **Reshaping Data with the reshape Package**. *Journal of Statistical Software* 21:12 (2007), 1–20. URL: http://www.jstatsoft.org/v21/i12/ (see page 50).

[95] Hadley Wickham. **The Split-Apply-Combine Strategy for Data Analysis**. *Journal of Statistical Software* 40:1 (2011), 1–29. URL: http://www.jstatsoft.org/v40/i01/ (see page 50).

[96] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo and Hiroaki Yutani. **Welcome to the tidyverse**. *Journal of Open Source Software* 4:43 (2019), 1686. DOI: 10.21105/joss.01686 (see page 50).

[97] Amy L Wilson and Stan Zachary. **Using extreme value theory for the estimation of risk metrics for capacity adequacy assessment**. *arXiv preprint arXiv:1907.13050v1* (2019) (see page 19).

[98] W Zhang and R Billinton. **Application of an adequacy equivalent method in bulk power system reliability evaluation**. *IEEE Transactions on power systems* 13:2 (1998), 661–666 (see page 9).

# A Supplementary Results Figures

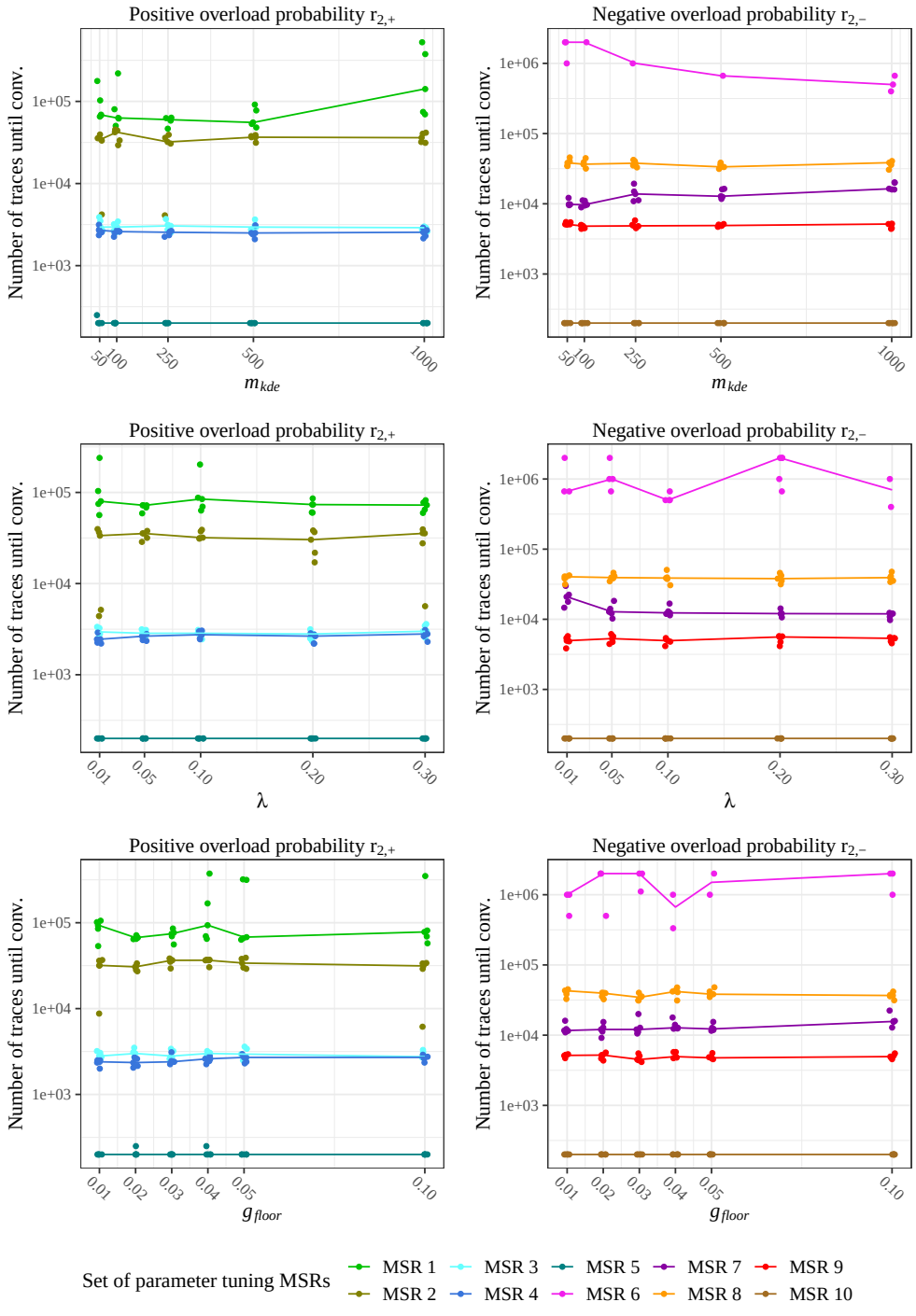


**Figure A.1:** Parameter variation results for the time step importance sampling algorithm. The number of traces needed to reach a relative error ≤ 10% is shown for each tested parameter value and replicate (points). A small amount of random noise has been added to the x-coordinate of the points to avoid overplotting. The medians of each set of five replicates are connected to highlight potential trends (lines).

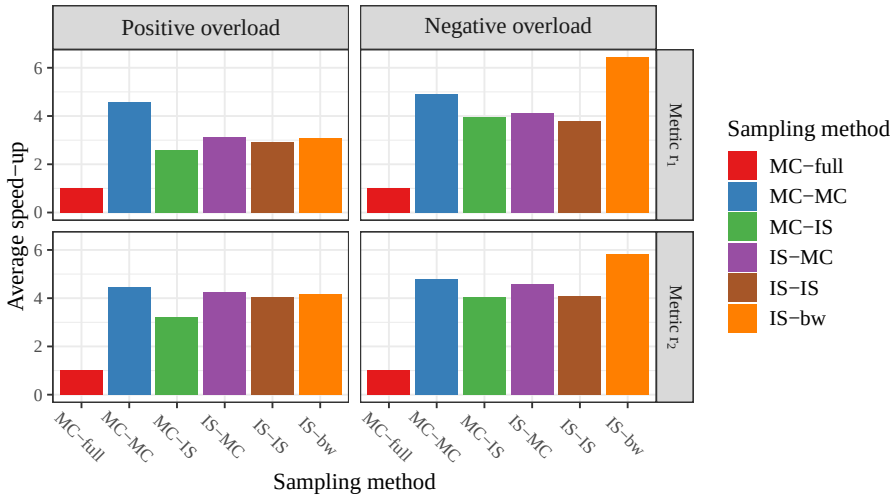


**Figure A.2:** Comparison of the average speed-up for all methods with respect to the reference method MC-full. Also the estimates whose accuracy was found to differ significantly from the estimates of the reference method were used to compute the average speed-ups for this figure. This is in contrast to the figure shown in the main body of the report. Estimates with zero probability were not included in the average speed-up computation.