Long-term planning of low voltage networks

Document status and date:
Published: 25/10/2017

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 08. Jul. 2024

Long-term Planning of Low Voltage Networks

Michiel Nijhuis

Long-term Planning of Low Voltage Networks

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

Voorzitter: prof.dr.ir A.B. Smulders
Promotor: prof. dr. ir. J.F.G. Cobben
Copromotor: dr. M. Gibescu
Leden: dr. P.M. Carvalho   (Instituto Superior Técnico)
                prof.dr. L.M. Nordström   (KTH – Royal Institute of Technology)
                prof.dr. P. Palensky   (Technische Universiteit Delft)
                prof.dr.ir. A.J.M. Pemen
                prof.dr.ir. J.G. Slootweg

Data shown on cover:
Topological data: ©Kadaster, 2017
Network data: Enexis open data, 2017

Summary

Long-term Planning of Low Voltage Networks

The energy system is in a transition from a fossil-fuel based towards a sustainable energy supply. The uptake of clean energy technologies like electric cars, heat pumps and photovoltaics has been rising in the past years and expected to continue to do so in the future. This brings challenges for the low voltage (LV) network, as these technologies are mainly connected at the household level. The scope of these challenges is dependent on the ability of the LV-network to handle these changes. This thesis focuses on the determination of the quality of the LV-network in light of this energy transition and the reinforcement of the LV-network to mitigate the problems that may arise.

First of all, the effects of the transition towards a sustainable energy supply on the LV-network have been mapped based on a literature review. The most important effects for the LV-network are the anticipated changes in power quality and the capacity of the LV-network, introduced by the increased penetration of both photovoltaics and electric vehicles. The loading of the network, therefore, needs to be modelled, with the inclusion of the technologies which govern the transition towards a sustainable energy supply. A bottom-up Markov Chain Monte Carlo household load model has been defined that can handle scenario inputs to model the load for the coming years. The probability density functions of the highly volatile household load can be accurately determined with the defined model.

The analysis of the LV-network is predominantly done by applying load flow calculations. The volatility of the household load calls for a more probabilistic approach to the calculation of the load flow. As the characteristics of the LV-network allow for the simplifications of the load flow calculations, a new probabilistic load flow algorithm specifically for LV-networks has been developed. Next to this, the integration of the requirements of the LV-network into a risk-based analysis framework has been performed. Based on this a single risk indicator can be determined for both current and new networks.

Based on the developed assessment frameworks and the current methods of assessing the adequacy of the LV-network of the Dutch DNO Liander is determined. The foreseen changes brought by the transition towards a sustainable energy supply are included in this assessment. Most of the LV-network does not run into any problems, however 36%

of the LV-network might require some sort of reinforcement based on all the analysed scenarios. From the level of adequacy in the network, the most important risks for the network are defined: the penetration of electric vehicles and photovoltaics> The LV-network characteristic that is most correlated with a high risk is the number of customers connected to the LV-feeder.

From the analysis of the risks within the LV-network, it becomes apparent that parts of the LV-network will have to be reinforced in order to handle the changes introduced by the transition towards a sustainable energy supply. The implementation of the volatility of the household load, as well as the uncertainty about the future loading in the LV-network expansion planning, has been analysed. Through the use of probabilistic load flows and scenario reduction, both these types of uncertainty can be included in the optimisation of the LV-network reinforcements. The inclusion of these uncertainties leads however to a computational intensive optimisation procedure. The genetic algorithm which is used for the optimisation of the network reinforcements has therefore been adjusted to a bilevel optimisation problem. By focussing more on the practical application of the LV-network expansion planning simplifications can be made within this bilevel genetic algorithm to allow for a computationally feasible optimisation. The use of demand response has been investigated as an alternative to conventional network reinforcements. The limited benefits, for all involved parties, make the application of demand response from an LV-network planning perspective not attractive.

Contents

1

Introduction

The energy system is in a transition from a fossil fuel based energy supply towards a sustainable energy supply. The generation of usable energy is increasingly shifting towards renewable energy, like photovoltaics (PV), while the end-use is shifting towards clean energy technologies based on electricity like electric cars (EV) and heat pumps. The uptake of these technologies has been rising in the past years and expected to continue to do so in the future. In Table 1.1 the number of EVs, heat pumps and installed capacity of PV is given for the years 2009-2014 in the EU [1]. From the table, it can be seen that the penetration of these technologies is increasing year after year. The rate at which these technologies are being implemented slowly increases for most years as well, displaying an acceleration in the adoption rate. These technologies will therefore, have an important role to play in the energy system of tomorrow.

Table 1.1: *The percentage of electric cars over the total amount of cars, heat pumps over the total amount of heating systems, and photovoltaics over the total electricity consumptions in the EU over the years 2009-2014 [1]*

	2009	2010	2011	2012	2013	2014
Electric cars	0.005%	0.005%	0.010%	0.014%	0.030%	0.052%
Heat pump	0.028%	0.031%	0.043%	0.047%	0.055%	0.064%
PV	0.10%	0.15%	0.31%	0.46%	0.54%	0.63%

Most of these technologies are connected to the electricity network at or near the premises

of residential customers. These appliances are considerably different from the current appliances predominately connected to the low voltage (LV) network. A large increase in electrical energy usage and/or the emergence of reverse energy flows are associated with the transition towards a sustainable energy supply. Most of the LV-networks in western countries are originally built more than 40 years ago and not designed for these changes. Reinforcing the LV-network requires investments in assets which have an economic lifetime of multiple decades. Taking only one or a few of these technologies into account when altering the LV-network generates the risk of not meeting the requirements for the coming decades but only for the anticipated developments. The current design paradigm followed by most distribution network operators (DNO) is aimed at creating a network which can last for at least forty years. This is illustrated in Fig. 1.1, where the date of construction of the LV-cables from the Dutch DNO Enexis is given.
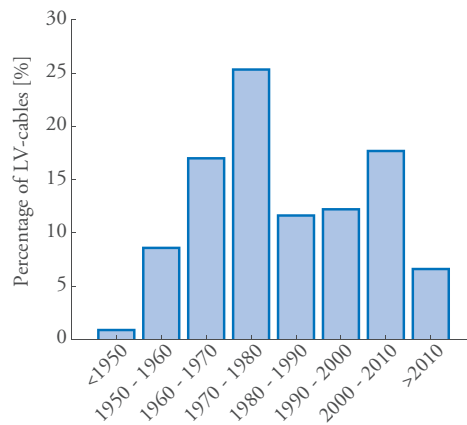
Figure 1.1: *Year of construction of the LV-cables as percentage of the total length of the LV-cables [2]*

From the figure, it can be seen that the majority of the LV-cables is at least 35 years old, with more than 20% already being older than 50 years. This is an indication that a very robust approach is applied to the design of the LV-network. A robust approach for the LV-network planning is used as there is very limited data available. The maximum loading of the LV-network is for instance, in most cases only measured once a year at the MV/LV transformer, and the loading of a household thirty years from now is almost impossible to predict. Therefore, LV-networks have a capacity far greater than they require in the coming years to account for all these uncertainties. This robust approach has made it possible for the LV-network to handle significant load increases without generating capacity constraints, however it also led to an LV-network which is much stronger than required in many places. Raising the question, whether the over-dimensioning of the LV-network is the most cost-effective approach.

The conditions in which the reinforcement of the LV-network occur are characterised

by capital-intensive investments, high uncertainty and a negative asset recovery value for most assets as large parts of the LV-networks in Europe consist of underground cables [3]. The uncertainties which are present when planning the LV-network are high. Typically, only one or two global scenarios for the coming decades are used to evaluate whether a certain design is strong enough. The local circumstances can vary greatly from a national scenario, which means only limited optimisation is possible because over-dimensioning is still necessary to account for the uncertainties surrounding the predictions of future loading and the implementation at the local level. The optimisation of the decision making in the LV-network expansion planning process is therefore of utmost importance as it is a major contributor to the economic effectiveness of a DNO. Due to the nature of the system, creating a design paradigm which will (to some extent) remove the need for over-dimensioning is hard without creating huge risks of the LV-network failing to comply with one of the requirements. The question remains, however, whether the current design paradigm comes close to an optimal level of robustness to be able to handle the future requirements, or if there is room for improvement of the current process. Next to this, there have been technological advancements which would also require a review of the long-term planning methodology for the LV-network.

1.1 Research questions

The changes in electricity usage combined with the fact that current LV-networks are ageing give rise to the question of how the replacement of these assets can best be performed. This leads to the main research question:

> *How should the distribution network operator change its approach to the design of the low voltage network in order to most efficiently handle the uncertainties in future loading introduced by the transition towards a sustainable energy supply?*

In order to answer this research question, the question is divided into a number of sub-questions. The first sub-question which arises when trying to answer the above research question has to do with the changes in the future loading of the LV-network. Whether or not the current approach to the design of the LV-network is capable of efficiently handling the changes introduced by the transition towards a sustainable energy supply needs to be determined. First, the changes which would have an effect on the LV-network should be investigated. So not only do these foreseen changes need to be characterised, but these changes should also be implemented in models capable of assessing the adequacy of the LV-network. Therefore the following sub-question has been drafted:

> 1) *Which of the changes introduced by the transition towards a sustainable energy supply has an effect on the long-term adequacy of the low voltage network?*

With the changes introduced by the transition towards a sustainable energy supply characterised, the effects of this transition on the LV-network need to be determined. If these changes have been determined, the next question which arises is how these changes can be implemented in the modelling of the residential load in order to generate the time evolution of the loading of the LV-network over the coming decades. This will be assessed by the following research question:

> 2) *How will the residential load change in the coming decades and how can this be modelled?*

Knowing how the residential load will change over the coming decades is an important step towards determining the adequacy of the current network. To be able to determine the network adequacy, the metrics which are necessary and the methods to calculate them are assessed based on research question number 3:

> 3) *Which metrics are important for the determination of the adequacy of the low voltage network and how can they be assessed?*

With the assessment method determined and the effects of the transition towards a sustainable energy supply implemented in the residential load, the current state of the LV-network and whether it can handle the future load changes can be assessed. This is done to answer the following sub-question:

> 4) *Is the current (Dutch) low-voltage network capable of handling the foreseen changes in the loading?*

The changes which are introduced by the transition towards a sustainable energy supply do not only jeopardise the adequacy of the network. Through the increase of controllable loads, power electronic interfaces and ICT, other options to ensure the adequate operation of the LV-network than the conventional ones become available to the DNO. These options should be assessed in combination with the currently acceptable options to reinforce the LV-network (e.g. adding or increasing the capacity of feeders and substations). In order to assess this, the following research question has been defined:

> 5) *How can low voltage network reinforcements best be implemented and can 'smart grid' solutions contribute to a more cost-effective low voltage network reinforcement?*

1.2 This thesis

The research questions posed in the previous section are addressed in the five main chapters of this thesis. In each of the chapters, one of the sub-questions is assessed. The chapters are organised as follows.

Chapter 2: The Low Voltage Network

In the second chapter, background information will be given to gain insight into the effects of the transition towards a sustainable energy supply on the LV-network. First of all, an overview of the distribution network and the main characteristics of the LV-network in particular, will be given. Within this overview attention will also be given to the requirements of the LV-network, to gain a better understanding of the main aspects the DNO should take into account when designing the LV-network. This chapter is finalised with a discussion of the main changes introduced by the transition towards a sustainable energy supply and how these changes can be characterised in a manner which is meaningful from an LV-network perspective, to give an answer to the first sub-question.

Chapter 3: Residential Load Modelling for LV-Network Assessment

The residential load (and generation) is one of the most important inputs for the assessment of the LV-network. The residential load is also one of the main system parameters affected by the transition towards a sustainable energy supply. The implementation of the characteristics of the sustainable energy transition (as defined in Chapter 2) into a model for the residential load, is discussed in this section. With the characterisation of the transition towards a sustainable energy supply and its implementation in the load model, the second sub-question can be answered.

Chapter 4: The Assessment of the LV-network

The analysis of the LV-network can be done in multiple ways, depending on the goal of the analysis. In this chapter several approaches to analyse the LV-network are given. The determination of the adequacy of the LV-network is preferably done based on a single indicator. The LV-network has to adhere to multiple requirements, however, to make the results on the performance of the network usable, a single metric will be defined from the long-term adequacy perspective. This should provide the results necessary for answering the third sub-question.

Chapter 5: The Adequacy of the LV-Network: A Dutch Case Study

With the assessment methods known and the metric to determine the adequacy defined, a closer look can be taken at the state of the current LV-network. The adequacy as defined in Chapter 4 is assessed for the LV-network originating from a Dutch DNO, taking into account the changes introduced by the transition towards a sustainable energy supply over the coming decades, as defined in chapter 2 & 3. Based on this analysis of the LV-network the importance of the various drivers of the transition towards a sustainable energy supply can be determined, giving an answer to sub-question four.

Chapter 6: The Reinforcement of the LV-Network

As Chapter 5 shows that not all parts of the LV-network are capable of handling the changes introduced by the transition towards a sustainable energy supply, the method for reinforcing the LV-network will have to be assessed. The current method for reinforcing the LV-network lacks in certain key areas, so a new bilevel LV-network expansion planning approach is defined. Thereafter, the implementation of the most promising 'smart grid' technologies: on-load tap changer (OLTC) control and demand response in the planning process, is also presented to answer the fifth sub-question.

With the discussion of all these topics fished, the thesis will be concluded with a review of the research sub-questions posed in section 1.1. Through this review, an answer to the main research question is formulated and the conclusions which can be drawn from it are given in the final chapter of this thesis.

2

The Low Voltage Network

In order to start answering the question how a DNO should change its approach to the design of the LV-network, a better understanding of the problem is required. In this chapter, the LV-network and its functional requirements are introduced. This is followed by a discussion on how the LV-network is affected by the transition towards a sustainable energy supply. First of all, an overview of the current electrical transmission and distribution network will be given, with a focus on the LV-network, followed by a discussion on the requirements the LV-network must adhere to. With this knowledge, the transition towards a sustainable energy supply can be characterised with the aid of scenarios, keeping in mind the expected impact of these scenarios on the loading of the LV-network, and hence on its functional requirements.

2.1 LV-networks: An introduction

To be able to analyse the LV-network, first a better understanding of this type of network is required. Therefore the LV-network and its place in the power system are characterised, mainly from the perspective of the current LV-network. When a better understanding of the characteristics of the LV-network is gained, the possible assumptions and most important parameters when it comes to modelling and analysing the LV-network will be discussed. Next to a discussion of the current LV-network, the requirements of the LV-network which have resulted in the creation of the current system are discussed. These requirements are of

importance when determining how the current approach to designing the LV-network can be changed, as the requirements the LV-network has to adhere to are assumed to remain constant throughout this research.

2.1.1 The LV-network as part of the power system

The power system is designed to transport electrical energy from the generators to the loads. The power system can be divided into two parts: the transmission network (HV-network), a system with high voltages designed for the bulk transport and exchange of electrical energy, with large generating plants and a few large customers connected, and the distribution network, a vast network originally designed to supply electrical energy to the many smaller consumers. The distribution network, in turn, can be split into two parts, the medium voltage MV-network, an intermediate network for the local transport of electrical energy with mid-sized consumers connected (0.3 MVA - 10 MVA) and the low voltage LV-network, the last meters to the many small consumers (< 0.3 MVA). To illustrate the differences between these networks Fig. 2.1, 2.2 and 2.3 have been created as well as Table 2.1.
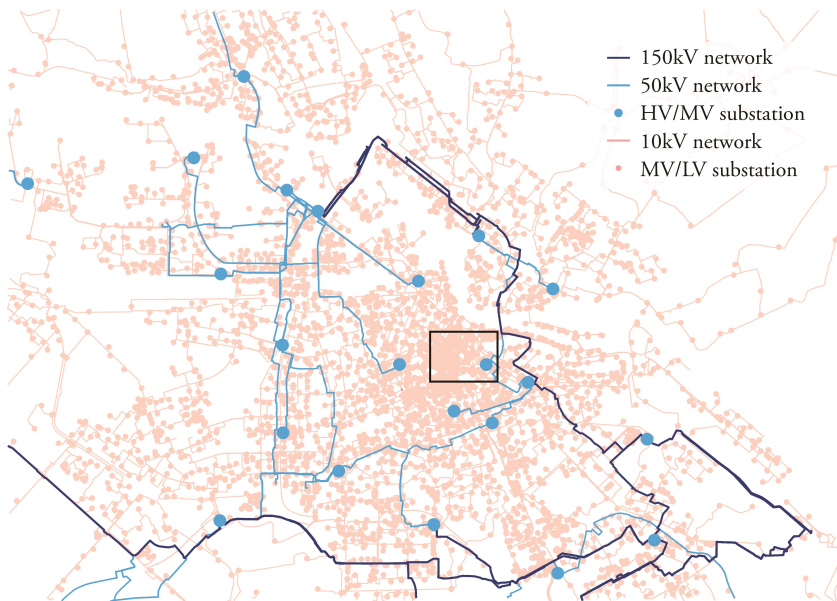
Figure 2.1: *Overview of part of the high voltage network in the Netherlands*

In the Fig. 2.1 part of the Dutch 150 kV HV-network and the MV-transmission network (50 kV) are depicted. The underlying 10 kV MV-distribution network is also included. From the figure, the difference between the transport network and the 10 kV network becomes apparent. For just 22 HV/MV substations the number of 10 kV stations is already in the
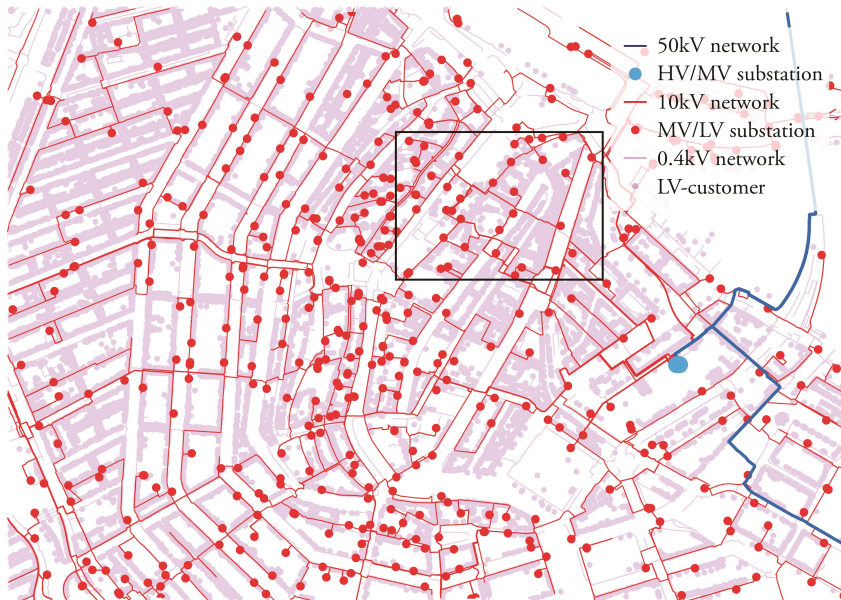
Figure 2.2: *Overview of part of the medium voltage network in the Netherlands*

thousands. The black box indicated in the figure is the area shown in Fig. 2.2. In this figure, the 10 kV network is highlighted with a single HV/MV substation. In the figure also the LV-network is included. The figure shows that even though the MV/LV substations are already plentiful, the number of connection points at the LV-network is still much greater than for the MV-network. In Fig. 2.3 the LV-network indicated by the black box in Fig. 2.2 is shown in detail. As the most consumers connected to the LV-network are household consumers, the location of the houses can already be determined from the network. The street pattern can also be observed from Fig. 2.3, as the LV-cables are mostly placed underneath the pavement next to the streets. In Table 2.1 an overview of some of the most important characteristics of these different networks is given [4]. From this, as well as from all the three figures, it becomes clear that the analysis of the LV-network should differ significantly from the analysis of the transmission network or even the MV-network. Bear in mind that the LV-network has more than 99.6% of the connections and more than 65% of the total cable length.

In addition to these large difference in size and number of connections compared to higher voltage networks, the LV-network also has other unique characteristics. These characteristics will lead to differences in the analysis of the LV-network. First of all, as the LV-network consists of such a large number of nodes, the network is often not monitored. Measurements of the maximum current are usually performed at the MV/LV transformer level and only collected yearly. The household connections for a large part still have meters without communication possibilities which are surveyed manually. The actual state of the LV-
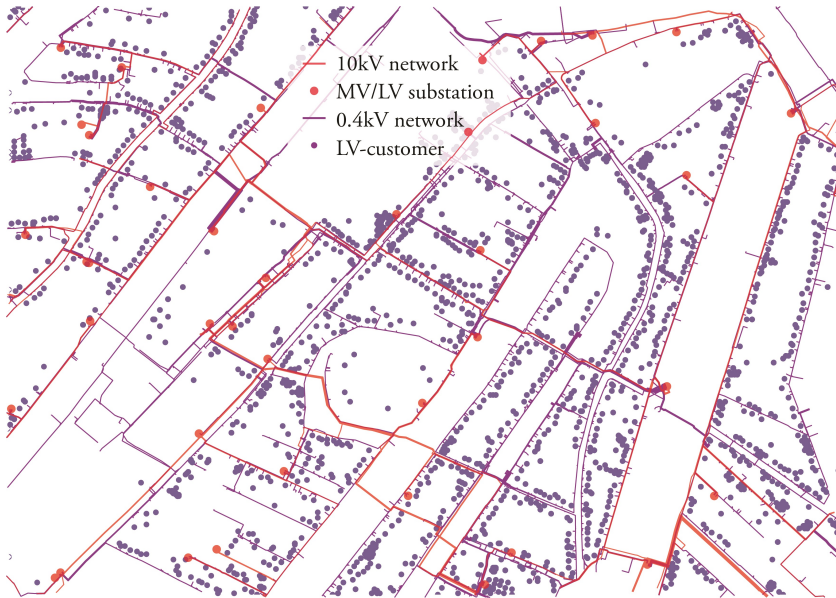
Figure 2.3: *Overview of part of the low voltage network in the Netherlands*

Table 2.1: *Number of connections and lines lengths for different voltage levels in the Dutch power system [4]*

	Voltage	Length [km]	Connections
Extra high voltage	220/380kV	2 760	99
High voltage	50/110/150kV	9 037	158
Medium voltage	3 t/m 25 kV	103 862	30 595
Low voltage	0,4 kV	223 511	8 073 554

network is, therefore, to a large extent unknown. With the current introduction of an advanced metering infrastructure, the observability of the network is also starting to improve. Secondly, much of the LV-network is purely radial in nature. The MV-network tends to be meshed, but radially operated, while the HV-network is meshed and also operated in a meshed configuration. The radial nature of the LV-network simplifies the analysis considerably as the direction of the flow of the current can be easily estimated. Next to the radial nature, the LV-network in European countries consists mainly of (underground) cables [3], resulting in much higher R/X ratios than for the MV and HV-network. This means the resistance is the most important factor determining the voltage and that the voltage angle is fairly constant throughout the LV-network.

The next important attribute of the LV-network has to do with the manner of reinforcing the

network. The LV-network needs to connect many consumers, all located in a relatively small area. The investments needed to reinforce the LV-network tend to be small, as problems can be solved locally instead of requiring an upgrade of a substantial part of the network.

The last important characteristic of the LV-network is that most connected customers are households. These household connections have a highly volatile nature of consumption. This is illustrated in Fig. 2.4.
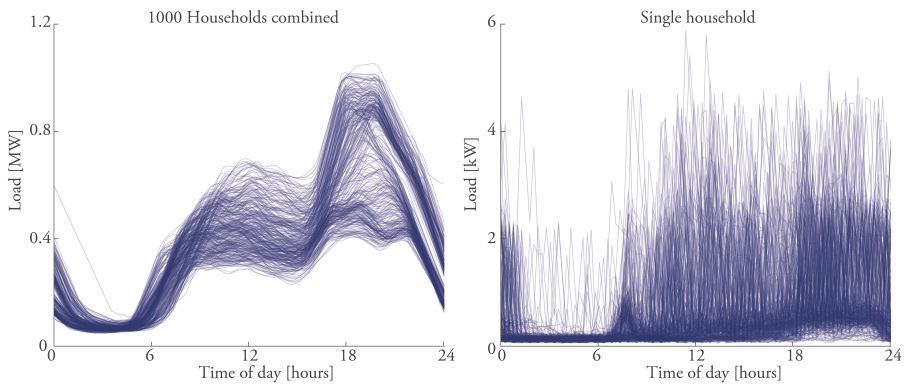
Figure 2.4: *The measured load curve for each day in a year for 1000 households combined (scaled from the reconciliation data [5]) and for a single household*

In the figure, the measured household load for each day in a year is plotted for both a single household (right figure) and for 1000 aggregated households (left figure). The load curve for the 1000 households is based on a scaled version of the load data from the reconciliation process [5] and the for the single household based on the measurements from a smart meter. From an MV-network perspective, a high level of aggregation would be visible, since on average about 200-400 households are connected to a single MV/LV substation in non-rural areas. This level of aggregation can already lead to a smooth load curve which follows a clear daily pattern, with a difference between the maximum and minimum loading at a single time of day not exceeding the 300%. From an LV-perspective a single household needs to be considered. In the right part of Fig. 2.4 a number of different times can be discerned: the night, the morning peak (around 7 AM), the day and the evening (from 6 PM). The differences between days can, however differ significantly, even when looking at the same time of day the difference in the load between consecutive days can be over 4000%. The modelling of the load and the analysis of the LV-network should thus take into account this high level of load volatility.

2.1.2 Requirements

In order to discuss the adequacy of the LV-network with respect to the changes which will occur in the future, the functional requirements of the LV-network should be known. An overview of these requirements is given in this section. The basic function of the LV-network is to provide customers uninterrupted access to safe electrical power of an adequate quality at the lowest possible cost. The function of the LV-network, can be expected to remain the same over the coming years and the resulting functional requirements can thus also be considered constant for the coming years. The requirements which can be derived from the main function of the distribution network (as with most other critical infrastructures) are usually classified based on the four A's: *availability, affordability, accessibility* and *acceptability* [6]. To allow for a more detailed and structured evaluation of the developments in the distribution network, for the purposes of this theses, the four cardinal requirements are sub-divided into eight categories: availability, capacity, cost, regulation, power quality, social context, safety and market facilitation. These eight requirements are discussed in more detail below.

Availability

An uninterrupted voltage at the point of connection with the customer is one of the most important goals of the DNO. Situations, when the voltage is interrupted, can be subdivided in planned and unplanned unavailability. The planned unavailability is mainly due to maintenance and can be kept to a minimum by local island operation whilst a higher network part is being maintained, or by increasing the possibilities of doing maintenance work live. For the design of the distribution network, the unplanned unavailability is the determining factor, which may be caused by component failure and overloading. For the DNO the chance of failure depends on the number of components between the point of connection and the high voltage/medium voltage (HV/MV) substation and the reliability of those components. From a consumer perspective, the availability is not only dependent on the distribution network, but also on their own installation and on upstream the transmission network.

Network Capacity

The required capacity of the network must be sufficient so that any user can draw their contracted current from the network at any time. As the coincidence of individual load profiles is low, especially for households, the actual network capacity is far lower than the summation of the individual contracted connection capacities.

Cost

The capital and operational costs of the DNO should be as low as possible, whilst fulfilling the other requirements. The cost for the DNO, excluding the overhead, can be split into four categories: expansion/replacement cost, impairment cost, energy losses and maintenance. From a consumer perspective, these costs should be optimised based on the lowest social cost (network cost plus economic externalities), while the DNO would like to minimise only its own costs. The main challenge associated with finding this optimum is the payback period of 20 to 40 years for investments and a technical life of assets often exceeding 60 years. As the uncertainty whether an asset will still be sufficient increases, with the transition towards a sustainable energy supply, it becomes more difficult to assess the investment options.

Regulation

The distribution network has two properties which make it hard for a competitive market to exist. The distribution network is characterised by its capital cost dependent nature (long payback periods and low operational cost) and its non-rivalry nature (access cost are to a large degree dependent on whether the good is consumed within the vicinity). This marks the distribution network as a *natural monopoly*. To keep the DNO from achieving supernormal profits as well as ensuring a high enough quality is delivered by the network operator, regulation is put in place. The DNO can however still take the role of market facilitator in combination within a local (retail) energy market, similar to the facilitation of the wholesale market done by the transmission system operator.

Power Quality

It is important that the voltage at the point of connection is of sufficient quality to ensure its compatibility with the equipment installed on the customer premises. The voltage quality can be defined by a number of indicators: harmonic distortion, asymmetry, transients, frequency level, flicker, voltage swells/dips and steady-state voltage level. The voltage quality is affected by both the network and the installations present at the customer site. However, not all these phenomena are considered and should be considered during the design of the distribution network. Harmonic voltage distortion in the network is induced by non-linear loads and affected by the harmonic impedance of the network. The mitigation options to decrease the harmonic distortion are usually either an improvement with respect to the harmonics currents produced by the load itself or the use of active or passive filters. Both of these options are applied locally and changes in the network are usually not considered; for this reason, the harmonics are not considered during the design of the distribution network. An imbalance between the phases is present, as households loads are mostly single phase loads. An imbalance arises if the loads are not equally distributed over the three phases. For the design of the distribution network, the imbalance is not considered as the distribution

of the load over the phases could be adjusted. The current policy to construct household connections is to lay the cables for all three phases and only connect the number of phases required by the customer. Voltage dips and swells are power quality phenomena which normally only occur during a short-circuit in the network. During the network design, other requirements like safety and availability are the driving factor which determines the expected magnitude and duration of dips and swells. The possible increase of RES in the distribution network can have a positive effect on the magnitude of the voltage dips, however (in unbundled power systems) the DNO is not allowed to place/operate RES and has little possibility to maximise the reduction in dip magnitude because of RES. There is no guideline given for the amount and magnitude of swells in the Dutch network code and the regulation on dips is limited to voltage levels of 35 kV and higher (though this will most likely be extended in the coming years). The DNO has thus a very limited incentive to improve its service with respect to dips, while the social cost may be high. Transients in the distribution network can be caused by the energising of a cable, capacitor or transformer or by the customer's own installations (the most common source of transients, lightning, is uncommon in distribution networks where all cables are buried). The transient propagation is dependent on the network design, however as it is not yet known to what extent transients cause problems and how often these problems occur in the distribution network, transients are not considered during the design phase of a distribution network. The frequency is controlled in the transmission network and thus considered a transmission system operator problem. The main power quality aspect which should be taken into account when reinforcing the LV-network is the (steady-state) voltage level, therefore this also gains the focus when assessing the developments in the distribution network.

Social Context

The DNO needs to operate within the context given by its service area. This social context gives a number of external boundary conditions with respect to the landscape and the distribution of the customers and with respect to the legislation. Legislation and governmental policies can determine to a high degree how the distribution network will be designed. The policies with respect to the built environment in the Netherlands determines that the assets of the distribution network should be hidden from view, which has resulted in the burying of almost all the low and medium voltage cables. Legislation with respect to the environment and safety can for a large part determine whether a certain solution is feasible or not. The location of the customers is another aspect which for a large part is given. As the DNO is a monopolist everyone has to be connected if the cost for the connection is within reason, therefore usually shallow connection cost is applied (only if the cost becomes uneconomical, the DNO can charge additional money to upgrade other parts of the network; deep connection cost). The DNO is thus obliged to adjust to the local geographic characteristics. The DNO is also dependent on the availability of qualified workers. The amount of work needed for certain network expansion options can differ greatly and this could have a large influence on whether a certain alternative is viable or not.

Safety

The distribution of electricity should come with the safe operation of the network for the DNO as well as for the end-users, this predominantly means the provision of adequate earthing. For low voltage (LV) networks the TN-earthing (the star point of the transformer is grounded and a connection is made between this star point and the installation of the customer) system is utilised in most European countries. This means that the DNO is responsible for providing an earth connection with a low enough impedance to ensure safe operation, during phase-to-ground faults. Another aspect of safety with regards to the distribution network is adequate fault clearing. If there is a short-circuit somewhere in the network, the current which flows will be large, leading to overheating of the components. To ensure that these components do not get damaged by the short-circuit current, they roughly need to be switched off within 5 seconds. As safety has little influence from the perspective of the loading of the LV-network, it is usually not taken into account when reinforcing the network.

Market Facilitation

The electricity markets in Europe and much of the developed world are liberalised markets, where only the distribution and the transmission network operators have a monopoly position. In order to have a functioning competitive electricity market, power exchange between producers, suppliers, and end-users should be possible. The distribution and the transmission network operators facilitate this marketplace by physically connecting these different market parties. From an end-user perspective, the facilitating of the electricity market implies the freedom to choose the electricity supplier best suited to their personal preference. The advantages with respect to supply-demand balancing and economies of scale have led to the creation of a completely interconnected network within Europe. At the moment market facilitation for the DNO is limited to measuring the amount of energy drawn from the LV and MV-networks by end-users. The DNO is a very large user of electricity (through the energy losses), this in combination with its knowledge of supply and demand and its possibility to influence the demand of other consumers, gives the DNO significant market power. To counter the use of its market power the DNO is excluded from the owning electrical energy production facilities.

2.2 Characterisation of the transition towards a sustainable energy supply

The transition towards a sustainable energy supply needs to be framed from the perspective of the long-term adequacy of the LV-network. The main influence of this transition from an LV-network perspective is on the residential electricity usage. The electrification of

transportation and heating can shift the energy usage of households away from fossil fuels and more towards electricity. At the same time, through locally generated renewable energy like PV, the residential consumers are becoming prosumers. The residential electricity usage is thus changing significantly through the transition towards a sustainable energy supply, both in terms of total electrical energy demand, and in terms of the time-varying profiles of this demand. Ensuring adequate network capacity and power quality are the main requirements the DNO should focus on to facilitate this transition[1]. The technologies which have the most influence on the LV-network are considered to be PV (photovoltaics), EV (electric vehicles), micro-CHP (micro combined heat and power) and heat pumps. The application of storage and power electronics within the network can also have a large effect on the LV-network. The cost of these technologies still inhibits their application in the LV-network and for the near future is not expected to decrease sufficiently. Storage and power electronics are therefore not taken into account in this thesis. The technology adoption scenarios of these technologies are modelled through the use of S-curves:

$$y = \frac{y_{max}}{1 + e^{-a(t-t_0)}} \tag{2.1}$$

in which y_{max} is the maximum penetration level, a determines the rate of adoption of a technology/appliance and t_0 is the year in which half of the maximum penetration is achieved. In this thesis, the values given in Table 2.2 are used for the generation of the scenarios. The values of these scenarios are based on a number of reports which focus on the forecasting of one or more of the different scenario elements [7, 8, 9, 10, 11, 12, 13, 14]. This approach is applied for all the scenario drivers. In Table 2.2 an overview of the resulting scenarios is given.

Table 2.2: *Overview of the different parts of a scenario and how they are implemented*

	High			**Medium**			**Low**		
	y_{max}	a	t_0	y_{max}	a	t_0	y_{max}	a	t_0
PV	63	0.25	2029	29	0.23	2028	4.6	0.15	2027
EV	84	0.25	2031	33	0.23	2031	7.2	0.18	2028
Heat pump	46	0.31	2034	5.2	0.31	2034	0.1	0.29	2033
micro-CHP	9.7	0.24	2031	2.7	0.24	2030	0.2	0.023	2031

Next to the increasing application of new technologies for heating, transportation, and self-generation of electricity, the residential load is also expected to change irrespective of the transition towards a sustainable energy system. This standard change in residential electricity

---

[1]A detailed discussion of the effects of the transition towards a sustainable energy system appendix can be found in A.1 and in the publication :

M. Nijhuis, M. Gibescu and J.F.G. Cobben, "Assessment of the impacts of the renewable energy and ICT driven energy transition on distribution networks", *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1003-1014, Dec 2015

usage should also be incorporated by the use of scenarios. Scenarios on the economic growth [15] and the change in efficiency of different groups of appliances (e.g. electronics, heating loads, etc.) [16] are also used. This leads to the scenarios for the individual drivers as depicted in Fig.2.5. These scenarios all have an effect on the household load. In the next chapter, in section 3.2.1, the method for implementing these scenarios into a household load model is discussed. In the remainder of this thesis, various combinations of these individual scenarios will be used to estimate the long-term load evolution.
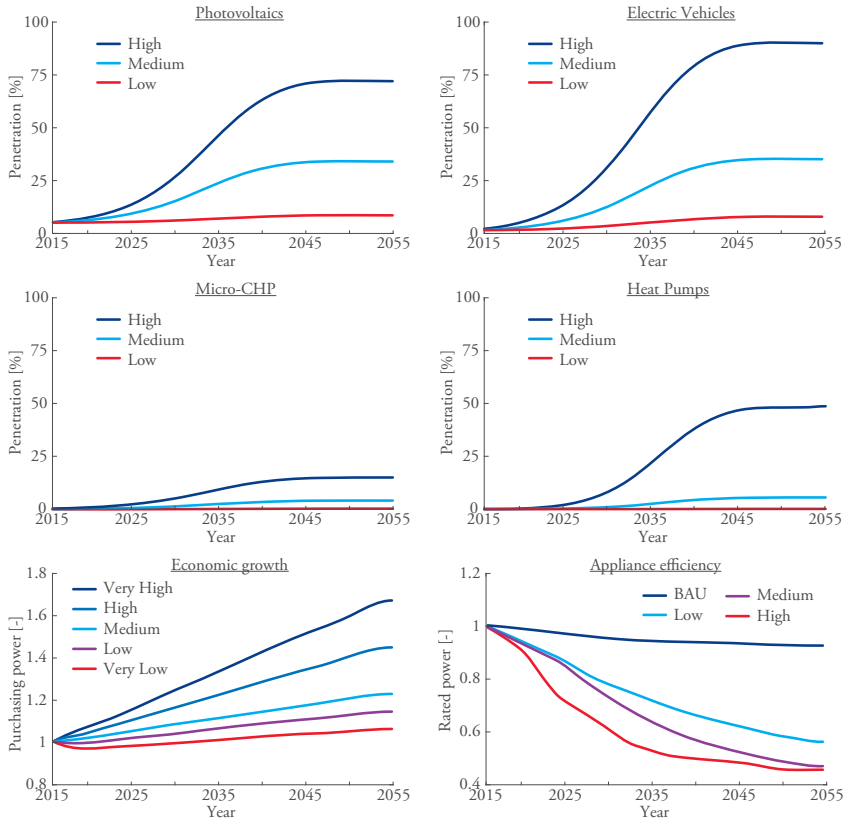
Figure 2.5: *Overview of the individual scenarios used in this thesis*

2.3 Conclusions

The most important technical characteristics of the LV-network include the radial structure with a high R/X ratio. The loads connected to the LV-network consist mostly of households,

which have a high load volatility. The change in the household load, through the electrification of transportation in combination with locally generated renewable energy like PV, can be considered the main change the transition towards a sustainable energy supply brings to the LV-network. The power quality and the capacity of the LV-network are the requirements that are affected the most through these changes in the residential load.

3

Residential Load Modelling for LV-Network Assessment

One of the most important inputs for the planning of LV-networks is the household load. LV-networks consist mostly of residential consumers. In the Netherlands, 94.6% of the connections are household connections[1], the other 5.4% are mainly small and medium-sized businesses. These residential loads are all connected to the current infrastructure. Changes in this infrastructure due to the building of new neighbourhoods is a small aspect of the developments within the LV-network. In most developed countries the renewal of the building stock tends to be quite low. For instance in the Netherlands on average the building of brand new dwellings only consists of 0.006% of the total building stock. The assessment of the LV-network would thus mainly be based on the reinforcement of the current infrastructure and not based on any expansion plans. To determine whether the

---

This chapter is based on the following publications:

M. Nijhuis, M. Gibescu and J.F.G. Cobben, "Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data", *Energy and Buildings*, vol. 112, pp. 121-129, Jan 2016

M. Nijhuis, R. Bernards M. Gibescu and J.F.G. Cobben, "Stochastic Household Load Modelling From a Smart Grid Planning Perspective", in *EnergyCon*, April 2016

M. Nijhuis, B.M.J. Vonk, M. Gibescu, J.F.G. Cobben and J.G. Slootweg, "Assessment of Probabilistic Methods for Simulating Household Load Patterns in Distribution Grids", in *International Conference and Exhibition on Electricity Distribution (CIRED)*, June 2015

[1] Based on data Liander LV-connections in Gelderland, 2010, and a number of households in Gelderland [17]

current LV-network remains adequate for the coming years it is essential that the change in the loading of the network elements and thus the change in the residential load is adequately modelled.

Modelling of the current and future household load remains necessary for the assessment of the LV-network. Though in the coming years every household will be equipped with a smart meter, which can send data about the household load profile to the DNO, the modelling of the household is still required for two reasons. First of all, the household load is very dependent on the behaviour of the members of the household. Similar households with similar economic backgrounds, household sizes, ages and dwelling size and type can have widely varying energy usage[2]. Through this dependence, changes in the occupants of a dwelling can have large effects on the adequacy of the LV network. Secondly, investments in the low voltage network are done based on investment horizons of 40 years or even longer. The assessment of the loading of the LV network solely based on current measurements would be insufficient, since the network should remain adequate for the foreseeable future as well.

The modelling of the residential load can be performed in multiple ways with multiple levels of accuracy and aggregation depending on the analysis which needs to be performed on the LV network. Detailed modelling can generate a more accurate picture of the loading of the LV network, however, a detailed model would require more computational power. To develop models which can balance accuracy and computational complexity it is important to define the requirements for the modelling of the household load based on the type of analysis which has to be performed in the LV-network. Based on these requirements, the modelling of the household can be discussed in more detail, followed by a description of the approach adopted for the modelling of the household loads in this thesis. Next, a detailed model will be defined followed by simplifications which can generate more computationally efficient residential load modelling.

3.1 Requirements for the modelling of household load

To determine how the residential load should be modelled, first of all, the requirements which arise from the subsequent analysis of the LV-network should be defined. Currently, the household load is modelled as a peak power consumed and a peak power produced per household, in combination with a coincidence factor. To obtain the load at some point in the future the current load is multiplied with a constant growth factor [19]. For the current expansion/reinforcement strategy, consisting of applying larger diameter cables, new LV-feeders or new MV/LV substations, this method remains sufficient.

With the advances in ICT, storage technology and power electronics, new options become

---

[2]Within the same neighbourhood the average yearly energy use can have a standard deviation of 2.8 MWh with a mean of 2.2 MWh [18]

available to ensure adequate network operation without using conventional network reinforcement options. These so-called *smart grid* alternatives require a different kind of load modelling. Next to the changes in reinforcement options, also the uncertainty about the future loading of the network is large. The pace of implementation of technologies like EV and PV can have large effects on the loading of the LV-network. This cannot be captured by the use of a constant load growth rate.

In light of these changes, the requirements with respect to the modelling of the residential load have to be redefined. In order to define the requirements, first of all the statistical properties of the residential load will be evaluated in more detail. Next, the implementation of future changes in the loading is discussed, followed by a discussion of the computational requirements which stem from the different kinds of analysis which have to be performed on the LV-network.

3.1.1 The residential load curve

A model of the residential load should generate load curves for households which resemble the actual load curves as much as possible. Depending on the purpose of the residential load model, the requirements may differ. For instance when it comes to transformer sizing only the aggregated amount of household load is of importance, while when it comes to assessing the voltage limitations within the network, the load during the peak hours at each household becomes important, while for assessing the possibilities of load shifting, a time series would be required. To determine how these different requirements can be introduced in the modelling of the household load, a closer look needs to be taken at the properties of the household load.

To illustrate the statistical properties of the household load curve, the probability density function of the load, measured with a fifteen-minute resolution during a 24-hour weekday, based on a data set from 200 smart meters is depicted in Fig. 3.1.

In the figure the loads are truncated at 4 kW, however the maximum values for each time step are between the 6 and 8 kW. From the figure, it can be seen that during the night the load is more homogeneous compared to the rest of the day. During the peak hours of the residential load (from 5PM to 9PM) there is a clear increase in the high percentiles (max, 99th, 95th 75th), while the low percentiles (min, 1th, 5th 25th) remain constant. In order to model the residential load from a network planning perspective, the load during peak hours is the most important for the sizing of the network. The larger variance during these peak hours should be considered if one wants to model the chance of voltage violations within the LV-network.

Another important property of the household load to examine is the autocorrelation of the household load curve. The autocorrelation for each of the measured load curves of 200 households is calculated. On average the loading of a household only shows a significant autocorrelation within the first four hours and at 24 hours time lags (the predictable daily
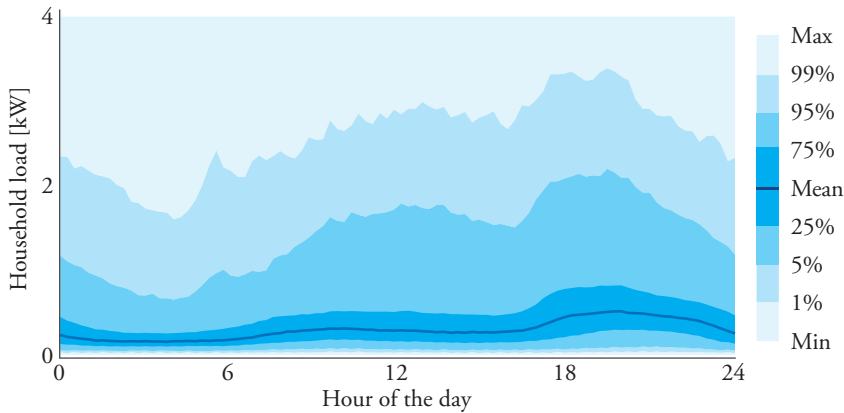
Figure 3.1: *The probability density function and the mean of 200 household load measurements throughout a summer day, truncated at 4 kW*

cycle). For the modelling of the load curve time series, this correlation should be preserved.

Lastly, the correlation between different households in the same neighbourhood is computed. This gives an indication of the level of dependence between the electricity consumption of two households. The correlation for the entire day was 0.258 with a p-value of 0.01, while if only the correlation at the evening peak hours (from 18 to 20) is considered the correlation coefficient drops to 0.012 with a p-value of 0.37 and to 0.014 with a p-value of 0.45 for the morning hours (from 7 to 9). Thus the households tend to follow a somewhat similar pattern on a daily level, while if only a small time frame is considered the household load curves become almost completely independent.

3.1.2 Scenario implementation

The residential load can be modelled based on the average load on the national scale. However, this only gives a rough estimation, since the residential load will be different depending on the local conditions and as time goes by. To be able to model the residential load, these alterations within the residential load should also be taken into account. In addition, the local context can have a profound effect on the residential load. The most important indicators for the residential load are the income, number and age of the household members, and the dwelling size of a household. To be able to model the residential load accurately, these parameters should be introduced within the modelling. These parameters are usually not known on the individual connection level but can be estimated for a single household depending on the statistics of the neighbourhood.

For the modelling of the residential load for network planning purposes the household

load curve should not just be generated for the present moment in time, but also for the coming 40 years. This would require the forecasting of the household load. The forecasting of developments on such time scales is usually done based on scenario assessment. The modelling of the household load should thus also be capable of dealing with the inputs from various scenario studies. The changes in the household load are mainly introduced by two different factors: changes within the appliances present in the household and changes in appliance usage. The drivers for these changes are however harder to completely define, as not only technological changes will have an effect, but also economic advancement and societal changes. This makes the forecasting of the household load curve an extremely difficult task. The implementation of many different kinds of scenarios is thus required to be able to generate a family of household load curves over the coming years.

3.1.3 Computational implications

The residential load can be modelled with a high level of detail, creating complete time series at all connection points, for the coming forty years with many different scenarios. Depending on the analysis, however, this can be an overload of information. The LV-network in the Netherlands consists of more than 200.000 km cable with about 8 million connections. To assess a network of this size, a detailed model of the household load is not always needed as the detail of the analysis is limited by computational requirements. The modelling of the residential loads needs to be adjusted for the goals of the subsequent analysis. To ensure that the results of the different kinds of analysis still stack up, the modelling of the residential load should be based on a common framework. This is further elaborated in the next section.

3.2 Residential load modelling

With the modelling requirements defined, these requirements can be used to build a residential load model. The requirements can, however differ significantly depending on the level of detail of the analysis for which the residential load is needed. To generate comparable results for both general and more detailed analyses, it is preferable that the residential load modelling is based on the same input data. In order to do this, first, a detailed model for the residential load will be defined. For this model, simplifications will be discussed which can generate different sets of input data which meet the requirements for the less extensive LV-network assessments.

3.2.1 Detailed Load Modelling

To generate a detailed model of the household load with similar probabilistic properties to measure data, a bottom-up approach is applied. When it comes to modelling the household

load basically four types of approaches can be applied: Top-down models; which focus on the loading of MV/LV transformers and generate load curves based on this aggregation level [20], bottom-up methods which employ statistical time use data to construct load profiles [21], constructing load curves based on smart meter data [22][23] and machine learning approaches [24][25]. For this thesis, the use of a bottom-up approach is preferred, as it is hard to implement scenarios in machine learning approaches, while top-down models tend to perform better at the MV/LV transformer aggregate level while lacking accuracy on the individual household level. The detailed modelling through a bottom-up approach is discussed in the next subsection. Afterwards, it is shown how scenarios can be implemented in the proposed load modelling appoach[3].

Bottom-up modelling

In a bottom-up residential load modelling approach each appliance in the household, as well as the behaviour of each resident, are modelled. Depending on whether household members are active, the time of the day and the appliances active in the previous time period, the switching on or off of appliances may occur with certain probabilities. In order to model this, a Markov Chain approach is utilised.

Time use surveys (conducted in the Netherlands), in which people note down on a 15-minute basis, how they are spending their time are used to generate the behaviour of the members of the household. The occupancy of the household is the most important factor which is determined from the time use data [4], as occupancy is the main driver for energy usage. The modelling of the occupancy is done based on a Markov Chain approach. In this approach, transition matrices are used to determine whether the state of an occupant in the household changes, whether he/she becomes active or inactive (not present in the household or asleep). The transition matrices are created by using logistic regression on the time use survey data, which are categorised depending on the household size and the age of the household members. The occupancy is modelled with the same resolution as the time use surveys, i.e. on a 15-minute basis for each household. The time series of the occupancy with a 15-minute time step is subsequently interpolated to form a time series with a 1-minute time step.

Next to the occupancy, a similar approach is used to compute transition probability matrices for activities which can be directly linked to the usage of a certain appliance (i.e. watching TV). With the occupancy of the household known, the appliances in the household need to be modelled. First, the appliances are distributed over the households. The households are characterised based on the average level of income of the neighbourhood and the number

---

[3]For a detailed discussion on the household load model the paper M. Nijhuis, M. Gibescu and J.F.G. Cobben, "Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data", *Energy and Buildings*, vol. 112, pp. 121-129, Jan 2016 can be consulted in Appendix A.2

[4]Sociaal en Cultureel Planbureau (SCP); Centraal Bureau voor de Statistiek (CBS) (2011): Tijdsbestedingsonderzoek 2011 - TBO 2011 v2. DANS. https://doi.org/10.17026/dans-zmp-jj2x

of household members, as these factors are the main factors which determine the number of appliances present [26]. For the allocation of appliances to a household, statistics on the penetration levels and the amount of appliances per household type (size and income) are used and normal distributions for these variables are assumed.

The appliances are divided into eight types: base-load, night load, heating and cooling loads, lighting, activity-linked loads, EV, PV and general loads. Based on the type of appliance, a different simulation approach is chosen:

Base-load (e.g. fridge) can be modelled independently of all external inputs, as these appliances are continuously turned on or are switched on and off with little user interference. The energy use of these appliances is thus modelled as constant or variable switched energy usage.

Night load (e.g. electric boilers, dishwashers) consist of loads which usually only run at night (or during off-peak hours). Depending on the occupancy level on the day, the appliances start to switch on after 21:00 with a probability based on an exponential distribution.

Heating and cooling loads (e.g. electric heating) are dependent on the weather conditions and the occupancy. The main weather conditions which influence the use of these loads are the outside temperature and the wind speed. To evaluate how much heating energy is required, data from the allocation process in the Dutch gas sector is used. The gas use in this approach is dependent on the wind speed adjusted temperature and is converted to a required heat to be delivered to the household. This data is combined with the occupancy in order to generate a lower heating demand when the household has no active occupants. The cooling loads are modelled based on the same approach, only the relation between the wind speed adjusted temperature and the cooling load is estimated based on measurements from AC systems.

Lighting is for the most part only used when the irradiance is low ($< 60W/m^2$). The conversion of the irradiance values to energy use is based on the number of active occupants in the household. The number of lights present in the household is dependent on dwelling size.

EV is modelled based on the occupancy in combination with driving distance data. When there is a change in occupancy there is a chance the EV will arrive at or leave the household. After the EV has arrived at the household, the decrease in EV state of charge is estimated from the distribution of the driving distances. [5]

PV is simply assumed to be a function of the irradiance, with the angles of the PV-panel taken into account in the conversion from irradiance to PV output power.

Activity-linked loads (e.g. TV) can be seen as loads which have a direct relation to behaviour

---

[5] More details on the modelling of the PV and EV can be found in: G. Ye, M. Nijhuis, V. Cuk, and J. F. G. Cobben, "Stochastic Residential Harmonic Source Modeling for Grid Impact Studies," *Energies*, vol. 10, no. 3, 2017.

(e.g. cooking, watching TV, etc.) reported in the time use surveys. The underlying patterns of behaviour are dependent on many variables, however as only the resulting energy use is of interest, the appliances are modelled with a Markov Chain, with transition matrices generated through logistic regression of the time use data.

General loads (e.g. water-cooker) are loads which have no clear link to behaviour reported in the time use survey but can only be turned on when occupants are active. These loads have a chance to switch on depending on the yearly average reported usage time of the appliance.

The total household load can be constructed by adding all the appliance loads together.

Scenario implementation

One of the main aspects of a residential load model which is suitable to be used in the assessment of the LV-network is the modelling of future load changes. In order to achieve this, scenarios on drivers which can have a significant effect on the household load should be implementable within the residential load model. These individual scenarios on the main drivers of the long-term evolution of the household load have been discussed in 2.2. The drivers on the household load can be split into two categories, socio-demographic and technological changes. Through altering the inputs of the bottom-up load model, both these types of scenarios can be implemented.

Technological changes, like the adoption of new technologies and the changes in appliance ratings, can be introduced by changing the appliances which are present in the household. This is done in the modelling by half yearly updating the appliances within the households. Depending on the wealth of the household, the owners will replace old appliances with appliances with changed ratings and/or new appliances will be introduced at the household and/or old appliances will be retired, depending on the scenarios.

The socio-demographic changes need to be introduced differently. General changes in behaviour (i.e. the increase in working from home) can be introduced by changing the probability of occupants switching from inactive to active, or by changing the overall probability that an occupant will switch a certain appliance on or off. Changes in more demographic parameters (i.e. age and household size) can be introduced by changing the input parameters for the assignment of appliances to the household.

The whole household load generation procedure is illustrated in Figure 3.2.

3.2.2 Load Modelling Simplifications

With the bottom-up Markov Chain model, a 1-minute time series of the residential load can be defined for each individual household and for many scenarios. To evaluate all these different load curves would be computationally expensive. An analysis which takes into
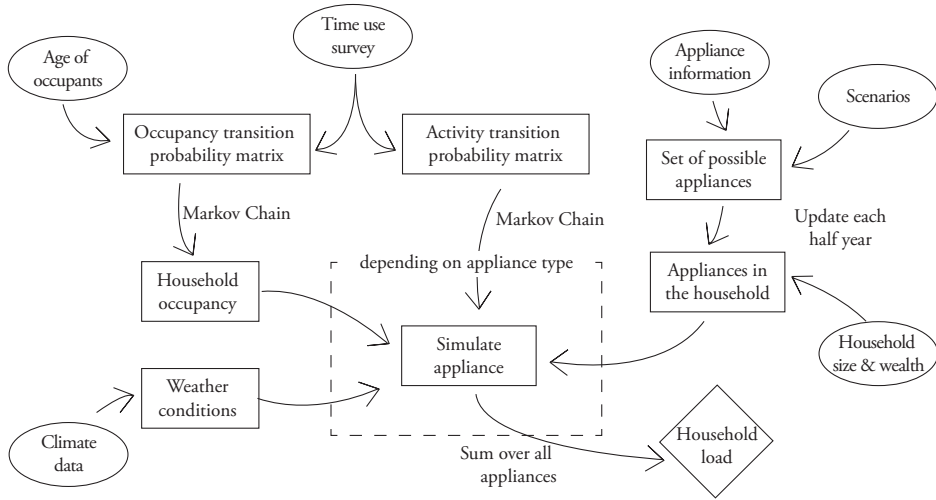
Figure 3.2: *Approach for the creation of the residential load curve*

account a lot of scenarios or many individual LV-feeders is not feasible to be performed using a residential load with this level of detail. Therefore simplifications have to be defined in order to allow for these analyses. Depending on the requirements of the analysis, different simplification methods have to be employed. The clustering of load curves can be applied when the time dependency within models is important, and the use of Gaussian mixture distributions can be applied for peak load estimates.

Clustering of load curves

When examining the effects of load shifting on the network, the time dependency of the load is of importance. For the assessment of the benefits of load shifting, optimisation of each load curve is required to obtain a realistic estimate of the resulting load curves. From a network planning perspective, many different distributions of the household load have to be assessed to generate adequate insight into the risk levels encountered within the network planning process. The assessment of the network based on each possible household load curve individually would become computationally too intensive. The number of load curves should be reduced while the time dependency should remain intact. In order to do this, a clustering approach can be used to generate a set of typical load curves. By performing a fuzzy k-means clustering, the number of load curves can be reduced from 200 to just 5 load curves without compromising too much on the accuracy of the results [6]. The probability of

---

[6] For a more detailed explanation please see: M. Nijhuis, R. Bernards M. Gibescu and J.F.G. Cobben, "Stochastic Household Load Modelling From a Smart Grid Planning Perspective", in *EnergyCon*, April 2016

a certain load curve occurring can be discerned from the clustering approach, allowing for a correct assessment of the risks within the LV-network.

Gaussian mixture estimation of loads

Certain technologies, like PV, can have a large effect on the adequacy of the network. It may, therefore, be interesting to assess at which level of penetration for a single technology the chance of overloading or over- or under-voltages becomes large enough to warrant new investments. To assess the effect of specific drivers on the LV-network, the load can be modelled based on a Gaussian mixture distribution. In Chapter 4 the advantages of the use of loads modelled through a Gaussian mixture distribution are discussed. The Gaussian mixture distribution is given by:

$$f(x) = \sum_{k=1}^{K} \omega_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

$$0 \leq \omega_k \leq 1 \text{ and } \sum \omega_k = 1$$

(3.1)

where K is the number of components in the mixture, ω_k is the weight, σ_k the standard deviation and μ_k the mean of the k-th component. The Gaussian mixture can only be used for a certain time step in the load profile, therefore it is important to first determine the time of interest with respect to the expected effects of the driver in question. The Gaussian mixture distribution is estimated for the initial values calculated with the MCMC method $f(x)_0$ at the current time and also with a 100% penetration rate of the specific appliance $f(x)_{100}$. The EM algorithm is applied, using a two-step iterative approach to find the maximum likelihood for the parameters of the two Gaussian mixture models. Assuming a Gaussian mixture model with K components with the weights ω_k and component distributions $p_k(x_i|\theta_k)$, the likelihood is given by:

$$\mathcal{L}(\Theta|X) = \prod_{i=1}^{N} \sum_{k=1}^{K} \omega_k p_k(x_i|\theta_k)$$

(3.2)

with N being the number of observations of the empirical data X and θ_k a vector consisting of the parameters for the normal distribution $\theta_k = \{\mu_k, \sigma_k\}$. The maximum likelihood estimate $\arg\max_\Theta \mathcal{L}(\Theta|X)$ cannot be determined analytically. The EM algorithm interprets the data X as incomplete and adds a binary vector Y. This vector indicates which observation of X belongs to which component in the Gaussian mixture. This vector Y is determined by using a K-means clustering approach. This results in the following likelihood:

$$\mathcal{L}(\Theta|X,Y) = \prod_{i=1}^{N} \sum_{k=1}^{K} y_{k,i} \omega_k p_k(x_i|\theta_k)$$

(3.3)
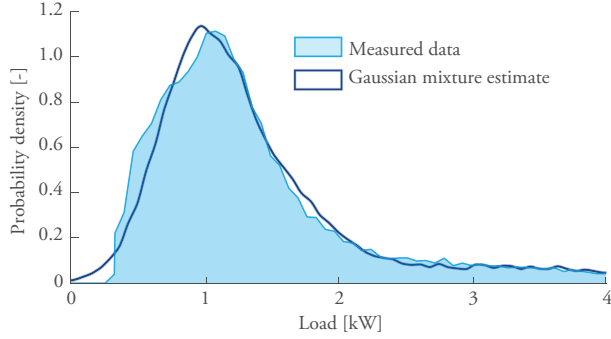
Figure 3.3: *A Gaussian mixture estimate of the household peak load*

The EM algorithm first step is to find the expected value of the complete likelihood given the parametrisation in the previous time step Θ^{p-1} through the Q-function:

$$Q(\Theta, \Theta^{p-1}) = E[\log p(\Theta|X, Y)|(Y, \Theta^{p-1})] \tag{3.4}$$

The next step is the maximisation of the expectation of the previous step $\Theta^p = \arg\max_\Theta Q(\Theta, \Theta^{p-1})$. By iteration of this algorithm, the Gaussian mixture distribution can be defined for both the $f(x)_0$ and $f(x)_{100}$ distributions. The following Gaussian mixture distribution can be constructed by combining the two distributions:

$$f(x) = (1 - p) \sum_{k=1}^{K_0} \omega_{0,k} \mathcal{N}_{0,k}[\mu, \sigma] + p \sum_{l=1}^{K_1 00} \omega_{100,l} \mathcal{N}_{100,l}[\mu, \sigma] \tag{3.5}$$

where p is the penetration rate of the appliance of interest. By applying this approach, a detailed analysis of the effects of the penetration ratio for a single device/technology can be conducted. [7].

An example of the resulting Gaussian mixture distribution of the household load is shown in Fig. 3.3. The main difference between the household during peak hours as measured by smart meter data and the Gaussian mixture estimate is in the loads which are close to zero. As the Gaussian distribution is a continuous distribution, loads close to zero and even negative loads can occur when estimating the household load through a Gaussian mixture distribution. For the analysis of the household load, this is generally no problem due to the very low associated probabilities; if it does lead to problems, then truncation can be applied to the individual components or the complete mixture distribution.

---

[7]For a case study see Appendix A.4

3.3 Conclusions

The modelling of the residential load is essential to be able to perform various types of analyses on the LV-network. To generate a detailed model for the household load in which future scenarios and different local conditions can be implemented, a bottom-up Markov Chain approach should be used. A detailed model has been defined based on time use surveys. In this model, scenarios can be implemented by dynamically updating the appliances present in the household, and by changing the socio-demographic inputs which govern these appliances within the household. As the detailed model generates household load curves with a 1-minute time resolution, simplifications have to be defined to be able to use these household load curves for computationally more extensive LV-network studies. The use of a clustering approach can generate generic times series of the household load curves, while the use of Gaussian mixture distributions can be used for peak loading assessments. With these simplifications, the bottom-up Markov Chain residential load modelling approach can be used as a starting point for most types of LV-network analyses.

4

The Assessment of the LV-network

With an approach for modelling the household load for the coming years in hand, the other main component which should be modelled in order to be able to assess the LV-network, is the LV-network itself. The modelling of the low-voltage distribution network can be done on many scales and at different levels of detail. Depending on the goal of the network analysis, the level of detail should be chosen. When it comes to the long-term adequacy of the LV-network, detailed modelling is usually not employed. The level of uncertainty in the loading of the network is so large, that it outweighs the errors introduced by the non-perfect modelling of the network itself. A simple single-phase equivalent circuit already offers enough detail for the assessment of most LV-network planning related problems. More detailed modelling to include, for instance, the imbalance between the phases or the thermal properties of the components can be of significant value when assessing an LV-network. These types of modelling require however detailed information on the distribution of the load per phase, or the soil conditions. When facing operation problems at a specific location, this information can be acquired, but in a more broad assessment of the adequacy of the (future)

LV-network, these data cannot be determined with sufficient accuracy.

The main difficulty when it comes to modelling the LV-network for the long-term assessment of its adequacy is the sheer size of the network. In the Netherlands, the LV-network consists of over 120 000 MV/LV substations with over 300 000 LV-feeders connected to them. These 300 000 feeders connect to a total of over 8 000 000 LV-consumers. To model this entire system is infeasible from a computational point of view. The modelling of the LV-network should, therefore, be based on a generalised network structure, or the modelling should only represent a small part of the network. In other to arrive at a representative model for the entire LV-network, a clustering approach can be applied to the network data. This is elaborated upon in section 4.1.

Next to the modelling of the LV-network, other important questions to be answered are: which metrics are important to measure the adequacy of a network and how should these metrics be calculated. As stated in section 2.1.2 there are multiple requirements for the distribution network, however, not all of these requirements are affected by the anticipated changes in the loading or the structure of the LV-network. Next to this, the requirements as stated in section 2.1.2 should be quantified to assist with the analysis of the LV-network. As this will result in multiple indicators for the network, these indicators are subsequently compounded into a single "risk" score for a given part of the LV-network.

The main method of analysing a distribution network is by using load flow calculations. The uncertainty about the loading of the LV-network is, however high. Hence, a deterministic approach with respect to the LV-network analysis can generate an unrealistic picture of the actual risks within an LV-network. The LV-network does have some characteristics (as described in section 2.1.1) which would allow for a fast calculation of the probability density functions of the branch currents and bus voltages. An approach to the calculation of this probabilistic load flow is given in the last section of this chapter.

4.1 LV-network modelling

The total electricity grid in the Netherlands alone consists already of over 300 000 km of cable and 8 000 000 connections. To analyse this complete network in one go is, through the sheer size of the network alone, already computationally very intensive. The network is therefore often cut up into smaller pieces to allow for an adequate analysis of the sub-networks. The LV-network can be partitioned into the sub-networks behind each MV/LV substation, or into the individual feeders connected to the MV/LV substation. This makes the network under focus small enough to be analysed quickly, however, the number of networks which have to be evaluated has risen now to over 150 000 sub-networks emanating from each MV/LV substation, or over 600 000 LV-feeders. The analysis of these individual networks can be feasible only for very simple network calculations. If one, for instance, wants to assess the performance of the LV-network over many different scenarios, the problem quickly

becomes computationally too intensive. Therefore, a clustering approach has been developed in order to reduce the number of LV-feeders which have to be assessed. Next to using a reduced network based on generic LV-feeders, the analysis of the whole LV-network can still provide useful insights. How this network can be modelled is discussed after the network reduction section.

4.1.1 Network reduction

The generation of a reduced LV-network from the complete network is discussed in this section. The generation of this reduced network is necessary from a computational point of view. To be able to design a good method for the reduction of the network, the subsequent analysis should be taken into account. Depending on how the reduced network is analysed, the method for creating the reduced network can differ. When it comes to the analysis of the reduced network, there are a number of important aspects which should be taken into account. First of all the results from the reduced network should be able to be reliably scaled up towards the complete network, to ensure the results from the reduced network carry meaning on a bigger scale. Secondly, the analysis of the network will often be based on load flow calculations, therefore a load flow calculation performed on the reduced network should give a result similar to the calculation of the load flow on the larger network. Taking both these aspects into account, a network reduction approach needs to be employed. The characterisation and clustering of electrical networks have previously been studied. For the evaluation of reliability and susceptibility to threats, clustering based on graph theory is already being used, especially for the transmission network [27, 28, 29]. In [30] a small number of networks are defined, based on the length of the feeder, the number of connected customers and the number of branches. Though some analysis can be performed on these representative networks, they are not classified based on enough detail to be usable for network planning. A more extensive approach is required to be able to create generic LV-feeders with a strong relation to the existing LV-networks. In this thesis, fuzzy k-medians clustering approach is applied to generate generic LV-feeders. [1] The approach is based on the main network parameters: impedances, cable lengths, number of branches and branch depth and the number and type of connected customers. These parameters are combined with the graph theory concepts of degree distribution, sequence and the centrality of the power, impedance and length. By using these parameters, representative feeders can be rebuilt from the cluster centres which would represent the structure and loading of the original feeders.

This network clustering approach has been applied to the LV-network of Liander, a Dutch DNO. This resulted in 94 generic LV-feeders, which can be used as a representative set of LV-feeders for the entire network. The network for a single MV/LV substations can be built from these generic LV-feeders. The LV-feeders in this generic set are chosen based on the cluster

---

[1] For a detailed discussion on the network clustering the paper M. Nijhuis, M. Gibescu and J.F.G. Cobben, "Clustering of Low Voltage Feeders From a Network Planning Perspective", in *International Conference and Exhibition on Electricity Distribution (CIRED)*, June 2015 can be consulted in Appendix A.3

	Generic	Original
Mean	0.990	0.990
Standard deviation	0.011	0.012
Difference between the mean of a cluster and its members		
Maximum	0.2%	
Average	0.08%	
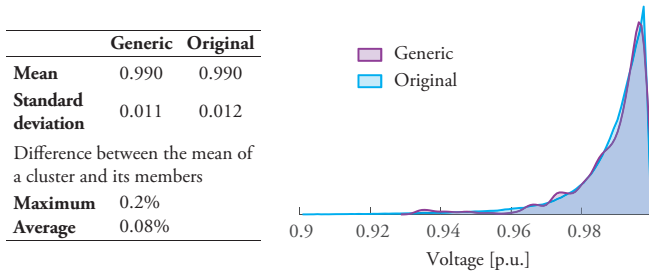
Figure 4.1: *Overview of the performance characteristics of the clustered networks versus the original network*

centres of the k-medians clustering approach. The LV-feeder with the smallest Euclidean distance to the cluster centre is chosen as a generic LV-feeder for that specific cluster. For a complete set of LV-feeder characteristics Appendix B can be consulted.

Whether the clustering approach yields a set of LV-feeders comparable to the original set in terms of voltage profiles still needs to be tested. This is done by comparing the resulting voltages from load flow calculations on both the original set of LV-feeders, as well as the set of generic LV-feeders. Fig. 4.1 has been created for the sake of this comparison. On the left side of the figure, the mean and standard deviation of the original and the generic LV-feeders are shown, as well as, the maximum and average deviation of the mean voltage level between the generic and complete set of LV-feeders. Both of these measures show a good fit with errors in the mean voltage being less than 0.5%. On the right hand side of Fig. 4.1 the kernel smoothing density estimate of the distribution of voltages from the original and the generic set of LV-feeders is shown. As the generic set only has a limited number of measurements (94) compared to the total set of LV-feeders (88 000) the distribution of the voltages of the generic is less smooth. The general shape of both distributions is however similar, showing that the clustering approach can be applied to generate a set of generic feeders with similar voltage profiles as the original LV-feeder set.

4.1.2 Full network analysis

Information on the complete LV-network is often of interest when we want to generate an overview of the status of the network. Depending on what the desired output is, two basic approaches to the analysis of the LV-network can be applied. First of all the generic LV-feeders as defined in the previous subsection can be used to represent the network. This approach will generate a detailed analysis of the LV-network which remains computationally feasible. Another approach is to use a simple set of indicators to evaluate the network. For the LV-network design, the most stringent requirements are related to the PE-conductor voltage rise (touch safety) and fast voltage fluctuations [31]. The adequacy of the LV-network with respect to these two requirements can be estimated directly based on the network

characteristics rather than by performing extensive load flow calculations. By looking at the impedance at the point of connection, an indication of the number of fast voltage fluctuations can be obtained. By looking at the maximum impedance in combination with the fuse ratings, the PE voltage rise can be estimated, in this way an estimation of the adequacy of the entire LV-network can be determined with respect to these two requirements [32].

4.2 LV-network metrics

When analysing the LV-network, next to the method used in the analysis itself, also the metrics which are assessed during the analysis are of importance. In this section, a closer look will be taken at these metrics, based on the requirements for the LV-network as defined in section 2.1.2. The main requirements which are affected by the transition towards a sustainable energy supply are given as: the LV-network capacity, the network losses (as the main indicator of the operating cost), the voltage level (as main power quality indicator) and the availability of the network. How these requirements can be assessed will be discussed in this section. To facilitate the analysis of the LV-network, the main essence of these requirements will be compounded into a single risk indicator for the LV-network.

4.2.1 Capacity

The capacity of the cables and transformers in the LV-network is one of the main requirements. The capacity of these components is limited by their thermal limits. A current higher than the rated current will, in time, heat up the component to unsustainable levels (usually these temperatures indicate that the insulation may start to melt). To accurately assess the capacity of the cables and transformers, the temperature of these components should be modelled. Modelling the component temperature would require the calculation of load flows in a time series set-up, as the thermal processes can take multiple hours to stabilise. For very detailed analysis, or when needing to make operational decisions, assessing the capacity from a thermal point of view makes sense. From a planning point of view, the requirement of performing a time-series analysis in combination with the differential equations for the temperature would be excessive from a computational perspective. The use of the rated values of the components in combination with a few snapshots during peak hours makes more sense. The assessment of the rated values can be extended to allow for a certain percentage of intervals which exceed these rated values, as it can be assumed they would not lead to breaches of the thermal limits.

4.2.2 Losses

The losses in the LV-network represent the largest part of the operational cost and the carbon footprint of the DNO. Though this cost will keep rising if the loading of the network increases, the losses are generally not considered when assessing the adequacy of an LV-network. When building a new LV-network the cost of the losses are considered, as they can differ significantly for different network designs and the cost of the losses is comparable to the cost of the cables over their lifetime. The losses are generally assessed through load flow calculations. For optimisation purposes, the losses need to be estimated to avoid calculating the probabilistic load flow at many time intervals. This can be done through the use of a network equivalent impedance, based on the number of connections, the location of the connections and the impedance at the point of connection [33].

4.2.3 Voltage level

The voltage level in the LV-network is limited to $U_n = 1 \pm 0.1[p.u.]$. In order to calculate the voltage at the nodes within the network, the load flow equations for a given set of node injections and withdrawals need to be solved. As the voltage deviation is dependent on where the highly loaded buses are in the network, the network should be assessed in a probabilistic manner. If the largest nodes are close to the substation, the voltage drop will be small, however, if they are far from the substation, then the drop will be high. The combination of loads can thus play an important role in the voltage deviations experienced in the LV-network. The use of a probabilistic load flow to assess the voltage levels within the LV-network should, therefore, be employed. In practice, a more stringent limit than the $\pm 10\%$ is usually employed in the LV-network. The main reason for this is that a voltage deviation from the MV-network will propagate to the LV-network, as on-load tap changers are usually not installed at MV/LV transformers. If the MV-network is not taken into account, the limits of $\pm 5\%$ are therefore often used to assess the LV-network. This is the approach taken in this thesis.

4.2.4 Availability

The availability of the LV-network is usually assessed as the amount of time the LV-network voltage at a customer connection point is interrupted. The detrimental effects of an interruption depend on a number of factors. With the time of day, the duration and the frequency of an interruption being the most important ones. In the LV-network the main cause of an interruption is an interruption at the MV-level. The damage for the consumers should still be taken into account when assessing the expansion or creation of an LV-network. The damage to consumers can be assessed in multiple ways. The main ways are accounting for the minutes lost or the amount of lost load. The amount of lost load is more difficult to take into account. the stochastic nature of the load makes a prediction of how much power is lost

if an interruption happens hard to estimate. Next to this, the availability of the LV-network needs to be assessed over long time horizons and the current loading situation will most likely change within this time horizon. For the assessment of the availability, the customer minutes lost are therefore used. Assessing these minutes lost only makes sense as the network structure changes, or when the average age of asset replacement needs to be determined. For these reasons, availability is not taken into account in the risk level definition below.

4.2.5 Risk levels

Based on the assessment of various requirements of the LV-network, a risk level for the LV-network is defined. This risk level represents the chance the LV-network would not be able to handle the changes introduced by the transition towards a sustainable energy supply. The two main requirements to assess the future tenability of the LV-network are the capacity and the voltage levels. As there is a large uncertainty about the actual loading of the network, the requirements should not be assessed in a deterministic way, but rather based on the distribution of possible loading situations for a single scenario and a single step in time. As the capacity is based on the thermal limits, overloading for a small amount of time should only add very limited risk. By contrast, even temporary overvoltages can quickly cause damage within the installation of a consumer, therefore a more stringent approach with respect to the voltage levels is chosen. This leads to the following definition of the risk within the LV-network [34]:

$$
\text{Risk} = \sum_{n=1}^{N}
\begin{cases}
\frac{1}{N}, & \text{if } V_{5^{th}} \leq 0.95 \vee V_{95^{th}} \geq 1.05 \\
& \quad \vee S_{99^{th}} \geq 1.2 \\
\frac{0.2}{N}, & \text{if } V_{1^{st}} \leq 0.95 \vee V_{99^{th}} \geq 1.05 \\
& \quad \vee S_{95^{th}} \geq 1 \\
\frac{0.05}{N}, & \text{if } V_{min} \leq 0.95 \vee V_{max} \geq 1.05 \\
& \quad \vee S_{99^{th}} \geq 1 \\
0, & \text{otherwise}
\end{cases}
\tag{4.1}
$$

where n is a single sample for the result of the probabilistic load flow, N is the total number of samples, V_{max}, $V_{99^{th}}$, $V_{95^{th}}$, $V_{5^{th}}$, $V_{1^{st}}$ and V_{min}, the maximum, 99th, 95th, 5th and 1st percentile and minimum voltage value respectively, and similar for the component's apparent power loading S. With this risk level, the adequacy of the LV-network with respect to the changes introduced by the transition towards a sustainable energy supply can be combined into a single number.

4.3 LV-Network analysis

Next to the modelling of the LV-network, the methods to quantitatively analyse the network to gain more insight into the metrics which determine the adequacy of the LV-network, should be assessed as well. The most common way of analysing the LV-network is through the use of load flow calculations, to obtain the voltages and currents within the LV-network for various network conditions. The uncertainties within the LV-network require a less deterministic way of analysing the LV-network. The standard probabilistic load flow can be altered to a variant which makes use of the characteristics of the LV-network to allow for fast computation of the voltages and currents as probability density functions. This is discussed in more detail in the first subsection. Next to obtaining the probability density function of the output variables, it is also important to note how these output variables are affected by changes in the load or in the cable parameters. With the sensitivities of these variables known, an estimation on which parts of the network are most vulnerable can be made. Therefore, the use of a global sensitivity analysis which can incorporate these uncertainties is discussed in the second subsection.

4.3.1 Probabilistic load flow for the LV-Network

When assessing the network it can be more interesting to know the probability density function of the voltages and currents within the LV-network as opposed to just deterministic values. Many variables within the LV-network are of a stochastic nature, therefore deterministic values for the voltages and currents can often give an incomplete view of the LV-network. The most common way of applying a probabilistic load flow is through the use of a Monte Carlo simulation. From a computational point of view, this can be an expensive process. Advanced probabilistic load flow methods are therefore necessary to determine the effects of the new loads on the LV-network.

There have been many different probabilistic load flow or optimal power flow formulations [35, 36, 37, 38] that are able to accurately and relatively quickly assess the distribution network. These methods generally reduce the required amount of samples in a Monte Carlo (MC) simulation or are based on two-point [39] [2] or multiple points estimates [40]. Gaussian mixture models have already been used to improve the probabilistic load flow computations as well. Application to MV-networks with the number of stochastic loads limited to mostly the distributed renewable energy sources has been performed [41][42]. For the DC load flow, a probabilistic load flow based on cumulants and Gram-Charlier expansion has already been applied [43]. However, for the LV network, the DC simplifications will not hold. The LV-network has in most cases a radial topology combined with a high R/X ratio, hence other simplifications to the load flow can be made to further improve the computational speed. For a rough initial network assessment, the trade-off

---

[2]The most commonly used load flow software for Dutch DNOs, GAIA, uses a two-point probabilistic load flow

between speed and accuracy is different, as explained below. With a quick initial assessment, the networks which truly require attention can be identified and more in-depth analysis can be performed on a limited subset of LV-networks.

As stated in the section 2.1.1 the LV-network consists mainly of radial networks with high R/X ratios. These characteristics can be used to improve the calculation of the probabilistic load flow. Two assumptions will, therefore, be made for the load flow calculations within the LV-network to develop a computationally efficient probabilistic load flow calculation. First of all, the voltage angle within the network is assumed to be constant. As the R/X ratio within the LV-network is high, the deviation of voltage angle will, therefore, be low. The second assumption is that the relation between the voltage deviation calculated during the first iteration of the load flow and the final voltage deviation when the load flow has converged is linear. The combination of these two assumptions allows for the direct implementation of loads described as a Gaussian mixture distribution (see section 3.2.2) in the load flow. This approach can generate an accurate representation of the probability density functions of the voltages and currents within the network in a computationally feasible manner.

The accuracy of this Gaussian mixture based load flow is evaluated with respect to the conventional Monte-Carlo based approach of using a Latin hypercube sampling based MC simulation [44]. This is done in combination with load flow calculations using a full backwards-forwards sweep. The load flows are calculated by using scaled versions of the distribution as shown in Fig. 3.3 for each load. For the testing of the Gaussian mixture based load flow approach, two of the Cigré US LV test feeders [45] (with the voltage level of the US feeder reduced to the more common 120V) and the IEEE European test feeder [46] are used. In the Fig. 4.2 the resulting PDF of the voltage at the farthest bus for the US residential feeder is plotted for the Gaussian mixture approach as well as for the MC approach with increasing sample sizes.

From the figure, it becomes clear that the probability density function (PDF) for minimum nodal voltage for the US residential case presents many modes. This can be explained by the low number of buses in the system, too low for the central limit theorem to apply. The MC sampling results with only 10^4 and to a lesser extent with 10^5 samples, show erratic behaviour. The MC method with 10^6 samples and the Gaussian mixture approach are close together. A more detailed analysis is required to judge the applicability of the Gaussian mixture-based load flow in comparison with the established MC sampling.

The result of a probabilistic load flow is a histogram of the possible outcomes. To assess to what degree two histograms are derived from the same underlying distribution, a number of statistical tools can be used. The three tests used will generate an outcome of maximum 1 if the histograms are exactly the same, while if there are discrepancies the result will be lower. To give a combined overview of the fit of histograms H and G the following compounded measure $D_{tot}(G, H)$ is defined for this thesis, which is the product of three individual distance measures:

$$D_{tot}(G, H) = D_W(G, H) \cdot D_B(G, H) \cdot D_{KL}(G, H) \tag{4.2}$$
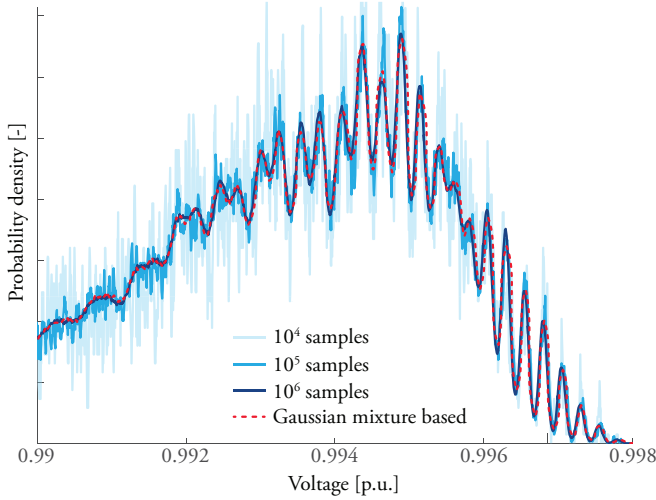
Figure 4.2: *Probability density function of the voltage at the farthest bus for the US residential feeder for the MC method and the proposed Gaussian mixture based load flow*

where $D_W(G, H)$ is the Wasserstein metric or earth movers distance [47], defined as follows:

$$D_W(G, H) = \frac{\min_{f_{ij}} \sum_{i,j} |f_{ij}| \cdot d_{ij}}{\sum_{i,j} f_{ij}} \tag{4.3}$$

where f_{ij} is a function for which the following holds: $G_i + f_{ij} = H_i$ and $G_j - f_{ij} = H_j$ the and d_{ij} the distance between i and j. $D_W(G, H)$ is the Bhattacharyya distance [48], defined as follows:

$$D_B(G, H) = -\ln \sum_i^n \sqrt{G_i \cdot H_i} \tag{4.4}$$

where G_i is the bin count of bin i in histogram G and H_i is the bin count of bin i in histogram H. $D_{KL}(G, H)$ is the Kullback-Leibler divergence [47] as defined by:

$$D_{KL}(G, H) = \sum_i^n G_i \ln \frac{G_i}{H_i} \tag{4.5}$$

By combining these three metrics both the absolute and relative error in bin count as well as the total variation distance will be assessed. The results for the fit of the histogram with respect to a distribution generated by an MC simulation with 10^7 samples and the required computational times (using an Intel i7-3770 3.4 GHz processor, 8GB ram computer running Matlab 2015a) are given in Table 4.1.

Table 4.1: *Evaluation of the MC and Gaussian mixture-based (GM) load flow methods for three feeders based on the calculation time and D_{tot} when comparing with a probability distribution generated with MC simulation with 10^7 samples*

US Residential [p.u.]	MC-10^4	MC-10^5	MC-$5 \cdot 10^5$	GM
D_{tot} [-]	0.869	0.961	0.994	0.996
Calculation time [s]	6.17	49.3	221	4.58

US Commercial	MC-10^4	MC-10^5	MC-$5 \cdot 10^5$	GM
D_{tot} [-]	0.879	0.965	0.995	0.992
Calculation time [s]	6.68	49.0	217	4.45

EU Residential	MC-10^4	MC-10^5	MC-$5 \cdot 10^5$	GM
D_{tot} [-]	0.868	0.969	0.995	0.996
Calculation time [s]	7.98	64.8	307	51.5

From the table, it can be seen that the Gaussian mixture-based method performs comparably to an MC simulation with 10^7 samples. The computational time is however significantly lower with only 2.1% of the computational time needed for the US test feeders, while the Gaussian mixture approach required 16.8% of the time for the EU case. The difference in calculation time between the US (both ±20 nodes) and EU (±200 nodes) test feeders indicates the scalability of the Gaussian mixture based is slightly lower compared to performing a Monte Carlo approach, due to the need for additional component reduction steps. The computational time to obtain a similar level of accuracy is however still much lower compared to a conventional MC approach.[3]

4.3.2 Sensitivity analysis

The stochastic nature of many variables within the LV-network needs to be taken into account when analysing how the LV-network reacts to changes in the load or branch impedance. With a sensitivity analysis, the reaction of the network to a change in one of the input parameters can be determined. The sensitivity analysis can be applied to a number of different problems, to assist in designing measurement networks [49], selecting the location for voltage control devices [50, 51], placing distributed generation units [52], alleviating voltage problems [53], improving transmission capacity [54], reconfiguring radial networks [55], and also minimizing energy losses [56]. The employed sensitivity analysis does not take into account the stochastic nature of many of the input variables of the LV-network, as most methods are based on a local sensitivity analysis (LSA). An LSA analyses the effect of a

---

[3]For a detailed discussion on the Gaussian mixture based load flow the paper M. Nijhuis, M. Gibescu and J.F.G. Cobben, "Gaussian mixture based probabilistic load flow for LV-network planning", *IEEE Transaction on Power Systems* (in press) 2017 can be consulted in Appendix A.4

single input on the output at a time, while treating the other inputs as deterministic values. An LSA locally estimates each sensitivity measure in the vicinity of a nominal value of the selected input. These LSA can be found in three variates, the first one uses the inverse of the power flow Jacobian matrix [57], [58], the second is the so-called 'perturb-and-observe' method which makes a small change to an input factor and records the effect [59], and the third one applies Tellegen's theorem and the concept of adjoint networks [60].

Instead of exploring only a fraction of the design space and neglecting the joint effects among inputs, a global sensitivity analysis (GSA) examines the sensitivity measure from the perspective of the entire range of each input's variation. Since many input variables in the power system are uncertain and can influence each other, a GSA is preferable over an LSA which gives an incomplete view of the sensitivities within a system. To perform a GSA, mainly four methods are used [61]: the non-parametric method, the regression-based method, the variance-based method, and the density-based method. Since the last decade, plenty of interpretations and discussions on the variance-based GSA have been presented [62], which highlights that the variance is a versatile and proper indicator to describe the output variability without any hypothesis on the linearity or monotonicity of the model. The variance-based GSA provides information on the contribution of each input variable to the variance of the selected output, either a single variable or in combination with others.

The variance-based GSA was first proposed by [63], known as the Fourier Amplitude Sensitivity Test (FAST). In the FAST, only the effects of individual inputs were calculated. Later, the total effects were also included by [64]. Meanwhile, the original FAST was developed further by [65]. In Sobol' method, the sensitivity is evaluated by Sobol' indices which are capable of measuring sensitivities for arbitrary groups of factors by Monte Carlo (MC) simulations. The extended FAST and Sobol' methods are similar in terms of functionality, the only difference lies in the way of the multidimensional integration. Sobol' indices are traditionally computed using random simulations which require a large number of model evaluations. Hence, the Sobol' method can hardly be applied to computationally expensive models, e.g. power flow equations of a network with large size and high complexity. In order to accelerate the computation of the Sobol' indices, the surrogate model, also named meta-model, comes into play. Commonly-used surrogate models [66] include artificial neural network, polynomial chaos (PC) expansion, kriging, space mapping, etc. From among these, PC expansion has been applied to GSA [61, 62, 63, 64, 65, 66, 67, 68]. The PC expansion intends to build a reduced response model which is statistically equivalent to the original one but is less computationally demanding. The advantage of the PC expansion for the variance-based GSA is two-fold: 1) analytical expressions for Sobol' indices can be obtained from the coefficients of PC expansion if the inputs are statistically independent; 2) numerical results can be estimated in an efficient manner if there is a correlation among the inputs. To show the difference in the LSA and GSA, the results for both an LSA and a GSA have been plotted in Fig. 4.3.

From the figure, it can be seen that the results from the LSA and GSA are very similar. With the GSA the difference between the sensitivities is easier to discern. In the GSA the
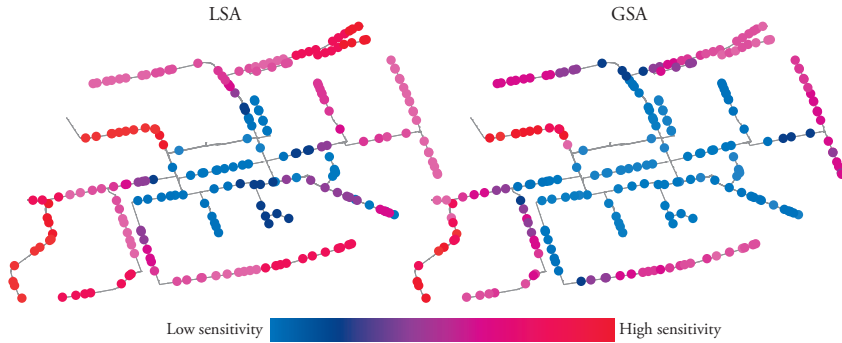
Figure 4.3: *The sensitivity of the voltage level with respect to changes in the load for both an LSA and a GSA*

loads are represented by a mean and a variance so that the volatility of the load can be taken into account. This is the most important benefit of applying a GSA rather than an LSA. In contrary to the LSA, the GSA can include the uncertainties present within a power system in the sensitivity analysis. The GSA can be used to describe a large set of possible input uncertainties. The main drawback of applying a GSA over the more conventional approaches is the computational demand. To be able to capture the output sensitivities correctly, the GSA requires Monte Carlo sampling. As the load flow equations are generally solved through an iterative process, these Monte Carlo simulations make the process computationally infeasible. Through the application of basis-adaptive sparse polynomial chaos expansion, a surrogate model can be created which can speed up the computation of the GSA. The application of the polynomial chaos expansion has another benefit, as the sensitivities can be obtained from the coefficients of polynomial chaos expansion if the inputs are statistically independent. In this way, a sensitivity analysis which is capable of incorporating the stochastic nature of the household load can be applied when the DNO wants to know how the network would react to a change in loading or network parameters. [4]

4.4 Conclusions

For the point of view of the long-term adequacy of the LV-network, the most important aspect is the capacity of the LV-network in terms of both voltage range as well as current carrying capacity. These two metrics should give the best indication of the adequacy of an LV-network. Assessing the capacity of the LV-network can be problematic because of its large size; the entire Dutch LV-network, for instance, consists of over 120 000 MV/LV substations

---

[4]For an elaboration on the use of the global sensitivity analysis the paper F. Ni, M. Nijhuis, P. Nguyen, and J.F.G. Cobben, "Variance-Based Global Sensitivity Analysis For Power Systems", *IEEE Transaction on Power Systems* (in press) 2017 can be consulted in Appendix A.5

with over 300 000 LV-feeders connected to them. In order to analyse this network, a generalised LV-network model needs to be developed. By generating this simplified model based on a fuzzy k-medians clustering approach, the link between the original network and the resulting generic LV-networks remains, allowing for the generalisation of the results obtained from a small set of generic feeders.

To assess the voltages and currents in a radial LV-network, a backwards-forwards sweep load flow can be applied. In the LV-network, the voltage angle can be neglected as R/X ratios in LV-networks tend to be high and the acquired voltage difference from the initial iteration is shown to be almost linearly related to the difference between the initial and final iteration voltage deviation. These two simplifications allow for the use of a Gaussian mixture distribution for the household load as an input to the load flow computation. Based on these computations the distribution of the voltage at each bus and power at each branch can be analytically determined, allowing for much faster probabilistic load flow computations.

5

The Adequacy of the LV-Network: A Dutch Case Study

Based on the future scenarios, the state of the current LV-network can be determined as well as how susceptible the network is to the increase in penetration of certain technologies. The network of Liander, which accounts for about 30% of the total LV-network of the Netherlands, has been used as a case study to analyse the current state of LV-networks. As this LV-network connects almost 3 million customers, the clustering approach as described in section 4.1 is applied to the whole network to distil the whole LV-network into 94 generic feeders. These 94 feeders are subsequently used to gain an idea of the effect of the transition towards a sustainable energy supply on the entire LV-network. To characterise the transition towards a sustainable energy supply the scenarios as described in section 2.2 are applied in combination with the modelling of the household load as described in section 3.2.1. With this modelling approach, many household load curves can be generated for each of the scenarios which describe the most important advances in the household load for the coming decades. These generated load curves are applied to each of the households connected to the 94 generic feeders, and by utilising conventional load flow calculations the state of the network for the coming decades can be determined. The definition of the risk in the LV-

network as given in section 4.2.5 is applied to convert the results from the many load flow calculations to a single metric which indicates the adequacy of the LV-network. The general results of these calculations are shown in Table 5.1.

Table 5.1: *Percentage of the LV-networks having a maximum risk level for multiple percentages of scenarios*

	Maximum risk		
	0	**≤ 0.1**	**≤ 0.5**
100% of all scenarios	63.8	71.8	75.4
99% of all scenarios	67.1	78.4	79.5
95% of all scenarios	67.7	78.8	83.1
75% of all scenarios	75.0	81.2	86.9
50% of all scenarios	79.5	84.5	93.0

This table shows the percentage of the LV-networks which has a certain level of maximum risk for a certain percentage of all the scenarios analysed. From the table, it can be seen that the majority of the LV-network will not experience any risk, even if all the possible scenarios are considered. For more than three out of every four networks the risk always stays under the 0.5 as can be seen from the table. When the most risk inducing scenarios are neglected, the percentage of the LV-network which experiences no, or only a small amount of, risk increases. It can be concluded that the largest part of the Dutch LV-network would be sufficiently strong enough to handle all the changes introduced by the transition towards a sustainable energy supply. Most of these risk occur due to voltage violations, if the band of the voltage limits can be extended to $\pm$ 10%, by means of voltage control at the MV/LV transformer, the risk can be reduced by as much as 76.4%.

In the rest of this chapter, the results of the network adequacy calculations are further evaluated. First of all, a closer look is taken at which of the network characteristics would be most indicative for the risks in the LV-network. Hereafter the time evolution of the risk within the network is evaluated, followed by an overview of the geographical distribution of the risks. In the final section of this chapter, the effects of the different scenarios on the risk within the LV-network is given.

5.1 Network characteristics

First of all, a closer look is taken at the network characteristics which would lead to higher risk levels. The correlation of the risk with a number of different feeder characteristics has been calculated. A correlation of 0.77, 0.26, 0.079 and 0.37 has been found for the number of customers connected to the feeder, the total feeder length, the average feeder impedance and the average impedance at the point of connection respectively. The correlation between the number of connected customers and the risk is more than twice as large as the correlation

between the risk and all the other network characteristics. The amount of change in the load will also be highest if the number of consumers is highest. This observation, in combination with the fact that the current LV-network is operating without violations, can explain why the correlation between risk and number of consumers is highest.

In table 5.2, the risk for subgroups created based on the feeder characteristics is given. The risk as shown in the table is for the year 2045 and is the average of the risks over the scenario space. The risk will have a value between 0 and 1, with an average risk level over all the different feeders of 0.25 for this moment in time. The LV-feeder characteristics which are chosen to be analysed are the number of customers the feeder has connected, the total length of the feeder, the average impedance of the feeder and the average impedance at the point of each of the household connections. This set of characteristics also corresponds to the characteristics used during the clustering approach. The differences between the feeders based on these characteristics should therefore also be higher, generating a better indication of how important these characteristics are for determining the risk within an LV-feeder.

Table 5.2: *The adequacy of the LV-network in 2045 split out based on subgroups based on feeder characteristics, with the average value of the network characteristic in question for each subgroup indicated between parentheses*

	Number of Customers	Length ([m])	Average cable Impedance ([mΩ/km])	Impedance at PoC ([mΩ])
5% Highest	0.82 (90)	0.21 (957)	0.45 (757)	0.60 (158)
25% Highest	0.69 (62)	0.41 (668)	0.45 (535)	0.49 (113)
50% Highest	0.46 (48)	0.36 (523)	0.39 (466)	0.37 (89)
75% Highest	0.31 (40)	0.29 (440)	0.29 (427)	0.28 (74)
95% Highest	0.25 (34)	0.25 (389)	0.26 (401)	0.25 (65)

From the Table 5.2 a number of observations can be made. When looking at the different groups of feeders, a clear difference between the different subsets can be seen. When adding feeders with a lower number of connections to the set, the risk decreases. The length shows some correlation with the risk level present in the LV-feeder, however far smaller than with the number of connections. Which is in line with what one would expect as the voltage drop is proportional to the length and to the number of customers squared (if one assumes all consumers are located at the end of the feeder). When looking at the subgroups, it can be seen that the longest feeders have a below average risk. The longest feeders are often placed in rural areas and the feeder is in these cases often limited by the constraints on flicker and safety. The feeders have therefore relatively low voltage deviations and current loading. An increase in loading will thus not directly increase the risk in these feeders. For the other subgroups, the expected trend of a decreasing risk with a decreasing feeder length is present. The correlation is lowest with the average impedance, which is an indicator for which kind of cables are used within the feeder. The average impedance can be an indication of the age of the network, with older networks generally having a higher impedance (tapering of the cable diameter towards the end of the LV-feeder has been gradually phased out). The impedance at

the point of connection, which is a measure of the distribution of the loads along the feeder, shows the second highest correlation. The risk values for the subgroups show also the trend that a decreasing impedance at the point of connection has a decreasing risk attached to it.

5.2 Time evolution

In addition to splitting out the risk based on the feeder characteristics, the evolution of the risk through time can be evaluated. In figure 5.1, the evolution of the risk for both the summer and winter load for the coming 40 years is depicted. The figure shows the probability density at each of the time steps, with darker colours indicating a higher probability. The risk in the picture is the average risk over all the feeders and all the scenarios. The risk has been split out into summer and winter, as certain technologies will have a large impact in the summer (PV) while others have their largest impact in the winter (heat pumps).
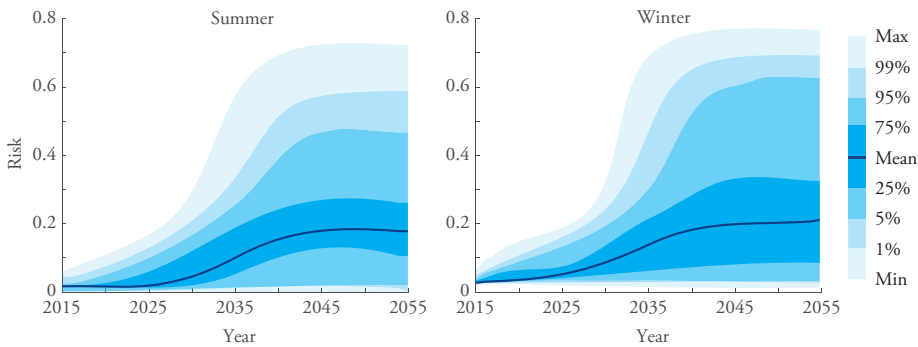
Figure 5.1: *Time evolution of the probability density of the risk in the network, for the summer (left) and winter (right) loading conditions*

From the figure 5.1 a clear increase in the risk as time goes by can be observed, assuming no network reinforcements are performed. The risk reaches it maximum value at the year 2045 and stabilises afterwards. When looking at the scenarios as defined in section 2.2 it can also be seen that most of the technologies will have reached their maximum penetration level at that time, only the economic growth and appliance efficiencies change significantly after this point. Both the summer and winter risk seem to start increasing more rapidly from the year 2026 onwards, within a few number of years the average risk increases by 0.15. This is comparable to 15% of all the feeders moving from a risk-free situation to a situation in which the risk is deemed too high to ensure operation within the branch current and nodal voltage limits. The replacement of 15% of the network within this limited period of time would be hard to achieve, therefore the estimation of when this increase in risk starts is important, so pro-active measures can be taken to reinforce the LV-network. The summer risk has a tighter distribution than the winter risk level. For the summer loading, PV technology is most likely

the single most important driver determining the risk, while for the winter loading this can be caused by multiple factors. Currently, the winter loading is the determining factor for the evaluation of the risk in the LV-network.

5.3 Geographical

The geographical areas where the risk becomes high or remains low are also of interest. As the risk has been calculated by using generic feeders extracted from the original network, all the feeders within the original network can be classified as one of the generic feeders. In this way, a geographical overview of the risk levels can be created. In figure 5.2 a map has been plotted and in this map, three variables are shown. In the top map, the average risk in the year 2035 is indicated, in the middle figure the current average energy use in 2015 [1] is depicted, while in the bottom figure the population [2] density is indicated.

From the figure, the geographical distribution of the risk for a future year can be seen. No obvious relation between the level of risk and the current energy use or the population density can be seen. Based on the data behind the map, the correlation between the population density and the risk has been calculated. This shows a correlation of 0.41 indicating that more densely populated areas have a higher risk. This is also what one would expect based on the correlation between the risk and the number of connections per LV-feeder. Based on the overview of the risk relative to other factors, the areas which should be reinforced first, as they have a higher energy dependency or more inhabitants, can be identified. In the figure, it can also be seen that most areas have a risk level lower than 0.3. This indicates that high-risk feeders are not necessarily geographically close together, as the average risk level is equal to 0.25.

5.4 Scenarios

The risk can also be split up based on the scenarios. As the scenarios are constructed as all possible combinations of the individual scenario drivers, the average risk can be determined on the basis of a single scenario driver. All the combinations of the other drivers are still taken into account, this allows for the determination of the risk of scenarios involving more technologies and penetration levels as well. In the following subsections, the main scenario drivers (PV, EV, Heat pumps and economic growth) are discussed individually based on their contribution to the overall risk. The risks induced by the other scenario drivers are insignificant by comparison and therefore not discussed in the following sections.

---

[1] Liander N.V., 2015
[2] ©Kadaster / Centraal Bureau voor de Statistiek, 2016

Risk
- No data
- 0 - 0.05
- 0.05 - 0.1
- 0.1 - 0.2
- 0.2 - 0.3
- 0.3 - 0.5
- >0.5

Yearly energy use
- No data
- <2500 kWh
- 2500 - 3000 kWh
- 3000 - 3500 kWh
- 3500 - 4500 kWh
- 4000 - 6000 kWh
- >6000 kWh

Population density
- <50
- 50-250
- 250-1000
- 1000-3000
- 3000-5000
- 5000-7500
- >7500
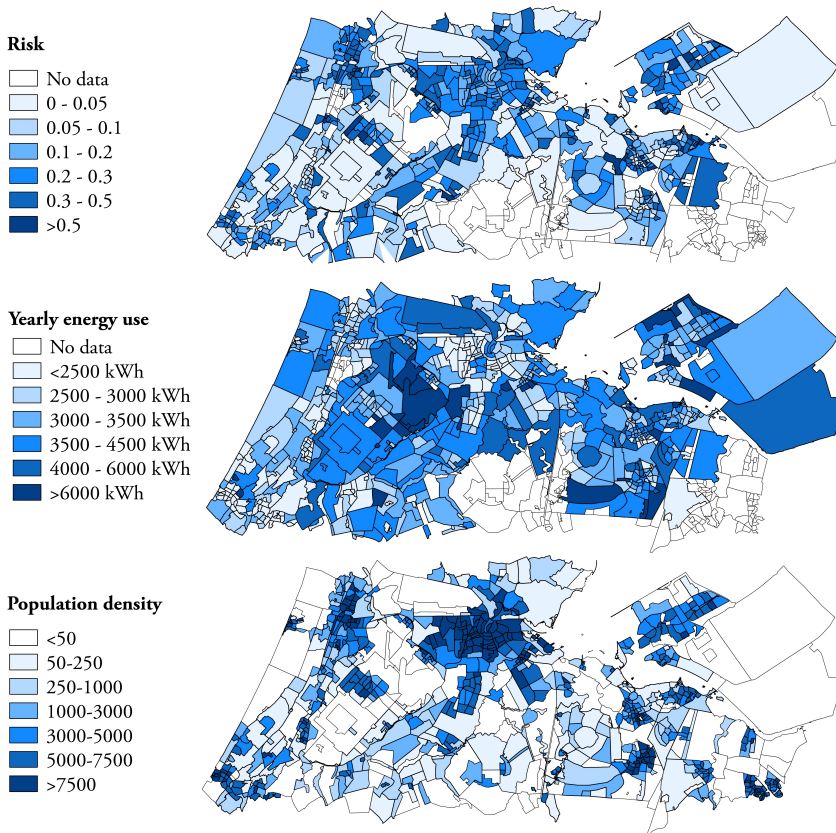
Figure 5.2: *Geographical overview of the adequacy of a portion of the Dutch LV-network (Amsterdam, Leiden, Almere) for 2035 (top) as well as the energy use and population density for 2015 (middle and bottom)*

5.4.1 Photovoltaics

First of all, a closer look is taken at the effects of the penetration of PVs on the risks within the LV-network. To show these effects, Fig. 5.3 has been created. In the left side of the figure, the relative risk for the three different PV scenarios is given. The relative risk is defined as the difference in risk level between the risk level of the complete set of scenarios and the risk level of a specific set of scenarios (e.g. high PV). The sum of the relative risks for a certain technology adds up to zero, with a negative relative risk indicating a lower risk with respect to considering all the scenarios. The right side of the figure shows how the level of risk within the LV-network is subdivided based on the scenarios other than PV, split out for the three PV scenarios (high, medium and low PV penetration).
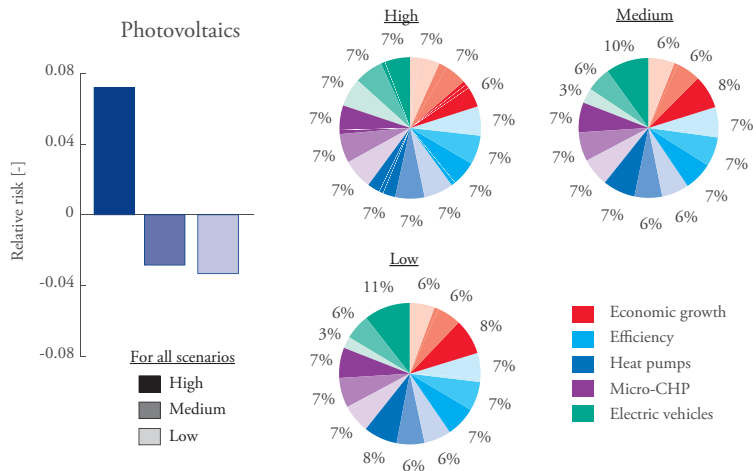
Figure 5.3: *Effects of photovoltaics on the network adequacy*

From the left side of the figure, it can be seen that the high PV scenario substantially increases the risks within the LV-network, while the medium and low scenarios both decrease the risks within the network. We conclude that small – and even medium – amounts of PV can be implemented within the network without inducing significant risks. Within the right side of the figure, it can be seen that, for the high PV scenario, the other scenario drivers all have a similar effect. Only a high level of economic growth induces a slight reduction in risk cost by the high penetration rate of PV. This can be explained by the increased number of appliances within a household which may consume power during the PV generation peak. The pie charts for the medium and low PV penetration scenarios show that the EV penetration is the main source of risk within these scenario sets.

5.4.2 Electric vehicles

For the penetration of EVs, the same type of figure is created as for the case of PV. The left side of Fig. 5.4 has only one difference, namely that the effect of direct charging in comparison to delayed charging is also indicated. The risk associated with direct charging is determined by starting the charging of the EV immediately when it returns to the household, while in the delayed charging cases, the EV charging does not commence until midnight.

Fig. 5.4 shows that direct charging has far higher risks associated with it, in comparison to the delayed charging. The figure also shows that even the medium scenarios significantly increase the risk within the LV-network compared to a low penetration scenario, even though the risk in the medium scenario is still lower compared to the average risk over all scenarios. This indicates that every increase in EV can already increase the risks within the LV-network,
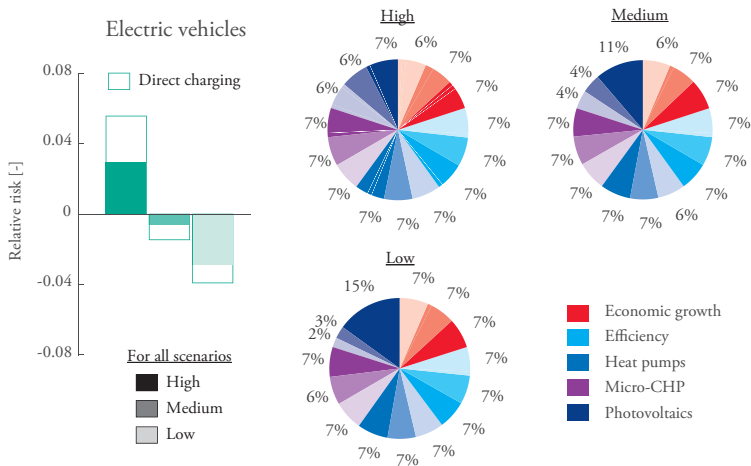
Figure 5.4: *Effects of electric vehicles on the network adequacy*

while the largest jump in the risk is associated with the move from the medium to high scenarios. This can also be explained by the far larger penetration rate reached between these scenarios (28% in comparison to 63%). The right side of Fig. 5.4 shows the effect of the PV on the risk level is increasing as the effect of the EV diminishes. The effect is larger than the other way around in Fig. 5.3, leading us to conclude that high PV-penetration levels generate the largest risks within the LV-network.

5.4.3 Heat pumps

For the penetration of heat pumps, a figure similar to the ones for PV and EV scenarios is made. Fig 5.5 shows the effects of the scenarios with an increased heat pump penetration on the risks within the LV-network. From the left side of the figure, it can be seen that the risk of the increasing penetration of heat pumps is much lower in comparison to the risks associated with the penetration of EV and PV. This is mainly because of the lower expected penetration rate in the high scenario of heat pumps (46%) when compared to EV (63%) and PV (82%). The right side of the figure shows that high penetration rates of both heat pumps and EVs induce small additional risks within the network, as the risk of the high EV penetration rate increases with the high heat pump penetration scenario.

5.4.4 Economic growth

The overview of the induced risks by different levels of economic growth is shown in Fig. 5.6. To keep the figure comparable to the others, only the very high, medium and very low
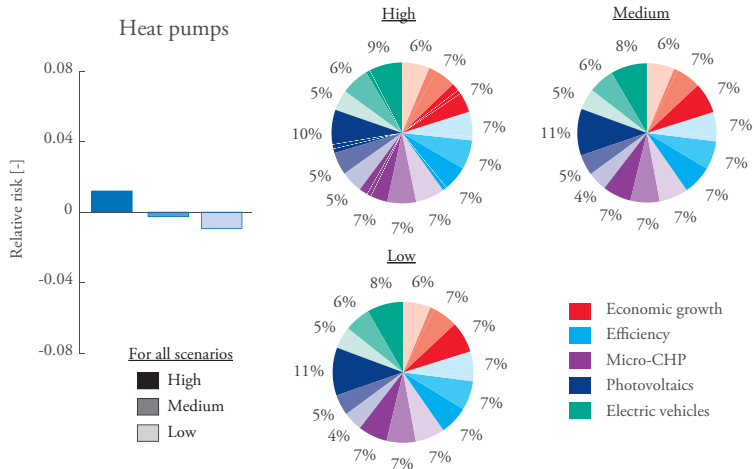
Figure 5.5: *Effects of heat pumps on the network adequacy*
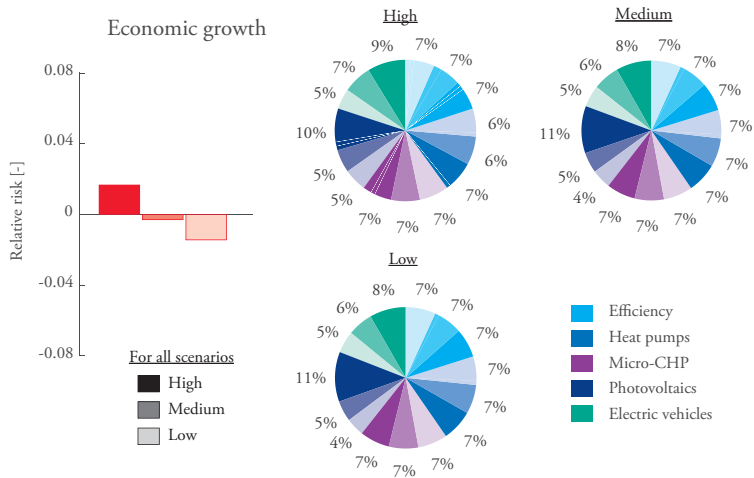
Figure 5.6: *Effects of the economic growth on the network adequacy*

scenarios of the economic growth as shown in Fig. 2.5 are depicted. The picture for the economic growth is comparable to the picture of the heat pumps. We note a significantly smaller impact of the economic growth on the LV-network adequacy compared to the impacts of PV and EV, and a correlation with the EV penetration.

5.5 Conclusions

The majority of the LV-network will not experience any problems, even if all the possible scenarios are considered. For more than three out of every four networks the risk never exceeds 0.5, thus the largest part of the LV-network would not need any reinforcing to be able to handle the changes introduced by the transition towards a sustainable energy supply. The number of consumers is the most important parameter for determining the risk within a certain LV-network. The feeder length also shows correlation with the risk level present in the LV-feeder, however far less than the number of connections. The longest feeders are often placed in rural areas and the feeder is in these cases often limited by the constraints on flicker and safety. When looking at these rural feeders with residential consumers a relatively small voltage deviation and loading are currently present, leading to little need for reinforcements. A large penetration of PV increases the risks within the LV-network the most, while small and even medium amount of PV can be implemented within the network without inducing significant risks. The increased penetration of EV is the second most important source of risk introduced by the transition towards a sustainable energy supply, however this risk can be for a substantial part mitigated by smart charging.

6

The Reinforcement of the LV-Network

The LV-network might not always be able to handle the changes introduced by the transition towards a sustainable energy supply, as shown in Chapter 5. The generic feeders were incapable of handling all the scenarios which describe the transition towards a sustainable energy supply. If an LV-network starts to operate outside the allowable voltage and current limits, the LV-network needs to be reinforced. For the determination of where the LV-network might need to be reinforced, the generic feeders can be used, while for the determination of the optimal reinforcements the actual network should be taken as starting point. In this chapter, new methods for the reinforcement or expansion planning of LV-networks are proposed, and their effectiveness is analysed. The currently applied method of waiting for the limits to be breached and applying a deterministic analysis when reinforcing the LV-network is not ideal when it comes to handling the requirements introduced by the transition towards a sustainable energy supply. In Chapter 5 it was shown that – for the Dutch case study – 36.2% of all the LV-networks are at risk to start operating outside the normal operating limits and would thus require reinforcements. In Chapter 5 it was also shown that the increase in risk can increase by a lot within only a couple of years. If these reinforcements are applied only when the LV-network is no longer operating within its limits, the number of replacements which have to be performed can vary significantly depending on the penetration rates of the different technologies. A more proactive approach with respect to the LV-network expansion planning is necessary. However, based on the current level of measurement data availability, a more proactive approach can be difficult to implement right away. Therefore in section 6.1, the benefits of additional measurement data in the LV-

network planning process will be discussed. The deterministic approach currently used for the network expansion planning is not designed for handling the uncertainties introduced by the transition towards a sustainable energy supply. The approach with respect to the LV-network planning, therefore needs to be changed from a deterministic approach to a more stochastic approach. How this can be done will be discussed in section 6.2. Finally, the possible reinforcement options are also developing to include on-conventional, "smart grid" technologies; these alternatives should be included in the LV-network planning optimisation. How this can be performed will be discussed in section 6.3.

6.1 Measurement data for LV-network planning

The measurements in the LV-network are currently often limited to a single yearly maximum power measurement at the MV/LV transformer. this introduces a lot of uncertainty for the LV-network, as the loading of the LV-network remains largely unknown. This complicates the planning of the LV-network as the current network loading needs to be estimated. As DNOs are already investing in an advanced metering infrastructure (AMI) mainly from simplified billing purposes, this infrastructure can be used to reduce the uncertainty of the loading of the network for the network expansion planning process. The AMI data can be used for a multitude of applications [69], from demand response to automating the meter reading process and more accurate LV-network expansion planning [70]. AMI data offers many opportunities for real time operation of the distribution network [71][72][73], however for the planning of the LV-network the available opportunities are not well documented. AMI data can be used to improve the load modelling [74] and load models based on AMI data can be implemented in the network planning process [75]. Privacy concerns about the use of AMI data require anonymisation. Depending on the implemented AMI communication infrastructure, a limited read-out frequency and/or measurement frequency of the AMI data may be required[76]. This limits the DNO in the amount of information which can be extracted from the AMI measurements or requires additional communication infrastructure investments[77].

To be able to make adequate choices about the extent of the implementation of an AMI, the benefits of the AMI data should be assessed. The valuation of the AMI data from a demand response perspective has already been performed [78]. The assessment of the automation of a business process like the surveying of conventional electricity meters can easily be assessed from the expenses of the DNO. The use of AMI data to accurately assess transformer loading has also been studied [79], as well as the creation of load forecasts based on AMI data[80]. The usefulness of the additional measurements to increase observability from an operational perspective, e.g. in state estimation, is a common research topic [81]. However, the data requirements from a long-term planning perspective are different. Preliminary research in using AMI data for the long-term network planning problem has been conducted [82]. An estimate of the value of AMI data from an LV-network expansion planning perspective is needed when determining the required characteristics of an AMI. From the point of view

of generation adequacy, studies on the value of additional measurement data have already been performed[83] [84], however for the LV-network expansion planning the effects of the additional measurement data still need to be quantified. A uniform network planning methodology which incorporates different types of LV-measurement data is therefore needed to assess the value of AMI measurements for the LV-network planning process. By adjusting the value of the applied loads in the LV-network expansion planning process the value of the measurement data can be determined.

Depending on the LV-network measurement approach, data with different levels of detail may be available. Based on the most logical locations of measurement devices, a number of different levels of data availability can be distinguished. These different levels are depicted in Figure 6.1.
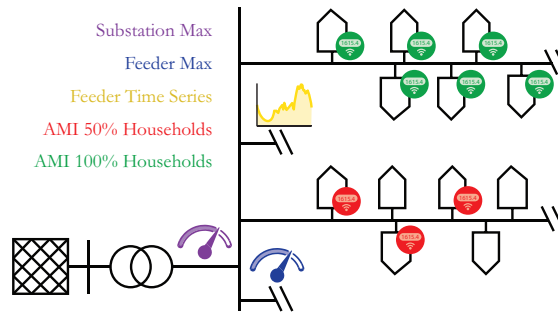
Figure 6.1: *Different levels of data availability in the low voltage network*

Figure 6.1 shows the possible locations of the measurement devices in the LV-network. The purple indicates the situation now, with just maximum current measurements at the MV/LV transformer. This can be implemented by assigning all the load the same value, but these values are drawn from a distribution of possible loads. The blue indicates the next step in implementing measurements, with the maximum current also measured at each of the outgoing LV-feeders; this reduces the variance of the distribution from which the equal load values are drawn. In addition to just measuring the maximum current at the LV-feeder, time-series measurements at the beginning of the feeder can also be employed, this is indicated by the yellow in the figure. If the time series approach is chosen, the actual loads of the households are randomly permuted over the households. The DNO can also choose to collect AMI data every two months for planning purposes. This option is evaluated at two different participation rates for consumers. The red in the figure indicates the situation if 50% of the consumers are willing to let the DNO use their energy usage data for improved network planning. With this option, 50% of the household have their real load, while the other 50% have an equal load based on the remaining MV/LV transformer loading. The green indicates the (ideal) situation if 100% of the consumers give their energy usage data to the DNO, all the loads are known to the DNO in this case.

The LV-network of Alliander has been used to compute the cost of the LV-network expansion planning with these different levels of data availability. The results show that an increase in measurement data can lead to €0-254 lower LV-network reinforcement costs per feeder. From the perspective of an AMI, the network cost can also be calculated with different percentages of households with AMI data available. The results show that 50% of the possible cost reduction can be achieved if 65% of the households have AMI data available, while if only 25% of the households make their AMI data available to the DNO, only a marginal cost reduction of 11% can be gained. Missing AMI data from a few households has limited effect on the achievable cost reduction as 90% AMI data availability corresponds to a cost reduction of 94%. The use of additional measurement data in the LV-network can and should thus be employed in the LV-network planning process, as additional benefits can clearly be gained.[1]

6.2 Network expansion planning

LV-network reinforcement (or network expansion) planning is performed by the DNO when the LV-network is deemed no longer able to supply a reliable voltage of sufficient quality or as capacity problems arise in the foreseeable future. The conventional approach to reinforcing or planning a new LV-network is reasonably straightforward. Generally, the current approach with respect to LV-network planning is based on minimising the investment cost (and sometimes the cost of the energy losses) [85][86][19]. The investment cost is evaluated with respect to a simple deterministic peak load with a constant load growth factor. The network is analysed for the situation in 30-40 years' time. The options of reinforcing the LV-network: adding an additional feeder, replacing (part of) a cable or adding an additional MV/LV substation are evaluated by an engineer, with the aid specialised software. This approach has worked so far but should be augmented to be able to most adequately handling the changes in loading introduced by the transition towards a sustainable energy supply. The process of the LV-network expansion planning is therefore analysed in this section and changes with respect to the handling of uncertainty (subsection 6.2.1) and the optimisation approach (subsection 6.2.2) are recommended, based on how the description of the uncertainties present in the LV-network as given in chapter 2.

6.2.1 Handling uncertainty

Reinforcing LV-networks is generally capital cost intensive, as the cost for the trenching is much larger than the cost of the actual LV-cable. When a DNO installs a new cable it wants to be sure that the cable would remain operational for the coming decades. The

---

[1]For an elaboration on the value of measurement data in the LV-network expansion planning the paper M. Nijhuis, M. Gibescu, and J.F.G. Cobben, "Valuation of Measurement Data for Low Voltage Network Expansion Planning" *Electric Power Systems Research* Vol. 151, Oct 2017, p. 59–67 can be consulted in Appendix A.6

technical lifetime of new LV-cables is so long that failure rate statistics collection is deemed uneconomical[87]. When reinforcements are required in the LV-network, the network needs to be analysed to ensure the chosen reinforcement option is robust enough. In order to do this, the DNO applies an expansion planning optimisation approach to the network. The main challenge with the evaluation of the LV-network is the large load uncertainty. This uncertainty consists of both the long-term load evolution as well as the volatile nature of the household load (including any local renewable sources such as rooftop PV). Therefore, the DNO should incorporate the load uncertainty into the LV-network expansion planning. How these uncertainties can be modelled and implemented within the confines of the network expansion planning problem should be defined.

For the distribution network, many expansion planning strategies under load uncertainty have been described. Most work has been done with respect to the inclusion of the uncertain output of distributed generation in the expansion optimisation process [88, 89, 90, 91]. Load uncertainty in these approaches is however not taken into account. The use of a fuzzy description of the load [92, 93, 94] or the type of load [95] is one of the ways to deal with the load uncertainty in the network planning. Load curves for different load magnitudes [56, 96, 97], and daily load curves [98],[99] are also employed to bound the uncertainty of the load. All these approaches will generate a limited amount of load uncertainty, which is suitable for modelling the household load for the MV network, as the loads are already aggregated to a certain extent, but not for analysis on the LV-network. Another way to deal with uncertainty is by using scenario analysis. This can be done in multiple ways, by incorporating scenario-based load growth in the conventional planning procedure [100] [101] or by adjusting the network optimisation under scenario-based load uncertainty [102]. The final future load can also be seen as a random variable, rather than a deterministic value [103][104]. Though these approaches are good for evaluating the future load from a system-wide perspective, the uncertainty on the level of the individual household loads as necessary for an LV-network planning exercise is not taken into account.

Household load volatility

In the LV-network, most connected loads are household consumers. Even though these consumers behave in fairly the same way when looking at a large group of consumers, their individual behaviour is highly volatile. From the perspective of medium and high voltage levels, the residential load can almost be seen as a deterministic parameter, as the number of households which are grouped at a single bus can easily be more than 200. However, for an LV-feeder this is not the case and the uncertainty should be taken into account.

With the introduction of smart meters, this uncertainty should be reduced for the existing network, as for each household a load curve would be available. The household composition in a neighbourhood is however constantly changing (people move away, family situations change, etc.). These changes would have a significant effect on the energy usage in the neighbourhood as the household energy consumption is primarily dependent on the

behaviour of the individuals who make up the household [105]. The use of measurement data to reduce this uncertainty is an adequate solution for the short time, though when looking further ahead the measurement data loses its applicability.

When it comes to the planning of the LV-network, the general approach to deal with this uncertainty in the load is to take the 95th or 99th percentile value of the distribution curve. The use of this method has one major drawback. All the household loads use the exact same peak value. However, through the distribution of the load over a feeder and the non-linearities in the relationship between power and voltage, the estimated voltage may not be at the same percentile value. This is because the effect of households at the end of a feeder on the maximum voltage deviation is much larger than for households at the beginning of the feeder. This generates a change in risk level between the percentile values of the load and of the resulting voltages.

In order to eliminate the change in the risk level due to the inherently non-linear mapping of the peak powers into nodal voltages, a stochastic approach to the evaluation of each individual household load should be employed. By employing a probabilistic load flow in the evaluation of the LV-network, a better risk assessment can be made. When a probabilistic power flow, as defined in section 4.3.1, is employed rather than a deterministic one, a probability density function will be obtained for the current in each branch and the voltage in each node[106][107]. The use of a probability density function allows for a more advanced approach with respect to modelling the volatility of the household load.

Long-term load evolution

In addition to the uncertainty due to the volatility of the household load, also the general trend with respect to the future load is uncertain. When reinforcing the LV-network, the implemented reinforcement option should last for at least the coming 30 years within the network. This is caused by the capital cost-intensive nature and long lifetime of underground cable networks, as explained in the introduction. The load of the households can be estimated based on scenario analysis as described in section 2.2.

These scenarios should be addressed by the network expansion planning approach, while incorporating the volatility of the household load. The Bottom-up Markov Chain Monte Carlo scenario-based method for generating the household load is therefore applied, as described in section 3.2.1. Based on the generated load curves, the distribution of possible household load values during conventional peak and PV-related peak times can be generated. These probability density functions would give a better estimation of the load than a simple peak value as they allow for variation in the generated scenario peak [108]. For the scenario options shown in Table 2.2 the peak load is shown in Fig. 6.2. In the figure also the typical times for a multi-stage investment approach are given, the initial time (T_0), the second investment time T_1 (after 10 years) and the final network time T_2 (after 30 years) are indicated.
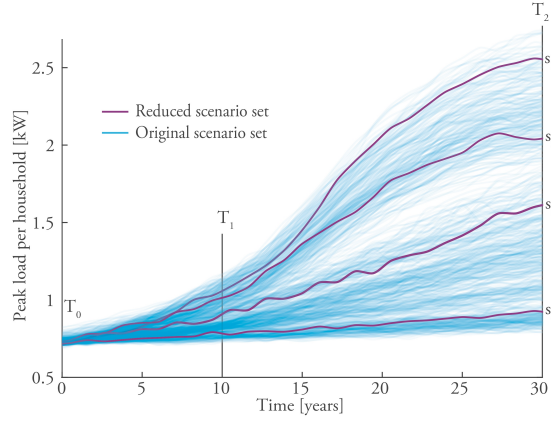
Figure 6.2: *Possible peak load scenarios for the household load and the four scenarios obtained through scenario reduction*

The figure is generated by using each unique combination of the different scenario drivers. There are 5 drivers, and each can take a low, medium or high value, resulting in 243 ($= 3^5$) scenarios. From the figure, the large spread among scenarios becomes apparent. The network cannot be analysed on all these possible scenario outcomes, as that would be computationally very expensive. Therefore, a subset of scenarios needs to be selected. The selection of the subset of scenarios is based on minimising the Wasserstein distance [109] between the original and reduced scenario subset. The Wasserstein distance can be seen as the minimum required transportation cost to move all the data from the original scenario set to the reduced subset. According to [110], this can be estimated by applying a k-medians heuristic [111] which leads to the following minimisation problem:

$$\min_{u \in \{1,\dots,N\}} \sum_{\substack{i=1 \\ i \neq u}}^{N} \mathcal{P}_i c(s_u, s_i) \tag{6.1}$$

where c is the cost to transport from an original scenario s_u to a scenario in the reduced scenario subset s_i and $\mathcal{P}_i$ is the probability of belonging to scenario s_i. The final subset of scenarios can be estimated by starting from the complete set of scenarios and performing a backwards scenario selection process. This entails applying the minimisation (6.1), generating the set of N-1 scenarios and repeating the selection procedure until the desired number of scenarios is obtained. This will generate a reduced set of k scenarios in which each of the scenarios will have a probability that is based on the probabilities of the closest scenarios from the original set. The approach is applied with the following metrics for each scenario: the average load, the peak load and the off-peak load, all three measured at both T_1 and T_2. In this thesis, a subset of $k = 4$ scenarios is used for the generation of the results. The resulting scenarios are also indicated in Fig. 6.2 as s_1, s_2, s_3 and s_4.

For the analysis of these different scenarios, it must be noted that scenarios are generated at a national level. This implies that if a certain scenario from the set of generated scenarios pans out, all of the reinforced networks will experience this scenario. Taking a risk-neutral expected value approach (as technically done with the local load uncertainty) is therefore not advisable as basically all the investments within the same scenario area have to be pooled and seen as a single investment. A more risk averse approach is therefore adopted. The different network reinforcement options are analysed based on the minimum regret over the selected scenarios. For the calculation of the minimum regret, the lowest cost per scenario is estimated by applying a deterministic perfect information approach to the network expansion problem to ensure equal comparison values when calculating the regret.

By applying both types of uncertainty – the household load volatility and the long-term load evolution – in the network expansion planning, the transition towards a sustainable energy supply can be incorporated when reinforcing the LV-network.[2]

6.2.2 Bilevel network expansion planning

When evaluating the reinforcement options, the DNO has to take into account the long economic lifetime of additional LV-cables ($>$30 years) and the negative asset recovery value (removing a cable which is no longer required comes at a cost for the DNO) for underground cables. The DNO, therefore, assesses the LV-network over a long horizon to ensure the network can handle almost all foreseen load changes. This LV-network expansion planning problem is, from an optimisation standpoint, a mixed integer non-linear problem. As the solution should span a large time horizon, multi-stage approaches are generally preferred, thus greatly increasing the complexity of the problem. The LV-network expansion planning is therefore often solved by using heuristic as opposed to classical optimisation techniques.

The main heuristic techniques which are applied for distribution network planning are simulated annealing [112], Tabu search [113], partial swarm optimisation [114], ant colony optimisation [115] and genetic algorithms [116]. These types of algorithms have been compared in a number of typical power system problems [117, 118, 119, 120]. These comparisons show that the actual implementation of a certain algorithm is more important than the actual type of algorithm which is used. A careful consideration of the description of the multi-stage LV-network expansion planning problem is thus required. The calculation of all the stages in a multi-stage approach is often done in a single integrated approach [90, 121, 122, 123], which, through the combinatorial nature of the problem, vastly increases the computational complexity of the problem. The solutions for a multi-stage integrated problem are not only the reinforcement options which are applied, but also the timing of the reinforcement options. Both the scheduling of the reinforcements [124] as well as the choice of which reinforcements are to be applied are already complex problems. From a

---

[2]For an elaboration on the handling of uncertainty the paper M. Nijhuis, P.M. Carvalho, M. Gibescu, and J.F.G. Cobben, "Modelling Load Uncertainty for Multi-Stage LV-network Expansion Planning Optimisation" (under review) can be consulted in Appendix A.7

practical point of view, only the investment in the first stage and an indication of the expected total cost is of interest. Therefore the problem can be split into two parts, the optimisation of the first investment and the optimisation of the subsequent investment decisions. By applying a bilevel optimisation the differences in the requirements with respect to the first stage versus the subsequent stages can be fully exploited. This kind of decomposition of the problem is commonly applied when distributed generation is included in the distribution network expansion planning through the use of an optimal power flow [125, 126, 127].

A multi-stage approach for assessing network reinforcement solutions is preferred over a deterministic single-stage approach. To determine how a multi-stage approach can be made computationally feasible, a closer look at why a multi-stage approach is preferable should be taken. The two main reasons are: the difference in magnitude of required subsequent reinforcements and the reduced set of subsequent reinforcement options available.

Depending on the chosen reinforcement option, the resulting network strength can differ significantly. If a single-stage approach is applied, the reinforcement should be sufficient to adequately handle the load at the evaluation time. Whether or not an additional incremental change in the load would lead to a violation of the constraints is not taken into account. A network capable of handling an additional load change would provide additional value; however, in the single stage approach, this is neglected.

When a certain reinforcement option is applied to a network, it changes the space of possibilities of future reinforcements. For instance, when depleting most incremental reinforcement options, the flexibility to deal with future load changes diminishes, while placing a new substation would lead to more incremental reinforcement options in the future. This should also be accounted for in order to accurately evaluate a certain reinforcement option.

Next to the above 2 issues, the use of a multi-stage approach also resembles closely how the network reinforcements are carried out in reality. The normal procedure when reinforcing LV-networks is to wait until the network becomes overloaded, experiences voltage problems or the cables are no longer reliable enough, and then take action. Assuming one will reinforce the network at set intervals is, therefore, a crude representation of the actual reinforcement procedure. When a reinforcement investment is applied at a later stage (i.e. T_1 to T_n) the network will be analysed at that time to determine the optimal reinforcements (e.g. the value of X_1 will be re-evaluated at T_1, instead of meticulously following the plan determined at time T_0).

The evaluation of the later stages of the best reinforcement options is thus mainly done to gain a better estimate of the expected value of the initial options over a longer period of time. The importance of the sequential investments $X_1, ..., X_n$ is reduced due to the time value of money and the uncertainty which exists about future required network strength and reinforcement options. In an LV-network expansion planning approach, it is, therefore, justified to focus on determining the first investment option most accurately. The proposed approach will focus on the first investment X_0 and determine the sequential investments
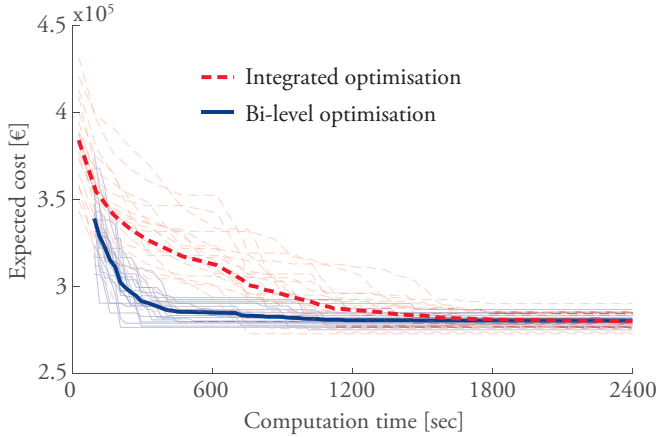
Figure 6.3: *The computational performance of the bilevel optimisation versus an integrated approach for LV-network planning*

$X_1, ..., X_n$ with a more computationally efficient calculation. For the solution of this bilevel optimisation problem featuring discrete decision variables, a genetic optimisation approach is very suitable [128].

To gain insight into whether the application of a bilevel optimisation approach is beneficial from a computational point of view without much loss of accuracy, the bilevel approach is compared to an integrated approach for the LV-network expansion planning. The data used for the optimisation is given in Appendix C. The convergence of the bilevel optimisation versus that of an approach wherein the first and sequential investments are integrated into a single level optimisation problem is plotted in Fig. 6.3.

Fig. 6.3 shows that both algorithms converge to the same level of the expected value, with the integrated approach generating a slightly lower expected value. As an advantage, the bilevel optimisation converges faster when comparing the computation time. The number of required generations before the solution has converged is much lower for the bilevel optimisation case, and this can already be seen from the initial starting point of both graphs. The starting point indicates how long the evaluation of a single generation takes, with the bilevel optimisation approach taking more than three times longer to evaluate a single generation. The bilevel optimisation takes longer for the evaluation of a single generation, as the nested optimisation problem has to be solved within that generation, nonetheless the overall computation time is shorter.[3]

---

[3]For an elaboration on the handling of uncertainty the paper M. Nijhuis, P.M. Carvalho, M. Gibescu, and J.F.G. Cobben, "Low Voltage Multi-Stage Network Expansion Planning Under Uncertainty; a bilevel evolutionary optimisation approach" (under review) can be consulted in Appendix A.8

6.3 Non-conventional planning alternatives

Next to the conventional ways of reinforcing the LV-network, the number of reinforcement alternatives can be expanded. There are many technologies which can assist in the reinforcement of the LV-network other than additional cables and substations. To give some initial insight in the applicability of these technologies, two of the most promising technologies which could have a large effect on the reinforcement LV-network are an On-Load Tap Changer (OLTC) and demand response are assessed. The application of demand response to defer LV-network investments will be discussed in the first subsection and in the second subsection, a closer look will be taken at how the OLTC can be implemented in the LV-network planning. If these technologies can reduce the cost for network reinforcements, they should still be implemented in the planning approach as described in Section 6.2.

6.3.1 Demand response implementation

Technology advancements in the ICT sector have resulted in smart metering infrastructures and new possibilities to control household appliances through the use of home energy management systems [129]. This will enable retail customers to alter their energy use based on real-time price signals and may result in monetary rewards when demand response (DR) programs are introduced [130] or when local renewable energy generation is fed back into the network [131]. The DNO can use these real-time peak signals to reduce the peak load, by giving incentives to the consumers for shifting their energy usages to off-peak hours. The DNO can pay for these incentives by the cost deferral of the otherwise necessary network reinforcements.

For household consumers in the Netherlands, the electricity prices are regulated through a constant electricity price or a two-tier time-of-use price (low price from 11PM till 7AM and normal price from 7AM till 11PM), plus a fixed network tariff. To assess the benefits of DR, and the shift to more time-varying pricing structures, multiple DR pilot projects have been started. The *Jouw Energie Moment*[132] pilot is operated by a distribution network operator with a goal to minimise network loading by peak shifting, while the *PowerMatching City*[133] pilot is evaluating a local market-based DR program. These two pilot projects can however, highlight the conflicting interests between actors in the unbundled electricity market. An energy supplier (ES) coordinated price-based DR scheme can introduce higher peak loads as users increase their loading during periods of low prices, while peak shifting based on network constraints could shift demand to times of higher energy prices for the consumer, however reducing the investments necessary in the distribution network.

The incentive function of the DR program is either based on the energy price or the network loading, or a combination of the two. For the ES-based DR program, the price plus incentive is simply replaced by a real-time energy price. To avoid continuous load changes on even the smallest of price changes, a dead band is introduced in the price signal. If there are

price deviations of less than 0.02€ from the average price no price signal is received by the households.

For a DNO-based DR program, a more complex incentive function needs to be defined. The incentive function should translate the energy consumption of a household to an incentive based on the impact on the network of that particular consumption profile. The DR approach is based on the loading of the households, as this information can be made available through the use of smart meters. These loading conditions are converted to a signal which indicates the relative amount of DR required to keep the voltages within the limits, as the voltage limits are violated before overloading occurs. Based on the reinforcement cost of the feeder this signal is converted to a price incentive for the customer. As the voltage limits should not be violated, the incentive requires an increasing rise when the voltage deviations get closer to the limits. If the customer reacts on the incentive from the DNO, the network will be restored to a position well within its limits. This reduces the incentive from the DNO and will then increase the loading of the households, causing the incentive to increase again. To avoid this oscillation the incentive from the DNO is only reduced after a time lag of 15 minutes and the rate in which the incentive can change is limited. Based on the above considerations, the proposed incentive functions for the ES-based and the DNO-based DR programs are given in the equations below:

$$\alpha(i)_p = \begin{cases} \overline{p_E} - p_E(i) - 0.02, & \text{if } |\overline{p_E} - p_E(i)| \geq 0.02 \\ 0, & \text{otherwise} \end{cases} \tag{6.2}$$

$$\alpha(i)_n = \begin{cases} \dfrac{(|1-V(i)|-0.04)^3}{T_{DR}} \dfrac{(C_{rf}-C_{rf}\cdot d)}{C_{lf}}, & \text{if } |V(i) - 1| \geq 0.04 \\ 0, & \text{otherwise} \end{cases} \tag{6.3}$$

Where d is the depreciation rate, $\alpha(i)_p$ is the incentive the households get from the ES and $\alpha(i)_n$ the incentive from the DNO, p_E is the real-time electricity price, $\overline{p_E}$ the yearly average electricity price, T_{DR} is a factor to ensure the incentives supplied by the DNO do not exceed the profits from the investment deferral $\left(T_{DR} = \sum\limits_{i=0}^{i_Y} |1 - V(i)| - 0.04 \right)$, i_Y the number of 15 minute intervals in a year, C_{rf} is the reinforcement cost and C_{lf} is a factor (between 1.02 and 1.05 depending on the network topology) to decrease the investment deferral with the expected reduction in losses when reinforcing the network. Examples of the incentive signals which the customer receives can be seen in Fig. 6.4b.

The simulations of the ES- and DNO-driven DR programs are done based on four selected feeders (from Appendix B feeders: 16, 47, 53 & 82), for one summer and one winter week, to include effects of seasonal variation in load as well as in electricity prices. The load is divided into flexible (EV, dishwasher, washing machine, and dryer) and non-flexible loads (all other loads). The flexible loads can be delayed for up to 24-hours, expect for the EV. The state of charge of the EV should be 100% by the time the EV leaves the household. An example of the incentive functions, the resulting load and the effect on the voltage deviations are given in Fig. 6.4 for a single feeder.

(a) *Network load*

(b) *Incentive function*

(c) *Minimum voltage*
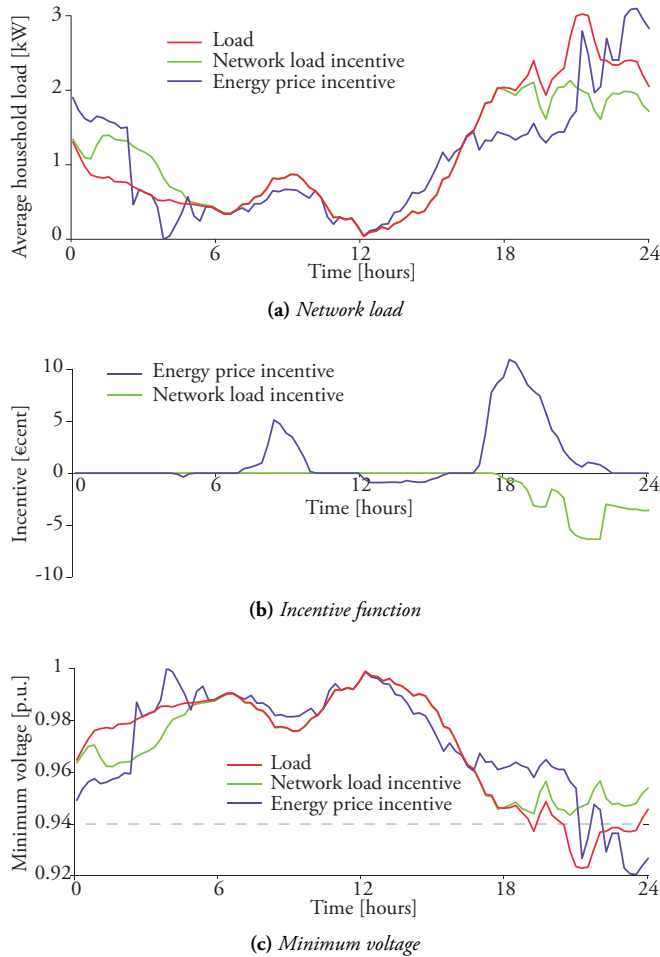
Figure 6.4: *The loading under the two DR programs and the incentive functions used to simulate the DR during a winter weekday for a single feeder*

The Fig. 6.4a shows the load of an average household and how this is affected by the DR programs. It can be seen that the loads are mainly shifted from the evening peaks to the night. The energy price incentive shows a much smoother profile compared to the incentive from the DNO, as can be seen in Fig. 6.4b. This is caused by the direct feedback loop in the incentive from the DNO, while for the incentive from the ES it is assumed that the changes in this LV-network are too small to have an effect on the price. The voltage profile in Fig. 6.4c shows that both the normal household load and the household load with DR based on the ES, violate the lower voltage limit of 0.94 p.u. (the actual voltage regulations allow
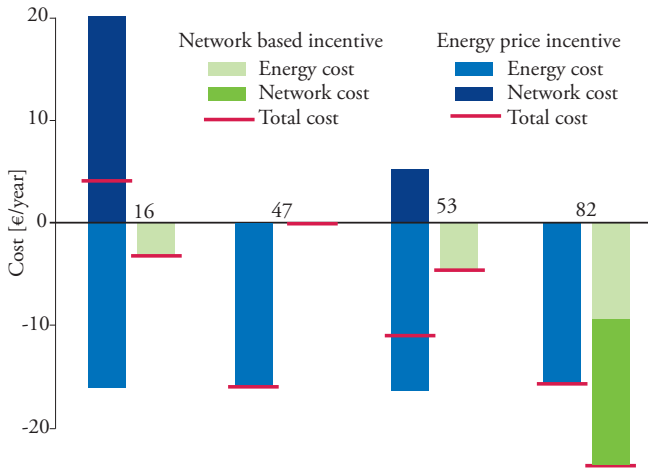
Figure 6.5: *Overview of the relative yearly energy cost per household for two DR programs with respect to the business as usual situation for the feeders A-D*

larger voltage deviations, but as the voltage deviations in the medium voltage network are not modelled, the 0.94 limit is used to account for this). The DR from the DNO point of view reduces the voltage deviations to within allowable limits. It can also be seen that the DNO will only start to give an incentive if the voltage goes below the 0.96 p.u. mark.

For the base case, the yearly energy cost is calculated for the four feeders without the use of any DR. This is simulated for every 15-minute value of both a winter and a summer week with all loads assumed to be non-flexible. This results in voltage deviations outside of the limits for feeder 82 (weakest feeder), the other feeders remaining within the limits. The same winter and summer weeks are then simulated with the inclusion of the demand response program of the DNO and the ES. The resulting yearly energy costs with respect to the base case are depicted in Fig. 6.5.

In the figure the difference in yearly energy cost with respect to the base case is depicted, so a negative cost means that the costs of the DR program are lower than without DR. The light bars indicate the contribution of the changes in energy cost due to the DR, the dark bars indicate the difference in network cost with respect to the base case and the and red bars indicate the combined effect of both the network and the energy cost on the yearly cost. The introduction of a DR program seems to have a generally positive effect for the customer, however the decrease in energy cost is limited. The exceptions are feeder 16 with a DR program from the ES point of view which is more expensive than the base case, and feeder 47 with a DNO controlled DR program, which results in negligible cost savings (< 1€ cent per year). For feeders 47 and 53, the ES- driven DR program is more profitable, while for feeders 16 and 82 the DNO-driven DR scores best. For the feeders 16 and 53, the cost of the network reinforcements increases as the ES-driven DR program increases the

peak loading. For feeder 53 this does not result in a higher total cost as the investment costs per connection are much lower due to the higher connection density. The DNO-driven DR program is capable of reducing the peak loading of feeder 82 enough, so that network reinforcement is no longer necessary. This leads to the largest total saving as the energy cost is also reduced due to the predominantly positive correlation between network loading and energy prices.

The cost savings from the perspective of a household are however relatively low and as shown in Fig. 6.5 the cost reduction with a DR scheme based on the network load, can oppose the cost reduction with a DR scheme based on real-time energy prices. This would require different prices for network access for each consumer, which is currently not allowed by regulation as it impedes non-discriminatory network access. These reasons make DR from a network planning perspective, not a very interesting option; only from a more operational point of view, DR can offer some benefits. When a large uncertainty over the local future load exists, the application of DR can delay the decision making, if a DR system is already in place.[4]

6.3.2 On-load tap changer implementation

Conventional MV/LV transformers are equipped with an on-load tap changer. With the transition towards a sustainable energy supply, the absolute magnitude of the voltage deviations in the LV-network can increase. To maintain these voltage deviations within the required limits (avoiding both under- and over-voltages), an OLTC can be installed. The use of an OLTC is therefore also often considered in smart grid concepts. The OLTC is currently getting applied more and more often, in LV-networks with a high amount of PV. The OLTC is not consistently and methodologically evaluated as one of the possible alternatives of reinforcements in the LV network. The implementation of an OLTC at the MV/LV transformer is now mostly performed on a case-by-case basis (mainly vacuum switch based) and for smart grid pilot projects (mainly power electronics based). To gain more insight into the value that an OLTC can have in the LV network, a framework to evaluate how the OLTC can be applied within the planning process is needed.

The use of an OLTC in distribution networks has been discussed before. However, the research has been mostly focussed on more advanced operational approaches of combining the control of the OLTC with the control of distributed generation [134][135] and/or electric vehicles [136], solving unbalance problems through individual phase voltage control [137], or using multi-objective optimisation approaches [138]. The control of the OLTC alone has also been a subject of research. While in [139] the control is based on state estimation from measurements within the distribution network, the work in [140] shows

---

[4]For an elaboration on the use of the implementation of demand response the paper M. Nijhuis, M. Babar, M. Gibescu, and J.F.G. Cobben, "Demand response: Social welfare maximisation in an unbundled energy market Case study for the low-voltage networks of a distribution network operator in the Netherlands" *IEEE Transaction on Industry Applications*, vol. 53, pp. 32-38, 2016 can be consulted in Appendix A.9

that a considerable part of these advantages can already be achieved by just measuring at the first bus. Different strategies for the application of an OLTC to mitigate over-voltages have been presented in [141]. Though these strategies can all increase the hosting capacity of the network, they are only assessed from an operational perspective. The implementation of an OLTC in the planning process has been touched upon by [142] in a specific case study. The generalisation of this approach and the implementation in the network planning optimisation problem however still need to be defined. Based on probabilistic load flow calculations for cases with and without an OLTC, the maximum admissible voltage in the underlying feeders can be determined for the different operating approaches. The admissible voltage is found to be independent of the network structure, allowing for an elegant implementation of the OLTC in the network planning optimisation. By relaxing the voltage constraint into a penalty function to be added to the objective function, the use of an OLTC can be evaluated within the same framework as the conventional network planning methods already introduced.[5]

6.4 Conclusions

The LV-network expansion planning is still often based on a deterministic approach. By switching to a probabilistic approach with the inclusion of both the uncertainty introduced by the forecasts of the residential load for the coming 40 years and the volatility of the household load itself, a more cost-effective approach to the network expansion planning can be designed. A probabilistic approach to the determination of the LV-network reinforcements has a far higher computational burden. A bilevel optimisation approach in which the first investment is determined in the upper optimisation loop and the subsequent investments are determined in the lower optimisation loop is found to be able to reduce the computational burden, while not discernible sacrificing on accuracy. The use of measurement data from smart meters can decrease the uncertainty in the LV-network planning and in turn lead to slightly lower LV-network cost. The implementation of demand response to defer LV-network investments shows only very small gains from a consumer perspective. The majority of the LV-network (>75%) has no too little direct benefits from demand response, while the potential profit for the DNO is very limited if the consumers need to be reimbursed for their change in demand.

---

[5]A case study on the implementation of an OLTC in the LV-network expansion planning, with different PV penetration scenarios shows that the OLTC can be used to defer network investments or as a complete smart grid alternative to network reinforcements, especially for situations with long, rural or suburban feeders where a scenario of high growth of PV is anticipated. More information can be found in the paper M. Nijhuis, M. Gibescu, and J.F.G. Cobben, "OLTC implementation in low voltage network planning" *in ISGT Europe*, 2017 can be consulted in Appendix A.10

7

Conclusions & Recommendations

The transition from a fossil-fuel based energy supply towards a sustainable energy supply is challenging for the planning and operation of the LV-network. In the previous chapters, the modelling, analysis and mitigation of the challenges in the long-term LV-network planning domain have been discussed. In this chapter, the conclusions which can be drawn regarding the approach of the DNO with respect to the LV-network planning are given. Thereafter some recommendations for future research directions are given.

7.1 Conclusions

To assess the effects of the transition towards a sustainable energy supply on the LV-network planning process the following research question was presented in chapter 1.

> *How should the distribution network operator change its approach to the design of the low voltage network in order to most efficiently handle the uncertainties in future loading introduced by the transition towards a sustainable energy supply?*

Based on the research presented in chapters 2-6 the sub-questions posed in chapter 1 can now be answered.

1) Which of the changes introduced by the transition towards a sustainable energy supply have an effect on the long-term adequacy of the low voltage network?

There are many changes that have an effect on the adequacy of the LV-network; the introduction of demand response, microgrids, PV, EV, power electronics, storage, changes in the conventional household load and the introduction of more communication possibilities within the power system all have an influence on the LV-network. The main influences of this transition from an LV-network point of view, the introduction of EV and PV, are both related to the residential electricity usage. The electrification of transportation, through the use of EV, can shift the energy usage of households away from fossil fuels towards electricity. While PV enables the local generations of renewable energy, leading to bidirectional energy flows. The residential electricity usage is thus changing significantly through the transition towards a sustainable energy supply. From a DNO point of view, the capacity and the power quality are the main requirements that are affected by this transition.

2) How will the residential load change in the coming decades and how can this be modelled?

The residential load will mainly change through the increased penetration of the technologies associated with the transition towards a sustainable energy supply. Through the increased use of PV, heat pumps and EV, the coincidence factor among the households will increase, leading to higher peak loads. However as appliances become more efficient, the overall energy usage and the peak load might also decrease. The residential load depends on how the future will unfold in terms of: penetration level of EV, heat pumps and PV, economic growth and appliance efficiency. A large scenario space can, therefore be constructed to reflect the possible changes in the residential load. The volatility of the household load is high and the behaviour of the loads at peak times can be considered independent, as the correlation during peak hours is only 0.01. Therefore the load curves must be constructed at the level of an individual household. To generate these load curves while allowing for the implementation of future load scenarios, an appliance based bottom-up Markov Chain Monte Carlo approach is required. Household load curves with a 1 min. time resolution with statistical characteristics validated with smart meter measurements can be generated in this way. Simplifications have to be defined to be able to use these household load curves for computationally more intensive LV-network studies. The use of a clustering approach can generate generic times series of the household load curves and the use of Gaussian mixture distributions can be used for peak load assessments.

3) Which metrics are important for the determination of the adequacy of the low voltage network and how can they be assessed?

For the point of view of the long-term adequacy of the LV-network, the most important aspect is the capacity of the LV-network in terms of both voltage range as well as current

carrying capacity. These two metrics should give the best indication of the adequacy of an LV-network. Assessing the capacity of the LV-network can be problematic because of its large size; the entire Dutch LV-network for instance consists of over 120 000 MV/LV substations with over 300 000 LV-feeders connected to them. In order to analyse this network, a generalised LV-network model needs to be developed. By generating this simplified model based on a fuzzy k-medians clustering approach, the link between the original network and the resulting generic LV-networks remains, allowing for the generalisation of the results obtained from a small set of generic feeders.

To assess the voltages and currents in a radial LV-network, a backwards-forwards sweep load flow can be applied. In the LV-network, the voltage angle can be neglected as R/X ratios in LV-networks tend to be high and the acquired voltage difference from the initial iteration is shown to be almost linearly related to the difference between the initial and final iteration voltage deviation. These two simplifications allow for the use of a Gaussian mixture distribution for the household load as an input to the load flow computation. Based on these computations the distribution of the voltage at each bus and power at each branch can be analytically determined, allowing for much faster probabilistic load flow computations.

> 4) *Is the current (Dutch) low-voltage network capable of handling the foreseen changes in the loading?*

The majority of the LV-network will not experience any problems, even if all the possible scenarios are considered. For more than three out of every four networks the risk never exceeds 0.5, thus the largest part of the LV-network would not need any reinforcing to be able to handle the changes introduced by the transition towards a sustainable energy supply. The number of consumers is the most important parameter for determining the risk within a certain LV-network. The feeder length also shows correlation with the risk level present in the LV-feeder, however far less than the number of connections. The longest feeders are often placed in rural areas and the feeder is in these cases often limited by the constraints on flicker and safety. When looking at these rural feeders with residential consumers a relatively small voltage deviation and loading are currently present, leading to little need for reinforcements. A large penetration of PV increases the risks within the LV-network the most, while small and even medium amount of PV can be implemented within the network without inducing significant risks. The increased penetration of EV is the second most important source of risk introduced by the transition towards a sustainable energy supply, however this risk can be for a substantial part mitigated by smart charging.

> 5) *How can low voltage network reinforcements best be implemented and can 'smart grid' solutions contribute to a more cost-effective low voltage network reinforcement?*

The implementation of LV-network reinforcements should be done based on a probabilistic evaluation of the LV-network; both the uncertainty introduced by the trends in the

residential load for the coming 40 years and the volatility of the household load itself should be included in the network expansion planning approach. A probabilistic approach to the determination of the LV-network reinforcements has a far higher computational burden than a deterministic one. Therefore, a bilevel optimisation approach, in which the first investment is determined in the upper optimisation and the subsequent investments are determined in the lower optimisation, should be employed to reduce the computational burden.

In addition, a number of 'smart grid' oriented approaches to the network expansion planning have been investigated. The use of measurement data from smart meters can decrease the uncertainty in the LV-network planning and in turn lead to slightly lower LV-network costs. The implementation of demand response to defer LV-network investments will only lead to very small gains from a consumer perspective even if all the benefits are passed on to the consumer. The majority of the LV-network (>75%) has no to little direct benefits from demand response, while the potential benefits from a DNO perspective are very limited as the consumers need to be reimbursed for their change in demand.

7.2 Recommendations

The work in this thesis gives an overview of how the LV-network can best be reinforced in the light of the transition towards a sustainable energy supply. There are however much more aspects of the LV-network planning in the light of the transition towards a sustainable energy supply which requires a more elaborate investigation. In this section, some of the possible future research directions are given.

- **Effects on higher voltage levels**
  The research in this thesis focuses on the LV-network. Changes in the LV-network can have a large effect on the adequacy and structure of the networks with higher voltage levels. Changes in the LV-network might also be resolved by changes in the MV-network which lead to less severe voltage variations at the LV-busbar. The optimisation with respect to the LV-network should take these effects on the higher voltage levels into account, however this would require much larger networks to be analysed, increasing the computational burden of the network expansion planning. How the MV-network can best be included in the LV-network expansion planning to correctly include the cost of changes to the MV-network remains to be analysed.

- **Demand response integration**
  The use of demand response for the deferral of LV-network investments has been analysed in this thesis. Even though it showed very limited potential benefits, demand response should still be integrated into the network expansion planning optimisation. The analysis in this thesis with respect to demand response is limited to a single future loading situation. When modelling the future load with the inclusion of uncertainty through the use of multiple scenarios, more detailed and possible different

results might be obtained. With such an approach, demand response might still offer significant benefits as the benefits of delaying a decision increase in more uncertain environments.

- **Network constraint handling**
  Within the optimisation of the network expansion planning the constraints on the node voltages and the branch currents are implemented as simple limits. Especially for the current limits, which are based on the heating of the cable, this is a crude representation of reality. For the temperature rise to become problematic, the peak load for which the network is assessed should be present for multiple consecutive time instances. Whether or not this would be the case should be investigated with successive load flows in addition to deciding how these thermal limits can best be implemented in a snapshot calculation of the LV-network.

- **Multi-objective planning**
  The planning problem formulation as given in this thesis is based on the transformation of all the objectives into a cost. This approach gives no indication if a slight change in altering the cost functions of these objectives would result in a very different optimal network. The transformation of the LV-network planning approach to a multi-objective approach can give a sense of the sensitivity of the obtained network to changes in the cost functions related to each of the objectives.

Bibliography

[1] Eurostat, "Database." [Online]. Available: http://ec.europa.eu/eurostat/web/energy/data/database

[2] Enexis, "Open data van Enexis," 2017. [Online]. Available: https://www.enexis.nl/over-enexis/open-data

[3] Eurelectric, "Power distribution in Europe - Facts and figures," Tech. Rep., 2013. [Online]. Available: http://www.eurelectric.org/media/113155/dso{_}report-web{_}final-2013-030-0764-01-e.pdf

[4] Autoriteit Consument & Markt, "Bijlagen x-factormodel regionaal netbeheer elektriciteit 2014-2016," Autoriteit Consument & Markt, Den Haag, Tech. Rep., 2014. [Online]. Available: https://www.acm.nl/nl/publicaties/publicatie/13306/Bijlagen-x-factormodel-regionaal-netbeheer-elektriciteit-2014-2016/

[5] NEDU, "Verbruiksprofielen," 2016. [Online]. Available: http://www.nedu.nl/portfolio/verbruiksprofielen/

[6] APERC, *A quest for Energy Security in The 21st Century Resources and Constraints*, 2007.

[7] J. J. C. Bruggink, "The next 50 years: Four European energy futures," Tech. Rep., may 2005.

[8] E. Veldman, "Power Play Impacts of flexibility in future residential electricity demand on distribution network utilisation," Ph.D. dissertation, dec 2013.

[9] Pwc, "Energie-Nederland Financial and economic impact of a changing energy market," Tech. Rep. March, mar 2013.

[10] C. Hellinga, "De energievoorziening van Nederland Vandaag (en morgen ?)," Tech. Rep., oct 2010.

[11] S. Teske, G. Onufrio, A. Gianni, E. Rainer, and H. Rahlwes, "Energy [r]evolution: A sustainable Netherlands energy outlook," Tech. Rep., may 2013.

[12] S. Moorman and M. Kansen, "Naar duurzaam wegverkeer in 2050 Een verkenning van mogelijke opties," Tech. Rep., nov 2011.

[13] DNV KEMA, "Rol van gas in nieuwe wijken," Tech. Rep. april, apr 2013.

[14] "Analyze Routekaart 2050," Tech. Rep., nov 2011.

[15] CPB, MNP, RBP, L. H. J. M. Janssen, V. R. Okker, and J. Schuur, "Welvaart en Leefomgeving: een scenariostudie voor Nederland in 2040," Tech. Rep., 2006. [Online]. Available: http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Welvaart+en+Leefomgeving{#}1

[16] W. Eichhammer, T. Fleiter, B. Schlomann, S. Faberi, M. Fioretto, N. Piccioni, S. Lechtenbohmer, A. Schuring, and G. Resch, "Study on the energy savings potentials in EU member states, candidate countries and EEA countries," Tech. Rep. March, mar 2009. [Online]. Available: http://ec.europa.eu/energy/efficiency/studies/efficiency{%}7B_{}{%}7Den.htm

[17] Centraal Bureau voor de Statistiek, "Huishoudens; grootte, samenstelling, positie in het huishouden," 2016. [Online]. Available: http://statline.cbs.nl/Statweb/selection/?DM=SLNL{&}PA=82905NED{&}VW=T

[18] Liander, "Beschikbare data." [Online]. Available: https://www.liander.nl/over-liander/innovatie/open-data/data

[19] P. van Oirsouw and J. F. G. Cobben, *Netten voor distributie van elektriciteit: Phase to Phase*. Arnhem: Phase to Phase, 2011.

[20] C. M. Colson and M. H. Nehrir, "An alternative method to load modeling for obtaining end-use load profiles," in *41st North American Power Symposium, NAPS 2009*, jan 2009, pp. 1–5.

[21] J. V. Paatero and P. D. Lund, "A model for generating household electricity load profiles," *International Journal of Energy Research*, vol. 30, no. 5, pp. 273–290, jan 2006.

[22] F. Wallin, C. Bartusch, E. Thorin, and E. Dahlquist, "Important Parameters for Prediction of Power Loads - A Bottom-Up Approach Utilizing Measurements from an Automatic Meter Reading System," in *IEEE PES PowerAfrica 2007 Conference and Exposition, PowerAfrica*, jan 2007, pp. 1–7.

[23] A. Grandjean, G. Binet, J. Bieret, J. Adnot, R. Edf, N. Supérieure, D. P. Mines, and B. Duplessis, "A functional analysis of electrical load curve modelling for some households specific electricity end-uses École Abstract," in *"6th International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL'11)*, vol. 6th, may 2011, pp. 1–24.

[24] A. Azadeh and Z. S. Faiz, "A meta-heuristic framework for forecasting household electricity consumption," *Applied Soft Computing*, vol. 11, no. 1, pp. 614–620, jan 2011.

[25] T. Zia, D. Bruckner, and A. Zaidi, "A hidden Markov model based procedure for identifying household electric loads," *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, pp. 3218–3223, jan 2011.

[26] NIBUD, "Energielastenbeschouwing: Verschillen in energielasten tussen huishoudens nader onderzocht," Utrecht, Tech. Rep., nov 2009.

[27] Å. . J. Holmgren, "Using graph models to analyze the vulnerability of electric power networks," *Risk Analysis*, vol. 26, no. 4, pp. 955–969, aug 2006.

[28] Y. Xu, A. J. Gurfinkel, and P. A. Rikvold, "Architecture of the Florida power grid as a complex network," *Physica A: Statistical Mechanics and its Applications*, vol. 401, pp. 130–140, may 2014.

[29] P. Schultz, J. Heitzig, and J. Kurths, "A random growth model for power grids and other spatially embedded infrastructure networks," *European Physical Journal: Special Topics*, vol. 223, no. 12, pp. 2593–2610, sep 2014.

[30] J. Dickert, M. Domagk, and P. Schegner, "Benchmark low voltage distribution networks based on cluster analysis of actual grid properties," *2013 IEEE Grenoble Conference PowerTech, POWERTECH 2013*, pp. 1–6, jan 2013.

[31] M. Nijhuis and J. Cobben, "Effects of power quality limits on LV-network design," in *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*. Belo Horizonte: IEEE, oct 2016, pp. 83–88. [Online]. Available: http://ieeexplore.ieee.org/document/7783341/

[32] M. Nijhuis, M. Gibescu, and J. Cobben, "Clustering of Low Voltage Feeders from a Network Planning Perspective," in *Cired 2015*, Lyon, France, jun 2015.

[33] P. Rao and R. Deekshit, "Energy Loss Estimation in Distribution Feeders," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1092–1100, jul 2006.

[34] M. Nijhuis, M. Gibescu, and J. Cobben, "Scenario analysis of generic feeders to assess the adequacy of residential LV-grids in the coming decades," in *2015 IEEE Eindhoven PowerTech*. IEEE, jun 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7232337/

[35] T. Williams and C. Crawford, "Probabilistic load flow modeling comparing maximum entropy and gram-charlier probability density function reconstructions," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 272–280, feb 2013.

[36] P. Siano and G. Mokryani, "Probabilistic assessment of the impact of wind energy integration into distribution networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4209–4217, nov 2013.

[37] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish-Fisher expansion," *Electric Power Systems Research*, vol. 89, pp. 129–138, apr 2012.

[38] J. Liang, D. D. Molina, G. K. Venayagamoorthy, and R. G. Harley, "Two-level dynamic stochastic optimal power flow control for power systems with intermittent renewable generation," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2670–2678, jul 2013.

[39] G. Verbč and C. A. Cañizares, "Probabilistic optimal power flow in electricity markets based on a two-point estimate method," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1883–1893, nov 2006.

[40] J. M. Morales and J. Pérez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1594–1601, nov 2007.

[41] G. Valverde, A. Saric, and V. Terzija, "Probabilistic load flow with non-Gaussian correlated random variables using Gaussian mixture models," *IET Generation, Transmission & Distribution*, vol. 6, no. 7, p. 701, jan 2012.

[42] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *International Conference on Power Engineering, Energy and Electrical Drives*, vol. 6, no. 2, p. 110, mar 2011.

[43] D. Villanueva, A. E. Feijóo, and J. L. Pazos, "An analytical method to solve the probabilistic load flow considering load demand correlation using the DC load flow," *Electric Power Systems Research*, vol. 110, pp. 1–8, jan 2014.

[44] H. Yu, C. Y. Chung, K. P. Wong, H. W. Lee, and J. H. Zhang, "Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 661–667, may 2009.

[45] K. Strunz, R. H. Fletcher, R. Campbell, and F. Gao, "Developing benchmark models for low-voltage distribution feeders," in *2009 IEEE Power and Energy Society General Meeting, PES '09*, jul 2009, pp. 1–3.

[46] IEEE PES Distribution System Analysis Subcommittee's Distribution Test Feeder Working Group. (2016, feb) Distribution Test Feeders. [Online]. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html

[47] A. L. Gibbs and F. E. Su, "On Choosing and Bounding Probability Metrics," *International Statistical Review*, vol. 70, no. 3, pp. 419–435, dec 2002.

[48] B. M. Pötscher, "Confidence sets based on sparse estimators are necessarily large," *Sankhya: The Indian Journal of Statistics*, vol. 71, no. 1 SUPPL. 1, pp. 1–18, jul 2009.

[49] D.-H. Choi and L. Xie, "Sensitivity Analysis of Real-Time Locational Marginal Price to SCADA Sensor Data Corruption," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1110–1120, may 2014.

[50] F. Tamp and P. Ciufo, "A Sensitivity Analysis Toolkit for the Simplification of MV Distribution Network Voltage Management," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 559–568, mar 2014.

[51] H. Ng, M. Salama, and A. Chikhani, "Classification of capacitor allocation techniques," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 387–392, 2000.

[52] D. K. Khatod, V. Pant, and J. Sharma, "Evolutionary programming based optimal placement of renewable distributed generators," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 683–695, may 2013.

[53] R. Tonkoski, L. A. C. Lopes, and T. H. M. El-Fouly, "Coordinated Active Power Curtailment of Grid Connected PV Inverters for Overvoltage Prevention," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 2, pp. 139–147, apr 2011.

[54] V. Calderaro, G. Conio, V. Galdi, G. Massa, and A. Piccolo, "Optimal Decentralized Voltage Control for Distribution Systems With Inverter-Based Distributed Generators," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 230–241, jan 2014.

[55] A. Gonzalez, F. M. Echavarren, L. Rouco, and T. Gomez, "A Sensitivities Computation Method for Reconfiguration of Radial Networks," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1294–1301, aug 2012.

[56] R. S. Rao, K. Ravindra, K. Satish, and S. V. L. Narasimham, "Power Loss Minimization in Distribution System Using Network Reconfiguration in the Presence of Distributed Generation," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 317–325, feb 2013.

[57] R. Kasturi and P. Dorraju, "Sensitivity Analysis of Power Systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-88, no. 10, pp. 1521–1529, oct 1969.

[58] W. Barcelo and W. Lemmon, "Standardized sensitivity coefficients for power system networks," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1591–1599, 1988.

[59] K. Christakou, J.-Y. LeBoudec, M. Paolone, and D.-C. Tomozei, "Efficient Computation of Sensitivity Coefficients of Node Voltages and Line Currents in Unbalanced Radial Electrical Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 741–750, jun 2013.

[60] R. Gurram and B. Subramanyam, "Sensitivity analysis of radial distribution network – adjoint network method," *International Journal of Electrical Power & Energy Systems*, vol. 21, no. 5, pp. 323–326, jun 1999.

[61] B. Sudret, "Global Sensitivity Analysis Using Polynomial Chaos Expansions," *Reliability Engineering and Systems Safety*, vol. 93, no. 7, pp. 964–979, 2008.

[62] A. Saltelli, M. Ratto, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis . The Primer*.   John Wiley & Sons, 2008.

[63] R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly, "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory," *The Journal of Chemical Physics*, vol. 59, no. 8, pp. 3873–3878, oct 1973.

[64] A. Saltelli, S. Tarantola, and K. P.-S. Chan, "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output," *Technometrics*, vol. 41, no. 1, pp. 39–56, feb 1999.

[65] I. Sobol', "Sensitivity Estimates for Nonlinear Mathematical Models," *Mathematical Modeling and Computational experiment*, vol. 1, no. 4, pp. 407–414, 1993.

[66] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker, "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, vol. 41, no. 1, pp. 1–28, jan 2005.

[67] M. M. Rajabi, B. Ataie-Ashtiani, and C. T. Simmons, "Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations," *Journal of Hydrology*, vol. 520, pp. 101–122, jan 2015.

[68] G. Blatman and B. Sudret, "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 95, no. 11, pp. 1216–1229, nov 2010.

[69] E. Liu, M. L. Chan, C. W. Huang, N. C. Wang, and C. N. Lu, "Electricity grid operation and planning related benefits of advanced metering infrastructure," in *2010 5th International Conference on Critical Infrastructure, CRIS 2010 - Proceedings*, sep 2010, pp. 1–5.

[70] J. Foosns, E. Tnne, and T. Pynten, "Power system planning in distribution networks today and in the future with smart grids," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, no. 1426, jun 2013, pp. 1426–1426. [Online]. Available: http://digital-library.theiet.org/content/conferences/10.1049/cp.2013.1233

[71] D. Bernardon, G. Iop, M. Gundel, and J. Fernandes, "A real-time operation analysis system for distribution networks," *Electric Power Systems Research*, vol. 78, no. 3, pp. 346–352, mar 2008.

[72] A. Alam, Jun Zhu, V. Frey, Ruili Zhao, Trong Hoang, S. Larson, S. Mundade, S. Ruelas, and A. Herasimava, "Improving Operational Planning study models with historical generation and load data," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, vol. 2016-July. IEEE, may 2016, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/document/7520038/

[73] A. A. Mohamed and O. A. Mohammed, "Real-time load emulator for implementation of smart meter data for operational planning," in *2012 IEEE Power and Energy Society General Meeting*. IEEE, jul 2012, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/6345619/

[74] X. Zhang, S. Grijalva, and M. J. Reno, "A time-variant load model based on smart meter data mining," in *2014 IEEE PES General Meeting | Conference & Exposition*, vol. 2014-Octob, no. October. IEEE, jul 2014, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/document/6939365/

[75] R. Brakken, T. Ait-Laoussine, S. Steffel, and D. Sikes, "Integrating Customer Load and GIS Data for Improved Distribution Planning & Operations (Updated)," in *2005/2006 PES TD*. IEEE, 2006, pp. 1056–1063. [Online]. Available: http://ieeexplore.ieee.org/document/1668648/

[76] W. Luan, D. Sharp, and S. Lancashire, "Smart grid communication network capacity planning for power utilities," in *Transmission and Distribution …*, apr 2010, pp. 1–4. [Online]. Available: 10.1109/TDC.2010.5484223{%}5Cnhttp://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5484223

[77] Wenpeng Luan, D. Sharp, and S. LaRoy, "Data traffic analysis of utility smart metering network," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/6672750/

[78] P. Siano, "Demand response and smart grids - A survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, nov 2014.

[79] Y. L. Lo, S. C. Huang, and C. N. Lu, "Transformational Benefits of AMI Data in Transformer Load Modeling and Management," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 742–750, apr 2013.

[80] B. P. Hayes and M. Prodanovic, "State Forecasting and Operational Planning for Distribution Network Energy Management Systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1002–1011, mar 2016.

[81] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement placement in distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 668–675, may 2009.

[82] G. Roupioz, X. Robe, and F. Gorgette, "First use of smart grid data in distribution network planning," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, jun 2013, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6683438

[83] J. Pan and S. Rahman, "Multiattribute utility analysis with imprecise information: an enhanced decision support technique for the evaluation of electric generation expansion strategies," *Electric Power Systems Research*, vol. 46, no. 2, pp. 101–109, 1998.

[84] A. H. van der Weijde and B. F. Hobbs, "The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty," *Energy Economics*, vol. 34, no. 6, pp. 2089–2101, jan 2012.

[85] H. Willis, *Power distribution planning reference book*. CRC Press, mar 2004, vol. 2.

[86] E. Lakervi and E. J. Holmes, "Electricity Distribution Network Design," p. 200, jan 1995.

[87] W. A. Thue, "Cable Performance," in *Electrical Power Cable Engineering*. CRC Press, dec 1975, pp. 395–400. [Online]. Available: http://dx.doi.org/10.1201/9781420001037.ch18http://www.crcnetbase.com/doi/abs/10.1201/b11507-22

[88] R. Hemmati, R.-A. Hooshmand, and N. Taheri, "Distribution network expansion planning and DG placement in the presence of uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 665–673, 2015.

[89] K. Zou, A. P. Agalgaonkar, K. M. Muttaqi, and S. Perera, "Distribution system planning with incorporating DG reactive capability and system uncertainties," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 112–123, 2012.

[90] C. L. T. Borges and V. F. Martins, "Multistage expansion planning for active distribution networks under demand and Distributed Generation uncertainties," *International Journal of Electrical Power and Energy Systems*, vol. 36, no. 1, pp. 107–116, jan 2012.

[91] M. E. Samper and A. Vargas, "Investment Decisions in Distribution Networks Under Uncertainty With Distributed Generation—Part II: Implementation and Results," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2341–2351, aug 2013.

[92] N. Kagan and R. Adams, "Electrical power distribution systems planning using fuzzy mathematical programming," *International Journal of Electrical Power & Energy Systems*, vol. 16, no. 3, pp. 191–196, 1994.

[93] I. Ramirez-Rosado and J. Dominguez-Navarro, "New Multiobjective Tabu Search Algorithm for Fuzzy Optimal Planning of Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 224–233, 2006.

[94] M. Skok, S. Krajcar, and D. Skrlec, "Dynamic Planning of Medium Voltage Open-Loop Distribution Networks under Uncertainty," in *Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems*, vol. 2005. IEEE, nov 2005, pp. 551–556. [Online]. Available: http://ieeexplore.ieee.org/document/1599323/

[95] M. A. Matos and M. T. de Leão, "Electric distribution systems planning with fuzzy loads," *International Transactions in Operational Research*, vol. 2, no. 3, pp. 287–296, 1995.

[96] M. Jalali, K. Zare, and M. T. Hagh, "A multi-stage MINLP-based model for sub-transmission system expansion planning considering the placement of DG units," *International Journal of Electrical Power and Energy Systems*, vol. 63, pp. 8–16, 2014.

[97] S. A. Taher and S. A. Afsari, "Optimal location and sizing of DSTATCOM in distribution systems by immune algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 60, pp. 34–44, 2014.

[98] M. Sadeghi and M. Kalantar, "Multi types DG expansion dynamic planning in distribution system under stochastic conditions using Covariance Matrix Adaptation Evolutionary Strategy and Monte-Carlo simulation," *Energy Conversion and Management*, vol. 87, pp. 455–471, 2014.

[99] B. Arandian, R.-A. Hooshmand, and E. Gholipour, "Decreasing activity cost of a distribution system company by reconfiguration and power generation control of DGs based on shuffled frog leaping algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 61, pp. 48–55, 2014.

[100] B. B. Souza, E. G. Carrano, O. M. Neto, and R. H. Takahashi, "Immune system memetic algorithm for power distribution network design with load evolution uncertainty," *Electric Power Systems Research*, vol. 81, no. 2, pp. 527–537, feb 2011.

[101] D. Wang, L. N. Ochoa, and G. Harrison, "Modified GA and data envelopment analysis for multistage distribution network expansion planning under uncertainty," *2011 IEEE Power and Energy Society General Meeting*, vol. 26, no. 2, pp. 1–1, jan 2011.

[102] E. G. Carrano, F. G. Guimarães, R. H. C. Takahashi, O. M. Neto, and F. Campelo, "Electric distribution network expansion under load-evolution uncertainty using an immun system inspired algorithm," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 851–861, jan 2007.

[103] E. G. Carrano, C. G. Tarôco, O. M. Neto, and R. H. Takahashi, "A multiobjective hybrid evolutionary algorithm for robust design of distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 645–656, dec 2014.

[104] N. C. Koutsoukis, P. S. Georgilakis, and N. D. Hatziargyriou, "A Tabu search method for distribution network planning considering distributed generation and uncertainties," in *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*.   IEEE, jul 2014, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/6960627/

[105] A. Druckman and T. Jackson, "Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model," *Energy Policy*, vol. 36, no. 8, pp. 3177–3192, jan 2008.

[106] A. M. L. da Silva, M. B. D. Coutto Filho, S. M. P. Ribeiro, V. L. Arienti, and R. N. Allan, "Probabilistic load flow techniques applied to power system expansion planning," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1047–1053, jan 1990.

[107] H. Yu, C. Y. Chung, K. P. Wong, and J. H. Zhang, "A chance constrained transmission network expansion planning method with consideration of load and wind farm uncertainties," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1568–1576, jan 2009.

[108] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, may 2010.

[109] G. Pflug, "Scenario tree generation for multiperiod financial optimization by optimal discretization," *Mathematical Programming*, vol. 89, no. 2, pp. 251–271, jan 2001.

[110] H. Heitsch and W. Römisch, "A note on scenario reduction for two-stage stochastic programs," *Operations Research Letters*, vol. 35, no. 6, pp. 731–738, nov 2007.

[111] M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys, "A Constant-Factor Approximation Algorithm for the k-Median Problem," *Journal of Computer and System Sciences*, vol. 65, no. 1, pp. 129–149, aug 2002.

[112] J. M. Nahman and D. M. Peric, "Optimal planning of radial distribution networks by simulated annealing technique," *Ieee Transactions on Power Systems*, vol. 23, no. 2, pp. 790–795, 2008.

[113] A. M. Cossi, L. da Silva, R. Lázaro, and J. Mantovani, "Primary power distribution systems planning taking into account reliability, operation and expansion costs," *IET Generation, Transmission & Distribution*, vol. 6, no. 3, p. 274, 2012.

[114] S. Ganguly, N. C. Sahoo, and D. Das, "A novel multi-objective PSO for electrical distribution system planning incorporating distributed generation," *Energy Systems*, vol. 1, no. 3, pp. 291–337, 2010.

[115] A. M. El-Zonkoly, "Multistage expansion planning for distribution networks including unit commitment," *IET Generation, Transmission & Distribution*, vol. 7, no. 7, pp. 766–778, jan 2013.

[116] V. Miranda, J. V. Ranito, and L. M. Proena, "Genetic algorithms in optimal multistage distribution network planning," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1927–1933, jan 1994.

[117] C. Gamarra and J. M. Guerrero, "Computational optimization techniques applied to microgrids planning: A review," *Renewable and Sustainable Energy Reviews*, vol. 48, pp. 413–424, 2015.

[118] R. Shi, C. Cui, K. Su, and Z. Zain, "Comparison study of two meta-heuristic algorithms with their applications to distributed generation planning," in *Energy Procedia*, vol. 12, 2011, pp. 245–252.

[119] P. Reche-López, N. Ruiz-Reyes, S. García Galán, and F. Jurado, "Comparison of metaheuristic techniques to determine optimal placement of biomass power plants," *Energy Conversion and Management*, vol. 50, no. 8, pp. 2020–2028, 2009.

[120] S. Kannan, S. M. R. Slochanal, and N. P. Padhy, "Application and comparison of metaheuristic techniques to generation expansion planning problem," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 466–475, 2005.

[121] S. Heidari, M. Fotuhi-Firuzabad, and S. Kazemi, "Power Distribution Network Expansion Planning Considering Distribution Automation," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1261–1269, may 2015.

[122] J. Salehi and M.-R. Haghifam, "Long term distribution network planning considering urbanity uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 321–333, nov 2012.

[123] M. Rahmani, R. Romero, and M. J. Rider, "Strategies to Reduce the Number of Variables and the Combinatorial Search Space of the Multistage Transmission Expansion Planning Problem," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2164–2173, aug 2013.

[124] E. Carrano, R. Cardoso, R. Takahashi, C. Fonseca, and O. Neto, "Power distribution network expansion scheduling using dynamic programming genetic algorithm," *IET Generation, Transmission & Distribution*, vol. 2, no. 3, p. 444, 2008.

[125] S. S. Al Kaabi, H. H. Zeineldin, and V. Khadkikar, "Planning active distribution networks considering multi-DG configurations," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 785–793, mar 2014.

[126] H. Falaghi, C. Singh, M.-R. R. Haghifam, and M. Ramezani, "DG integrated multistage distribution system expansion planning," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 8, pp. 1489–1497, oct 2011.

[127] G. P. Harrison, A. Piccolo, P. Siano, and A. R. Wallace, "Hybrid GA and OPF evaluation of network capacity for distributed generation connections," *Electric Power Systems Research*, vol. 78, no. 3, pp. 392–398, mar 2008.

[128] Y. Yin, "Genetic-Algorithms-Based Approach for Bilevel Programming Models," *Journal of Transportation Engineering*, vol. 126, no. 2, pp. 115–120, mar 2000.

[129] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, jan 2011.

[130] V. S. K. Murthy Balijepalli, V. Pradhan, S. A. Khaparde, R. M. Shereef, V. S. K. M. Balijepalli, V. Pradhan, S. A. Khaparde, and R. M. Shereef, "Review of demand response under smart grid paradigm," in *2011 IEEE PES International Conference on Innovative Smart Grid Technologies-India, ISGT India 2011*, jan 2011, pp. 236–243.

[131] M. A. A. Pedrasa, T. D. Spooner, and I. F. MacGill, "Coordinated Scheduling of Residential Distributed Energy Resources to Optimize Smart Home Energy Services," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 134–143, jan 2010.

[132] (2015, feb) Jouw Energie Moment. [Online]. Available: http://jouwenergiemoment.nl/

[133] (2015, feb) PowerMatching City. [Online]. Available: http://www.powermatchingcity.nl/site/pagina.php?

[134] K. M. Muttaqi, A. D. T. Le, M. Negnevitsky, and G. Ledwich, "A Coordinated Voltage Control Approach for Coordination of OLTC, Voltage Regulator, and DG to Regulate Voltage in a Distribution Feeder," *IEEE Transactions on Industry Applications*, vol. 51, no. 2, pp. 1239–1248, mar 2015.

[135] D. Ranamuka, A. P. Agalgaonkar, and K. M. Muttaqi, "Online voltage control in distribution systems with multiple voltage regulating devices," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 617–628, mar 2014.

[136] M. A. Azzouz, M. F. Shaaban, and E. F. El-Saadany, "Real-Time Optimal Voltage Regulation for Distribution Networks Incorporating High Penetration of PEVs," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3234–3245, nov 2015.

[137] J. Hu, M. Marinelli, M. Coppo, A. Zecchino, and H. W. Bindner, "Coordinated voltage control of a decoupled three-phase on-load tap changer transformer and photovoltaic inverters for managing unbalanced networks," *Electric Power Systems Research*, vol. 131, pp. 264–274, oct 2016.

[138] J. R. Castro, M. Saad, S. Lefebvre, D. Asber, and L. Lenoir, "Optimal voltage control in distribution network in the presence of DGs," *International Journal of Electrical Power and Energy Systems*, vol. 78, pp. 239–247, dec 2016.

[139] S. N. Salih and P. Chen, "On coordinated control of OLTC and reactive power compensation for voltage regulation in distribution systems with wind power," *IEEE Transactions on Power Systems*, vol. Accepted f, pp. 1–10, nov 2015.

[140] C. Reese, C. Buchhagen, and L. Hofmann, "Enhanced method for voltage range controlled OLTC-equipped distribution transformers," in *IEEE Power and Energy Society General Meeting*, jul 2012, pp. 1–8.

[141] C. Long and L. F. Ochoa, "Voltage control of PV-rich LV networks: OLTC-fitted transformer and capacitor banks," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 4016–4025, apr 2016.

[142] A. Navarro-Espinosa and L. F. Ochoa, "Increasing the PV Hosting Capacity of LV Networks : OLTC-Fitted Transformers vs . Reinforcements," in *IEEE PES Innovative Smart Grid Technologies (ISGT-America)*, feb 2014, pp. 1–5.

A

Papers

Assessment of the impacts of the renewable energy and ICT driven energy transition on distribution networks

M. Nijhuis [a,*], M. Gibescu [a], J.F.G. Cobben [a,b]

[a] Electrical Energy Systems, Eindhoven University of Technology, Den Dolsch 2, Eindhoven, The Netherlands
[b] Liander N.V., Utrechtseweg 86, Arnhem, The Netherlands

ARTICLE INFO

ABSTRACT

The shift to more renewable electricity generation, electrification of heating and transportation and the rise of ICT and energy storage lead to changes in the distribution of electricity. To facilitate the transition towards a clean sustainable power system distribution network operators are required to act proactively to these changes. To optimally capitalise on these changes it is imperative to have a clear and complete overview of the main developments and their effects on the distribution network. The technical, social, regulatory and economical effects of the renewable energy and ICT driven energy transition are discussed based on requirements which the distribution network should comply with. The discussed technologies all have effects on multiple requirements for the distribution network and the mitigation of unwanted effects should be assessed on these areas simultaneously. This can ensure that the future distribution network is best equipped to deal with the renewable energy and ICT driven energy transition. The overview of the requirements shows that the capacity, regulation and power quality are the main requirements which are affected by the renewable energy and ICT driven energy transition.

© 2015 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author.
  E-mail address: m.nijhuis@tue.nl (M. Nijhuis).

1. Introduction

The electrification of transportation and heating has the potential to make the energy usage of households more sustainable. In combination with the shift from consumers to prosumers through locally generated renewable energy like photovoltaics, the energy system is getting more sustainable. This transition towards a more sustainable energy system needs to be facilitated by the distribution network operator (DNO). Developments like the possibility of the implementation of storage systems into the electricity network and the developments in the ICT (information and communication technology) sector, advanced metering schemes are becoming more profitable and real-time operational actions in distribution networks are becoming possible. The distribution network no longer needs to be operated as a single system, but smaller parts of the network can be autonomously operated as microgrids. These developments are changing the way the distribution network can be operated and planned.

There has been much research performed to gain more insight into the changes which these developments individually bring to the distribution network. The power system can be seen as a complex system [1]. An analysis of all the individual effects of the renewable energy sources (RES) and ICT-driven energy transition might thus give significantly different results than the analysis of the combined effects. In order to be able to perform a study on the combined effects of the RES and ICT-driven energy transition the generation of a clear overview of the impacts of this transition is necessary.

In this paper an overview of the effects of the RES and ICT-driven energy transition is presented. To structure this overview, the requirements of the distribution network are used to assess the effects of the RES and ICT-driven energy transition. These requirements are discussed in Section 2. As the requirements have been established, the effects of the energy transition are discussed in Section 3 based on the driving technologies: Electric vehicles, energy storage, demand response, ICT, photovoltaics, power electronics and the changes and electrification of household loads. After these technologies have been discussed, the conclusions with respect to the planning of the distribution network are presented in Section 4 as well as directions for further research.

2. Requirements

The effects of the energy transition for the distribution network operator are multifaceted: Technical, social, regulatory and economical effects are all to be expected. To be able to generate a clear overview of these effects, they are framed according to the requirements of the distribution network. An overview of these requirements is given in this section. This overview of the requirements is used to structure the discussion of the effects of the energy transition.

To determine the adequate requirements, the function of the distribution network is chosen as starting point. The function of the distribution network is to provide access to safe electrical power of an adequate quality at the lowest possible cost. The requirements which can be derived from the function of the distribution network (as with most other critical infrastructures) are usually classified based on the four A's: availability, affordability, accessibility and acceptability [2]. To allow for a more detailed and structured evaluation of the developments in the distribution network these four requirements are sub-divided into eight categories: availability, capacity, cost, regulation, power quality, social context, safety and market facilitation. These requirements are discussed in more detail below.

Availability: Availability for the distribution network is defined as an uninterrupted voltage at the point of connection with the customer. The main reasons for interruptions in the voltage supply are: maintenance, overloading of components and short-circuits.

Capacity: The capacity of the distribution network translates to how much current the costumers can draw from the distribution network without overloading. The capacity in the distribution network is based on the limited by the maximum current which is contracted and the coincidence factor of users in the same area.

Cost: The capital and operational cost of the DNO should be as low as possible whilst fulfilling the other requirements. The cost

for the DNO can be split into four categories: expansion/replacement cost, impairment cost, energy losses and maintenance.

Regulation: The distribution network is characterised by its capital cost dependent nature (long payback periods and low operational cost) and its non-rivalry nature (access cost are to a large degree dependent on whether the good is consumed within the vicinity). This marks the distribution network as a natural monopoly, which needs regulated to achieve cost-efficient performance.

Power Quality (PQ): The voltage at the point of connection should have sufficient quality to ensure its compatibility with the equipment of the customer. Harmonic distortion, asymmetry, transients, frequency level, flicker, voltage swells/dips and steady-state voltage level are the main phenomena which define power quality.

Social Context: The distribution network operates within the context given by its service area. This social context imposes a number of external boundary conditions on the distribution network, such as the integration of the distribution network within the build environment or the prioritising of sustainable resources.

Safety: The distribution of electricity should be safe, for the DNO as well as for the end-users. Which mainly means the provision of adequate earthing at the point of connection and when maintenance on the distribution network is performed.

Market Facilitation: In order to have a functioning competitive electricity market, power exchange between producer, supplier, and end-user should be possible. The distribution network needs to facilitate this marketplace by physically connecting the market parties.

2.1. Overview

The requirements discussed above can interact with each other. For the analysis of the distribution network as a complex system these interactions are important. An overview of the interactions between these requirements is given in Fig. 1. The requirements which are strongly linked are connected by the blue lines. The interaction between availability and regulation for instance is generated through quality regulation which compliment or penalise DNO's based on their average outage duration per costumer. For more information about the interactions of the requirements
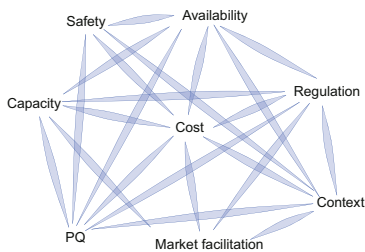

Fig. 1. Overview of the requirements of the distribution network and their interactions. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
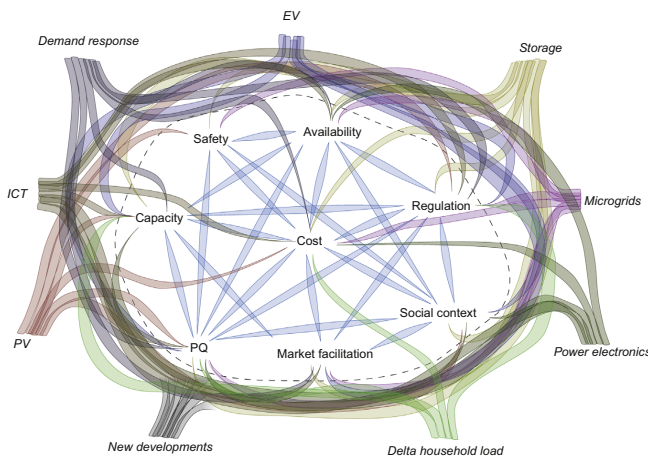

Fig. 2. Overview of new developments and the requirements with whom they interact in the distribution network.

when it comes to the planning of distribution networks [3–5] can be considered.

3. Developments

There are many developments which have an impact on the requirements of the distribution of electricity. The introduction of electric vehicles, energy storage, demand response, ICT and power electronics within the distribution network, photovoltaics, as well as changes in the household load are the main developments of the RES and ICT driven energy transition. Most of these developments have been separately studied, however as there are inter-actions between the requirements (as shown in Fig. 1) a complete overview of the effects of these developments is required. The qualitative effects on the distribution system of these developments are depicted in Fig. 2.

The figure shows the requirements and their interactions in the middle. Around these requirements the developments are de-picted. If a development has a direct impact on a requirement a line is shown between the development and the affected require-ment. In order to generate a more clear picture the effects of the developments and their effects on the distribution network are discussed in more detail in the following sections.
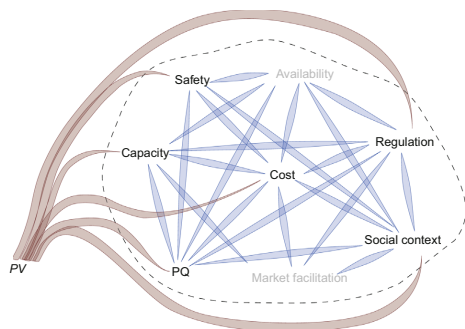
Fig. 3. Overview of the requirements which are affected by an increase of PV in the distribution network.
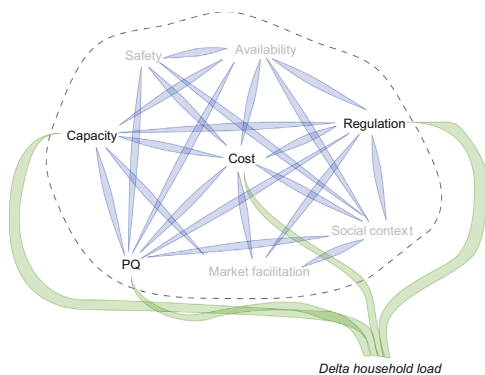
Fig. 4. Overview of the requirements of the distribution network which are affected by changes in the household load.

3.1. Photovoltaics

The main source of distributed generation (DG) at the low voltage level for the future seems to be rooftop photovoltaics (PV). The introduction of micro-CHP (combined heat and power) at household level is not thriving and windturbines are often con-nected at the HV level or on a separate feeder from the high voltage (HV) to medium voltage (MV) substation. Therefore it is assumed that PV will be the largest source of distributed genera-tion in the distribution network. In Fig. 3 an overview is given of which requirements affected by the increase of PV in the distribu-tion network.

The lines from PV to the requirements indicate a relationship between the requirement and the introduction of PV. These relationships are discussed per requirement in more detail below. The greyed out requirements in the figure are not directly affected by the introduction of PV, however through interactions with the other requirements the introduction of PV could also have an effect on the availability and market facilitation.

3.1.1. Power quality

The introduction of PV could lead to problems with the voltage level in the distribution network as well as form a new source for flicker.

Voltage deviations at the point of connection due to the voltage rise induced by the reversed power flow yield no problems for penetration levels of up to 0.5 kVA per household [6]. Though a standard household connection can accommodate PV-systems of upto 17 kVA, most PV-systems do not exceed 2 kVA (based on the 2013 data from the Dutch DNO Liander). On rural feeders however problems with the voltage rise could occur [7] as the grid impedance at the point of connection can be much higher than in non-rural cases.

An increase in flicker due to cloud induced PV fluctuations is possible, dependent on the local weather conditions and the impedance at the point of connection. As shown by seemingly contradictory studies, in [8] there was only very limited influence of PV on the level of flicker while in [9] a contribution larger than the local flicker limits was discovered.

3.1.2. Safety

Reverse power flows in the distribution network as well as the possibility of islanded operation can pose a safety risk.

Reverse power flows in some parts of the network could generate overloading of cables which would not be noticed by the protection devices, as protection relays usually only rely on the current at the beginning of a feeder [10]. This blinding of the protection relays could introduce unsafe situations.

Unintentional islanding is also a safety hazard with DG. As PV can keep parts of the network energised that have been isolated. Which makes it more difficult to assess for a technician whether or not the network is safe to work on. The current grid-tied inverter are required to trip at certain frequency or voltage deviation [11], which minimises the risk of islanding.

3.1.3. Capacity

PV might require additional capacity as the coincidence between peak production of different PV-systems is very high.

For PV installed on the rooftops of houses the coincidence factor of PV in a single neighbourhood is close to unity [12]. The diff-erences in orientation between houses and the angle of the roof creates the differences in PV power output between the houses. The network capacity planning for PV should not be based on the current coincidence factor but take into account the full rated power of all PV installations. The rated power of PV systems which

Fig. 5. Overview of the requirements which are affected by the introduction of storage systems into the distribution network.

are installed in the low voltage (LV) network is for a large part ($> 50\%$ of the installations are 2 kVA or smaller) small enough to not require any added capacity compared to the current design criteria. However for future scenario's with larger PV installations, the capacity of the LV-network might no longer be sufficient.

3.1.4. Cost

The placement of distributed generation has an impact on the energy losses in the distribution network.

An increase in PV could lead to a significant change in the cost of the losses, both higher or lower depending on the penetration level. For an UK MV-network it is shown that coordinating the DG placement the losses could be reduced by 20% [13]. As the DNO is not allowed to influence the DG placement, the most efficient combination of DG and the network design will not be achieved within the current regulatory framework.

3.1.5. Regulation

The reimbursement for adjusting the distribution network to facilitate PV integration is lacking in the current regulatory framework.

As most PV-units will be installed at existing LV-connections no connection charges or very limited shallow connection charges apply. The network operator would not be reimbursed for the network reinforcements which may occur due to the connection of PV. This leaves the network operator with no way of covering the cost for the reinforcements except by socialising the cost over all connections, which is from a PV placement and cost distribution point of view not preferable [14]. The inclusion of location based incentives in the regulation is not embraced, as it could reduce the affordable and non-discriminatory access to energy.

3.1.6. Social context

The connection of PV and other distributed generation should have priority due to the benefits which it generates for society [15]. The most important benefits are the absence of greenhouse gas emissions during electricity production and the increase energy security as the dependence on fossil fuel imports decreases.

3.2. Household load

In the coming years there are many developments which affect the household load. The proliferation of electric vehicles, the

introduction of home energy management systems and the electrification of heat generation are all examples of developments which will alter the household load. Next to the changes in the magnitude and time evolution of the loading of a household, the characteristics of the loads in the household will also change and the load profile can also be affected by societal trends such as an increase in working from home. These changes affect the power quality, capacity, cost and regulation requirements, as shown in Fig. 4.

3.2.1. Power quality

The change in household load could introduce change in the effect of poor power quality on the household appliances as well as reduce system stability.

The change in household load could result in a change in immunity to certain power quality effects. The phasing out of the incandescent light bulbs and their replacement with LED or compact fluorescent lighting for instance increases the flicker immunity [16], however if dimmers are applied to these lamps their flicker immunity levels decrease significantly [17].

The household load is in an increasing amount DC oriented, due to increasing penetration of consumer electronics, which indicates a change to predominantly constant power loads. These constant power loads have a negative differential impedance, meaning a decrease in voltage will lead to an increase in current, which could destabilise system voltage and frequency and increase voltage fluctuations and flicker [18].

3.2.2. Capacity

A change in the household load can introduce changes in required capacity as the peak load or the coincidence factor can rise.

New loads will not only have a large impact on the total household load level, either increasing or decreasing it, but also on the load profile. LED lighting for instance could generate a reduction in average load and peak load [19], while the introduction of many loads connected via power electronic converters has led to an increase in standby (base) load. These changes in the load curve have an effect on the required capacity and the asset utilisation.

The electrification of household energy use could lead to the increased usage of electricity especially for heating and transportation purposes. These types of loads (when present) are a substantial part of the total household energy consumption and have a relatively high coincidence factor of 0.5–0.7 [20,21] meaning a much higher feeder and transformer capacity is required when these types of loads are present in a neighbourhood.

3.2.3. Cost

If changes in the household load generate a higher peak load, or an increase in coincidence factor an investment to reinforce the distribution network could be required. The tariffs the DNO may charge are however based on gradual investments and the DNO has no method to recover these cost except by raising the tariffs.

3.2.4. Regulation

If the changes in household load generate more diversity among users, the household connection charging could become untenable.

Difference in load between two households could increase due to the differences in electrification, lifestyle and wealth. The current LV network connection charges allow for a large difference in loading (< 17 kVA peak) while maintaining the same tariff. This generates an inequality between the paid tariffs and the actual network usage and gives rise to questions about the tenability of the current LV connection classes [22]. For the MV-network these problems will not occur as the connection charges are already dependant on the actual maximum loading of the customer.
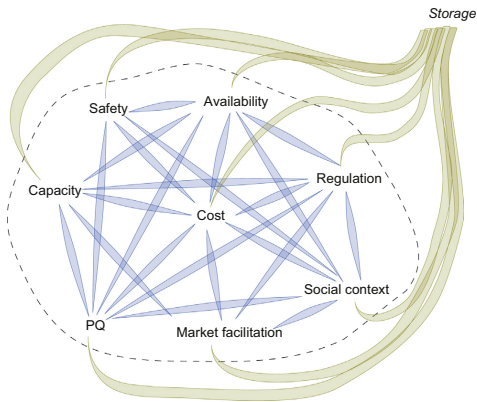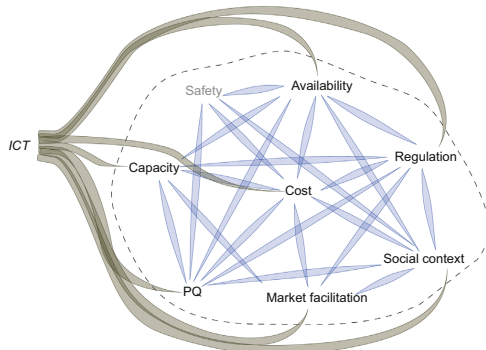
Fig. 6. Overview of the requirements of the distribution network which are affected by increased ICT usage in the network.

3.3. Energy storage

With the expected rise in electric vehicles the cost for energy storage is expected to decrease and the use of energy storage as a distribution network component is becoming viable [23]. This could have implications on all of the requirements of the distribution network as shown in Fig. 5.

3.3.1. Power quality

Energy storage can be used to increase the local power quality. The use of energy storage can providing reduce flicker and eliminate voltage dips by delivering power to the network when these phenomena occur. Energy storage can assist in voltage control, by smoothing out the load curve by storing energy during off-peak times and releasing energy during peak times. Voltage fluctuations of wind turbines [24] and of PV [25] can for instance be alleviated by the use of energy storage , which would allow for a larger amount of RES to be connected to the distribution network.

3.3.2. Availability

By storing enough energy to allow for the possibility of island operation, especially in combination with distributed generation such as photovoltaics, the provision of electricity could continue during fault conditions in the network, thus increasing the availability [26]. This solution could be applied at the connection level thus giving rise to differentiation on availability between customers.

3.3.3. Safety

Island operation can be used to increase availability, however the presence of part of the network which are still live can pose a safety risk for mechanics. Current regulations do not allow for this islanding mainly due to safety concerns. The regulations should be adjusted to achieve the islanding while maintaining the safety of the workers [27].

3.3.4. Capacity

Peak loads which determine the required network capacity only happen for a limited number of hours a year. By using storage to charge during off peak hours and discharge during peak hours we can shift the peak. This peak shifting can reduce the required network capacity and result in investment deferral [28].

3.3.5. Cost

Compared to classical network components, distributed storage has a lower economic lifetime and higher recoverable cost thus requires a different investment approach. However storage systems tend to benefit multiple actors instead of only the DNO. This requires including these actors in the investment or investments from a different actor than the DNO [29].

3.3.6. Regulation

The use of storage could generate a local increase in capacity. This reduces the non-rivalry nature of the distribution network. Depending on the strength and utilisation of the network, storage systems can alter the peak capacity for a single user, without any changes to the network. A change in peak capacity can thus be achieved without any benefits for the other connected customers. The peak capacity of a connection can thus no longer be seen as a purely public good. This could give rise to intra network competition between market parties with the DNO only supplying the basic connection service [30].

3.3.7. Social context

Energy storage systems have an impact on the social context. Storage systems often have hazardous materials, which could lead to opposition to integration of storage systems close to the costumers. Next to this the integration of energy storage in the network requires more space than the current situation. In urban and suburban areas the placement of assets is already problematic. Municipalities are less willing to offer space for the power system and have higher demands regarding the integration in the build environment.

3.3.8. Market facilitation

The usage of storage to shift the electricity usage influences the prices and transaction volumes in the electricity market. This interferers with the normal functioning of the electricity market. Therefore the DNO is currently not be allowed to trade in the electricity market. The contracting of "peak shifting" services from third-parties however is allowed.

3.4. ICT

In the recent years the possibilities offered by ICT have increased tremendously, while the costs have been dropping. Though distribution automation has been around for many years, up until recently the automation in the distribution network was limited to HV/MV substations. In the rest of the network the DNO has to rely on on-site technicians to perform measurements or switching actions. A description of the current state of distribution automation can be found in [31]. In this section the new opportunities and challenges of advanced automation in the distribution network are discussed. The requirements affected are depicted in Fig. 6.

3.4.1. Power quality

Through the use of more measurement points in the network more insight can be gained about the actual power quality in the network. As measurements are now only at the HV/MV substation level, the measured extremes in voltage deviations, harmonics and flicker levels tend to be much lower than at the end of the low voltage feeders. With the information on power quality also at the low voltage feeders a much better estimate can be made on how strong the network needs to be to provide adequate power quality [32]. The use of advanced metering schemes also gives more insight in the voltage levels in the LV-network.
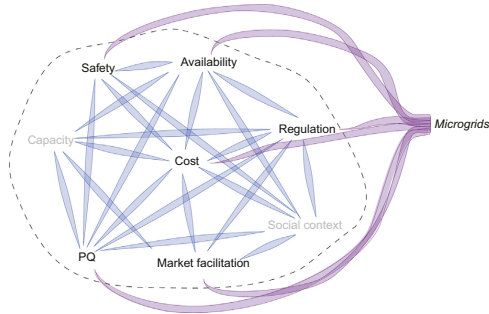
Fig. 7. Overview of the requirements of the distribution network that change due to the introduction of autonomously operating microgrids.

Fig. 8. Overview of the requirements of the distribution network which are affected by changes in the household load.

3.4.2. Availability

ICT can have much effect on the availability of the distribution network. More information and remote control outage can increase the availability while the increased dependence on ICT can increase the number of outages or outage duration.

By measuring at more places in the distribution network, faults can be faster recognised and localised [33]. Better knowledge about the location of a fault reduces the time technicians need to find the faulted section. This allows for faster isolation of the faulted section and thus increases the availability.

The availability of the MV network can be increased by automatic network reconfiguration. By the use of remote circuit breakers a faulted network section can be isolated automatically. This can reduce the average outage time for a MV interruption from 90 min to under a minute. As the cost for ICT go down, the cost of switches which can be operated remotely lowers. The dependency on the circuit breaker costs are also high [34] and can remain a barrier for the implementation of automatic network reconfiguration.

Reliability of the ICT systems remains one of the key challenges of the integration of ICT in electricity networks [35]. ICT can be used to optimise various aspects of power delivery and the interdependence between the two infrastructures is increasing [36], however a failure of the ICT system should not induce an interruption of power delivery. When designing the ICT system this should be taken into account in order to not decrease the availability.

3.4.3. Capacity

The capacity of the network which is actually utilised is unknown for most of the distribution network. The introduction of smart meters generates more information about the loading of the network for the distribution system operator. With this information a better estimate about the loading of the network can be made [37], which allows for less over-dimensioning during the network planning as the uncertainty about the current state is reduced; however the uncertainty over the future remains considerable.

The extra insight into the current loading level of the network and the ability to remotely perform control actions makes it possible to operate less conservatively, closer to the physical limits, and hence allows a better utilization of existing infrastructure and investment deferral for future infrastructure.

The meshed operation of distribution networks can become easier to implement through the use of ICT. The main barrier of meshed operation is the demands it places on the protection scheme to safely isolate a short-circuit. As wireless communication technologies continue to advance (latency of machine to machine communication in current mobile networks can be less than 50 s [39]), their use in differential protection relays becomes possible [38]. These protection relays can be retrofitted in the existing network structure, without the need for the creation of additional communication lines.
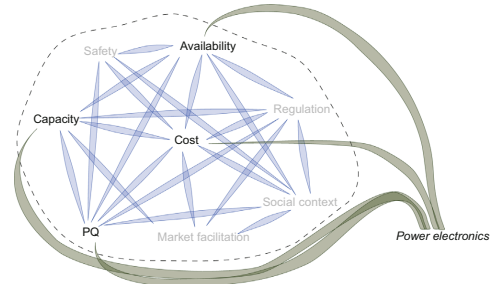
3.4.4. Cost

The use of ICT in the distribution network gives insight into the condition of the network and allows for better component monitoring [40]. This can be used to set more appropriate maintenance intervals and prolong component lifetime thus reducing the costs for the DNO. By having a better understanding of the failure and residual lifetime of the network components unplanned outages can be reduced in favour of the more cost efficient planned maintenances.

ICT components in power systems tend to have a typical economic lifetime between the 7 and 10 years [41], this is much lower than the traditional power system assets which have an economic lifetime close to the 50 years. The approach to the evaluation of the investments in ICT and the replacement schemes for ICT components are therefore harder to assess compared to classic network planning.

3.4.5. Regulation

ICT enables the use of different tariff structures. Through the use of ICT a closer link can be created between the tariffs which consumers pay and the actual price of electricity on the wholesale market. Real-time, historical and forecasted price data become easier to communicate to the electricity consumers and the possible data transfer rates allow for the tariff structure to approximate the actual market price for a given moment in time [42].

3.4.6. Social context

An increase in ICT in the network gives rise to privacy issues as some consumers do not want to give other parties insight into their electricity usage data. The concern is that from a detailed profile of the electricity usage it becomes possible to extract lifestyle patterns. This has already led to the creation of algorithms to mix and anonymise the electricity usage data as presented in [43]. This may reduce the usefulness of the smart meter data for planning purposes. When the extremes in the data are no longer present or the exact measurement location in the network is lost, the planning still has to rely on simplifying assumptions as it does nowadays, leading to over- or under-estimates of the capacity needs.

3.4.7. Market facilitation

The data which becomes available with the use of ICT technologies could give rise to a more transparent and well-functioning market. A market design which gives consumers a wide array of choices concerning energy trading would become possible. As the data on the consumption and generation of electricity becomes accessible through the use of advanced metering schemes [44], consumer insight into the origin of their electricity use increases as well as a more accurate allocation and reconciliation process for the energy suppliers.

3.5. Microgrids

As distributed generation becomes more and more common in the distribution network, the possibility of operating a small part of the network as a microgrid becomes possible. A microgrid is a physically connected part of the LV-network (or MV-network) which can at times be operated autonomously, isolated from the rest of the distribution network. The creation of microgrids could be driven by the desire to be self-sufficient, to collaborate as a neighbourhood on the energy market or to be independent from the network operator. Fig. 7 gives an overview of the requirements affected by the creation of microgrids.

3.5.1. Power quality

The decoupling of a microgrid from the local network operator can reduce the propagation of power quality phenomena. The harmonics and flicker form the higher voltage levels will no longer affect the consumers connected to the microgrid [45]. The creation of a microgrid does however mean that frequency and voltage stability should be managed locally. As there is usually little system inertia, available generation capacity reserve and the on-line generation has an intermittent nature, the stability of a microgrid is harder to maintain in autonomous operation. This can result in more volatility in the voltage and frequency [46]. The low system inertia might also require the use of flicker mitigating measures to have the flicker index within the limits [47].

3.5.2. Availability

If there is a fault in the higher voltage levels the microgrid would be able to decouple and continue to supply power to its users. In this way the effects of MV and HV failures can be eliminated resulting in down-times which can be up to ten times

lower than the current network [48]. The microgrid is however reliant on local generation which makes the availability more dependent on the available DG and energy storage systems and the location of these systems within the microgrid [49]. This limits the availability gains within the microgrid.

3.5.3. Safety

Microgrids have only a relatively small amount of generation connected, this results in low short circuit currents. However the short-circuit power flow direction could differ for the same short-location, depending on pre-fault generation at that time. This makes the protection scheme which is employed in conventional networks unfit for microgrids and could result in dangerous situations due to higher touch voltages [50]. A different protection approach which uses more advanced distribution automation is thus required to ensure safe operation of the microgrid.

3.5.4. Cost

The cost for the creation of a microgrid are higher than the realisation of a conventional network. Investments in additional power electronics devices and storage or controllable generation is needed for a microgrid. The DNO has only very limited possibilities to recover these investments. Incentives for the DNO to create a microgrid come from regulations with respect to power quality and availability. These incentives are currently not enough to recover the investments. However if the distribution system operator could charge extra for additional quality, reliability and the facilitation of renewable energy, a microgrid could have a positive business case [51]. This also makes the social and geographical context decisive whether or not a microgrid would be economically feasible.

3.5.5. Regulation

Microgrids have very different operational requirements than a conventional distribution network, and they also have a very local focus. This gives rise to the question if an incumbent DNO would be the right party to manage a microgrid. The microgrid benefits are shared between the distribution network operator, local energy prosumers and society at large (e.g. through the environment). Therefore a new regulation framework needs to be established in order to have a fair distribution of the benefits [52].

3.5.6. Market facilitation

A microgrid debilitates wholesale market facilitation, since during autonomous operation the local market is limited and only local generators can deliver energy to the microgrid customers. This would impose a single buyer/supplier market approach for the microgrid or an agent-based local marketplace, which both would require extra regulation. This extra regulation should ensure that it remains possible for customers to change energy supplier [53] which may not be physical be able to deliver the energy or to ensure that the local market power cannot be abused.

3.6. Power electronics

Technology advances have lowered the cost of many power electronic components. This could have implications for the distribution network as the business case for including power electronics in the network to for instance allow for more control over the power flows has become more positive. With the introduction of D-FACTS (distributed flexible AC transmission system), FACTS devices can also be introduced at distribution voltage levels [54]. Power electronics could be introduced to replace the traditional components or improve the capabilities of these components. A transformer for instance could be equipped
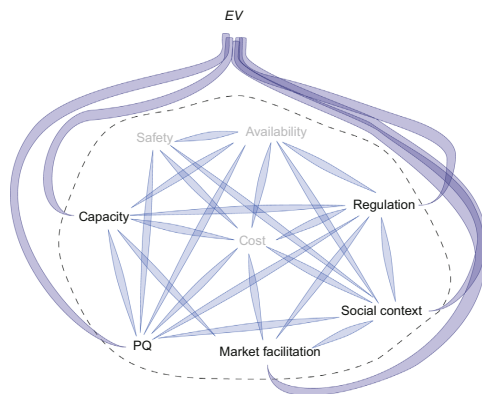

Fig. 9. The effect of an public charging infrastructure on the requirements of the distribution network.

with a power electronic tap changer to improve the voltage range [55] or could be completely replaced by a power electronics variant [56]. This has an effect on many of the requirements of the distribution network as can be seen in Fig. 8.

3.6.1. Power quality

The application of power electronics could have both a positive and a negative effect on the voltage quality. Power electronics allow for the conversion of AC to DC and DC to AC which generates control over the quality of the output voltage regardless of the input voltage quality [57]. This would make much higher levels of equivalent grid impedance feasible while maintaining an adequate level of power quality at the customer's point of connection. On the other hand, if these power converters are not designed properly they may inject high amounts of harmonics into the network [58]. This problem can be solved by using better designed power electronic converters or with active compensation with additional power electronics.

3.6.2. Availability

The distributed generation which has a significant contribution to the short-circuit power is usually required to switch off during fault conditions. Faults in feeders connected to the same busbar can trigger the protection relay of a generator connected via a power electronics interface. By introducing fault current limiters in the network or by installing a chopper in conjunction with the power electronics interface at the DG the fault ride through capabilities of DG can be increased [59]. This increases the DG availability and the system stability. For most DG connected to the MV level fault ride trough is already required from a regulation point of view. For DG connected to the LV level this is not yet the case (though it is included in the latest standards).

3.6.3. Capacity

Power electronics create the possibility of controlling the power flows in the electricity network [54]. As most of the distribution network is radially operated at the moment the need for controlling power flows is not there. With the ICT developments protection relays and communication possibilities continue to improve and meshed operation might become more advantageous. In this scenario, the need to control the power flow will arise in order to avoid overloading of cables and reduce losses.
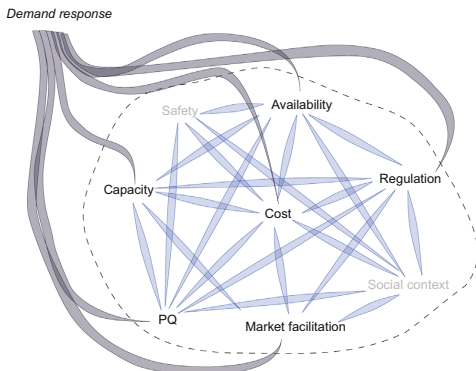
Fig. 10. Overview of the requirements of the distribution network which are affected by the increase in demand response at the household level.

3.6.4. Cost

The economic lifetime of power electronic devices is only 10–20 years [60]. This is much lower than the 40–50 years for conventional power system components. Compared to investments in cables to increase the network strength, power electronic devices can be placed above ground. This generates a positive instead of a negative asset recovery value (a power electronics converter can be placed in another part of the network for a cost lower than its value, while for underground assets the cost of excavation is much higher than the remaining value of the component). The use of power electronics in the distribution network would thus require a different approach for the investment decision making approach than most other conventional solutions.

3.7. Electric vehicles

The change from fossil fuel oriented mobility to electric mobility in the form of electric vehicles (EV) will have an influence on the distribution network. The increase in household load due to uncontrolled charging at home does not have a significantly different effect on the requirements than other energy-intensive changes in the household load. To limit the impact of EV, smart charging strategies must be applied. Smart charging can be seen as a form of demand response and is therefore included in the demand response development discussion. The changes which EV bring to the requirements on the network are evaluated purely from a charging infrastructure point of view in this section (same as the changes with respect to the uncontrolled and smart charging at home are discussed in other sections). About 90% of all urban households do not have a private parking space [61]. If these households adopt EV, they will need to rely on public charging infrastructure. A public charging infrastructure would thus require a large increase in capacity and connections to allow for the adoption of EV in urban areas. For suburban and rural areas the need for a charging infrastructure is much lower as charging from home would be possible for most EV-drivers. The requirements which are affected by this public charging infrastructure are indicated in Fig. 9.

3.7.1. Power quality

The current legislation with respect to the electricity supply gives different requirements for DNOs and privately owned charging stations when it comes to safety [62]. The requirements for a privately owned charging station are based on the general electrical requirements of low voltage installations, while a charging station owned by the DNO has to also comply with the grid code. The DNO for instance has to ensure flicker and voltage variations are within the limits, while a privately owned charging infrastructure must only ensure that the power quality is adequate for the EV-charging station to operate.

3.7.2. Capacity

The capacity required by an EV charging station differs from ordinary household load as the charging profile would almost be block-shaped (if no smart charging options are applied). The public charging station also has a larger coincidence factor than ordinary household loads where there will only be charging when people are at home. This results in a coincidence factor between the 40–60% [21], which is high compared to the current coincidence factor of household loads which is about 20%. The required capacity of a public charging station could be accurately predicted as the driving behaviour of people can be reasonably estimated from historical data [63]. This allows for smaller over-dimensioning as the capacity prediction error is smaller than with conventional loads.

	Capacity	PQ	Safety	Regulation	Market facilitation	Cost	Availability	Social Context
Demand response	+ || -	+ || -		?	-	+ || -	-	
ICT	+	+		+	+	+	+ & -	-
Microgrids		+ & -	-	?	-	?	+	
PV	+ & -	-	-	-		+ & -		+
EV	+	-		?	+ || -			?
Power electronics	+	+ & -				+	+	
Storage	+	+	-	+	-	+	+	-
Delta household load	+ & -	+ & -		+ || -		-		

+ Positive   - Negative   ? Unknown/neither
+ || - Positive or negative   + & - Positive and negative

Fig. 11. Overview of the effect of the new developments on the requirements in the distribution network. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3.7.3. Regulation

A public charging infrastructure should be regulated, as the placement of charging stations in the public space would not be possible in a completely competitive manner as there is limited parking space available [64]. There are several possible options for the role of a DNO in the charging infrastructure depending on the level of competition which would be desired in the charging infrastructure. The DNO can either be directly involved in the infrastructure which would allow for better integration with the existing network or the network operator could be one of the parties or the regulating party when the charging infrastructure is tendered [65].

3.7.4. Social context

The creation of a public charging infrastructure requires standardisation of the charger-EV interface to ensure that all EVs could be charged at the public charging station. At the moment multiple options exist for the charging interface [66] as well as for charging stations [67].

3.7.5. Market facilitation

The role a DNO plays in the public charging infrastructure could have significant effects on the requirements for market facilitation. Since the DNO is not allowed to act on the energy market (except for the purchasing of the energy to cover for transportation losses), an electricity supplier needs to be contracted by the users of the charging infrastructure. To guarantee a well-functioning market, DNOs have to ensure multiple providers can use the same charging station [68] in a non-discriminatory manner.

3.8. Demand response

Through the introduction of variable energy prices and ICT it becomes possible to influence the demand based on market conditions or grid conditions. Nowadays demand response exists in many countries in the form of a multiple tariff structure (time of use) and for large industrial customers. For the introduction of more load flexibility, communication between loads and electricity suppliers/traders or/and DNOs needs to be developed. After such a system is installed, and pricing agreements are made, demand response could be introduced at the household level. This will have an influences on the requirements of the distribution network depicted in Fig. 10.

3.8.1. Power quality

By having loads switch on/off based on a signal either given by the electricity prices or the local network capacity constraints, many loads can start-up at the exact same time. Refrigerators, air-conditioning and heat pumps can for instance be used as flexible loads. However they have start-up currents multiple times their steady-state current. The regular switching of these devices could lead to an increase in flicker [69]. The demand response signals introduce a large coincidence among the start-up times of these loads due to an automated reaction to price signals. The chance that these loads introduce flicker is thus heightened by demand response.

3.8.2. Capacity

Demand response can be used as a measure for peak shifting and shaving. This could be beneficial from both a network operator perspective as well as from an electricity supplier perspective. The timing of a demand response signal given based on the electricity or imbalance price could differ significantly from a signal given based on the network utilisation [70]. Depending on the form of implementation of demand response; electricity price based or network utilisation based, the effects on the capacity of the network could be either positive (peak shaving during high asset utilisation or asset overloading) or negative (a higher coincidence factor of loads due to their similar response to electricity prices) [71].

3.8.3. Availability

The availability can increase due to demand response, however this is mainly from a generation perspective as indicated by [72]. There it is shown that a system with nodal pricing and limited exchange and generation capacity would benefit from the possibility of reducing load during peak consumption. The introduction of demand response from an asset utilisation perspective could lower the reliability. This is because a system capable of performing demand response would be extremely complex [73] and ICT failures could lead to lower availability. Next to this the higher asset utilisation which can be obtained has a higher risk of overloading during extreme circumstances. These risks also arise if no flexible loads are present at the moment demand response is needed from an asset utilisation perspective [74].

3.8.4. Cost

The usage of demand respond for increased asset utilisation could be considered as alternative to network reinforcement [75]. The cost structure of implementing demand response is significantly different with respect to the current cost structure [76]. From high initial investments, long payback periods and almost no operational cost, the structure is changed to a high one-time demand response system investment after which there are almost only operational costs. This change in cost structure requires a different cost-benefit optimisation approach from a network operator. The optimisation primarily based on the required upfront investment should change to a continuous optimisation of the operational cost.

3.8.5. Regulation

The current regulation in the Netherlands and in most other countries with unbundled electricity markets does not allow the network operator to utilise demand side management in order to increase their asset utilisation [77] or postpone investments. With the introduction of demand response a part of the network capacity has the characteristics of a private good [30]. The question arises on what level the network operator as a monopolist can be involved in demand side management. Their involvement can reduce the social cost, but a large proportion of this cost reduction can also be reached if a demand response scheme is operated by market parties instead of the DNO.

3.8.6. Market facilitation

The usage of demand response in order to implement peak shaving or shifting from an DNO perspective influences electricity markets by increasing the local limits on the transported power at certain moments in time. With the current market structure in Europe, this is undesirable as the capacity costs of the network are only considered when it is inadequate and the regulation structure in place encourages the DNO to solve capacity constraints. However the market will not generate an incentive to solve the capacity constraints, as market parties may even benefit from the capacity constraints. As incentives from the demand response system could boost their profit, thus leading to the aggravation of the capacity constraints [78].

4. Conclusion

The effects of the main technology developments related to the energy transition have been discussed. These effects are shown to be significant enough to be taken into account when planning the distribution networks. The technologies all have effects on multiple requirements and the mitigation of unwanted effects should thus be assessed on these areas simultaneously, to ensure that the planned distribution network is capable of dealing with all technological changes. An overview of the effects on the requirements is given in Fig. 11.

If a requirement is affected by a certain development it is indicted whether this effect would be positive, negative, both, neither or unknown from the perspective of the DNO. The shading indicates the requirements which are affected the most by the developments. The blue shading indicates requirements which the DNO can directly influence, while the red shading indicates the requirements which the DNO cannot directly influence. The capacity and the power quality are the main requirements the DNO should focus on to facilitate the energy transition, while the regulatory agency should determine whether the current regulatory framework is still tenable in the light of the RES and ICT driven energy transition.

Though the effects of over-voltage and overloading due to EV or PV are mostly well-researched, there are numerous other effects on the distribution network. Many of these effects have not been investigated in detail. In addition, how these requirements influence each other also remains an open subject. The energy transition will have an impact on all the requirements of the distribution network in addition to just the technical impacts. Important examples include financial effects such as adequate remuneration of the DNO for the reinforcement and expansion of the network, or regulatory effects such as the tenability of the current tariff structure for LV-consumers. The overview of the requirements allows for a better estimation of the effects which developments and combination of developments can have.

4.1. Future work

A number of effects of new technologies on the distribution network still require more research. The implications for the energy market if a DNO starts utilising a demand response scheme to alleviate congestion in the network, or the effects on the fast voltage fluctuations if the coincidence of device switching is increased due to the reaction of these devices to control signals.

In this analysis only developments in technology have been analysed. To gain a more complete picture next to these technological changes, also social changes have to be taken into account. The increase in people working from home for instance not only changes the loading of the distribution network during the day, but also alters the social impact of outages in residential neighbourhoods during the day.

The resulting effects on the distribution network of a subset of these changes taken together can be different from the sum of their individual contributions due to emergence in a complex system. As part of future work, the modelling of individual households and their behaviour as agents combined with the introduction of various technologies will be undertaken to gain more insight into the interactions between the users, the new technologies and the distribution network. With this analysis, a better understanding of the expected effects on the distribution network can be obtained.

References

[1] Mei S, Zhang X, Cao M. Power grid complexity. Springer Science & Business Media; 2011.
[2] (APERC) APERC. A Quest for Energy Security in the 21st century. Japan; 2007.
[3] Lakervi E, Holmes EJ. Electricity distribution network design. IET 1995;21.
[4] Willis HL. Power distribution planning reference book. New York: CRC Press; 2004.
[5] Sallam AA, Malik OP. Electric distribution systems, vol. 68. Singapore: John Wiley & Sons; 2011.
[6] Thomson M, Infield DG. Impact of widespread photovoltaics generation on distribution systems. IET Renew Power Gener 2007;1(March (1)):33–40.
[7] Canova A, Giaccone L, Spertino F, Tartaglia M. Electrical impact of photovoltaic plant in distributed network. IEEE Trans Ind Appl 2009;45(January (1)):341–7.
[8] Ari GK, Baghzouz Y. Impact of high PV penetration on voltage regulation in electrical distribution systems. In: 2011 International conference on clean electrical power (ICCEP); 2011. p. 744–48.
[9] Lim YS, Tang JH. Experimental study on flicker emissions by photovoltaic systems on highly cloudy region: a case study in Malaysia. Renew Energy 2014;64(January):61–70.
[10] Kadurek P, Cobben JFG, Kling WL. Overloading protection of future low voltage distribution networks. In: 2011 IEEE Trondheim PowerTech; 2011. p. 1–6.
[11] Autoriteit_Consument_en_Markt. Netcode Elektriciteit Voorwaarden als bedoeld in artikel 31, lid 1, sub a van de Elektriciteitswet 1998; 2014.
[12] Kern EC, Gulachenski EM, Kern GA. Cloud effects on distributed photovoltaic generation: slow transients at the Gardner, Massachusetts photovoltaic experiment. IEEE Trans Energy Convers 1989;4(January (2)):184–90.
[13] Ochoa LF, Harrison GP. Minimizing energy losses: optimal accommodation and smart operation of renewable distributed generation. IEEE Trans Power Syst 2011;26(January (1)):198–205.
[14] Nieuwenhout F, Jansen J, van der Welle A, Olmos L, Cossent R, Gómez T, et al. Market and regulatory incentives for cost efficient integration of DG in the electricity system; 2010.
[15] Frías P, Gómez T, Cossent R, Rivier J. Improvements in current European network regulation to facilitate the integration of distributed generation. Int J Electr Power Energy Syst 2009;31(January (9)):445–51.
[16] Cai R, Kling WL, Cobben JFG, Blom JH, Myrzik JMA. Flicker responses of different lamp types. IET Gener, Transm Distrib 2009;3(January (9)):816–24.
[17] Azcarate I, Gutierrez JJ, Lazkano A, Saiz P, Leturiondo LA, Redondo K. Sensitivity to flicker of dimmable and non-dimmable lamps. In: 2012 IEEE international instrumentation and measurement technology conference proceedings; 2012. p. 344–47.
[18] Perera D, Meegahapola L, Perera S, Ciufo P. Flicker propagation analysis in distribution networks with embedded generation. In: 2012 IEEE international conference on power system technology (POWERCON); 2012. p. 1–6.
[19] Martinot E, Borg N. Energy-efficient lighting programs. Energy Policy 1998;26 (January (14)):1071–81.
[20] Pedersen L, Stang J, Ulseth R. Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems. Energy Build 2008;40(January (7)):1124–34.

[21] Richardson P, Flynn D, Keane A. Optimal charging of electric vehicles in low-voltage distribution systems. IEEE Trans Power Syst 2012;27(January (1)):268–79.
[22] Borenstein S. Wealth transfers among large customers from implementing real-time retail electricity pricing. Energy J 2007;28(January):2.
[23] Ribeiro PF, Johnson BK, Crow ML, Arsoy A, Liu Y. Energy storage systems for advanced power applications. Proc IEEE; 2001 January; 89 (12):1744–56.
[24] Arulampalam A, Barnes M, Jenkins N, Ekanayake JB. Power quality and stability improvement of a wind farm using STATCOM supported with hybrid battery energy storage. IEE Proc—Gener, Transm Distrib 2006;153 (January 6):701–10.
[25] Denholm P, Margolis RM. Evaluating the limits of solar photovoltaics (PV) in electric power systems utilizing energy storage and other enabling technologies. Energy Policy 2007;35(January (9)):4424–33.
[26] Hu P, Billinton R, Karki R. Reliability evaluation of generating systems containing wind power and energy storage. IET Gener Transm Distrib 2009;3(January (8)):783–91.
[27] Hanley C, Peek G, Boyes J, Klise G, Stein J, Ton D, et al. Technology development needs for integrated grid-connected PV systems and electric energy storage. In: 2009 34th IEEE photovoltaic specialists conference (PVSC) 2009; p. 001832–37.
[28] Oudalov A, Cherkaoui R, Beguin A. Sizing and optimal operation of battery energy storage system for peak shaving application. In: 2007 IEEE Lausanne power tech 2007; p. 621–5.
[29] Oudalov A, Chartouni D, Ohler C, Linhofer G. Value analysis of battery energy storage applications in power systems. In: 2006 IEEE PES power systems conference and exposition 2006; p. 2206–11.
[30] Künneke RW. Electricity networks: how natural is the monopoly?. Util Policy 1999;8(January (1)):99.
[31] Northcote-Green J, Wilson RG. Control and automation of electrical power distribution systems. Boca Raton: CRC Press; 2006.
[32] Makinen A. Power quality monitoring as integrated with distribution automation. In: 16th international conference and exhibition on electricity distribution (CIRED 2001); 2001. p. 1–5.
[33] Repo S, Della Giustina D, Ravera G, Cremaschini L, Zanini S, Selga JM, et al. Use case analysis of real-time low voltage network management. In: 2011 2nd IEEE PES international conference and exhibition on innovative smart grid technologies; 2011. p. 1–8.
[34] Popovic DS, Glamocic LR, Nimrihter MD. The optimal automation level of medium voltage distribution networks. Int J Electr Power Energy Syst 2011;33 (January (3)):430–8.
[35] Yan Y, Qian Y, Sharif H, Tipper DA. Survey on smart grid communication infrastructures: motivations, requirements and challenges. IEEE Commun Surv Tutor 2013;15(January (1)):5.
[36] Luksz Z, Deconinck G, Weijnen MPC. Securing electricity supply in the cyber age. vol. 15 of exploring the risks of information and communication technology in tomorrow's electricity infrastructure. Springer; 2010.
[37] Löf N, Pikkarainen M, Repo S, Järventausta P. Utilizing smart meters in LV network management. In: CIRED, vol. 21; 2011. p. 1050.
[38] Parikh PP, Sidhu TS, Shami AA. Comprehensive investigation of wireless LAN for IEC 61850-based smart distribution substation applications. IEEE Trans Ind Inform 2013;9(August (3)):1466–76.
[39] Nikaein N, Krea S. Latency for real-time machine-to-machine communication in LTE-based system architecture In: 2011 wireless conference 2011-sustainable wireless technologies (European Wireless); 2011. p. 1–6.
[40] Campos J. Development in the application of ICT in condition monitoring and maintenance. Comput Ind 2009;60(January (1)):1.
[41] Prüggler N. Economic potential of demand response at household level-are central-European market conditions sufficient? Energy Policy 2013;60 (January):487–98.
[42] Albadi MH, El-Saadany EF. Demand response in electricity markets: an overview. In: 2007 IEEE power engineering society general meeting; 2007. p. 1–5.
[43] Kalogridis G, Efthymiou C, Denic SZ, Lewis TA, Cepeda R. Privacy for smart meters: towards undetectable appliance load signatures. In: 2010 First IEEE international conference on smart grid communications; 2010. p. 232–7.
[44] Kumar R, Datta Ray P, Reed C. Smart grid: an electricity market perspective. In: ISGT 2011; 2011. p. 1–8.
[45] Lava J, Cobben JFG, Kling WL, Overbeeke F. Addressing LV network power quality issues through the implementation of a microgrid. In: ICREPQ; 2010. p. 1–4.
[46] Barklund E, Pogaku N, Prodanovic M, Hernandez-Aramburo C, Green TC. Energy management in autonomous microgrid using stability-constrained droop control of inverters. IEEE Trans Power Electron 2008;23(January (5)):2346–52.
[47] Prabaakaran K, Chitra N, Kumar AS. Power quality enhancement in microgrid—a survey. In: 2013 international conference on circuits, power and computing technologies (ICCPCT); 2013. p. 126–31.
[48] Mitra J, Vallem MR. Determination of storage required to meet reliability guarantees on island-capable microgrids with intermittent sources. IEEE Trans Power Syst 2012;27(January (4)):2360–7.
[49] Bae IS, Kim JO. Reliability evaluation of customers in a microgrid. IEEE Trans Power Syst 2008;23(January (3)):1416–22.
[50] Jayawarna N, Jenkins N, Barnes M, Lorentzou M, Papthanassiou S, Hatziagyriou N. Safety analysis of a microgrid. In: 2005 international conference on future power systems; 2005. p. 1–7.
[51] Morris GY, Abbey C, Wong S, Joos G. Evaluation of the costs and benefits of microgrids with consideration of services beyond energy supply. In: 2012 IEEE power and energy society general meeting; 2012. p. 1–9.
[52] Costa PM, Matos MA, Peças Lopes JA. Regulation of microgeneration and microgrids. Energy Policy 2008;36(January (10)):3893–904.
[53] Morgan MG, Zerriffi H. The regulatory environment for small independent micro-grid companies. Electr J 2002;15(January (9)):52–7.
[54] Divan D, Johal H. Distributed FACTS-A new concept for realizing grid power flow control. IEEE Trans Power Electron 2007;22(October (6)):2253–60.
[55] Hamidi V, Smith KS, Wilson RC. Smart grid technology review within the transmission and distribution sector. In: 2010 IEEE PES innovative smart grid technologies conference Europe (ISGT Europe); 2010. p. 1–8.
[56] Ronan ER, Sudhoff SD, Glover SF, Galloway DLA. Power electronic-based distribution transformer. IEEE Power Eng Rev 2002;17(April (2)):537–43.
[57] Majumder R, Ghosh A, Ledwich G, Zare F. Power management and power flow control with back-to-back converters in a utility connected microgrid. IEEE Trans Power Syst 2010;25(January (2)):821–34.
[58] Enslin JHR, Heskes PJM. Harmonic interaction between a large number of distributed power inverters and the distribution network. IEEE Trans Power Electron 2004;19(January (6)):1586–93.
[59] Rodriguez P, Timbus AV, Teodorescu R, Liserre M, Blaabjerg F. Flexible active power control of distributed power generation systems during grid faults. IEEE Trans Ind Electron 2007;54(January (5)):2583–92.
[60] Alabduljabbar AA, Milanović JV. Assessment of techno-economic contribution of FACTS devices to power system operation. Electr Power Syst Res 2010;80 (January (10)):1247–55.
[61] Nemry F, Brons M. Plug-in hybrid and battery electric vehicles: market penetration scenarios of electric drive vehicles. Luxembourg; 2010.
[62] Chang M, Wink M, Graafsma S, Visser R. Onderzoek Verlengde Private Aansluiting; 2013.
[63] Aabrandt A, Andersen PB, Pedersen AB, You S, Poulsen B, O'Connell N, et al. Prediction and optimization methods for electric vehicle charging schedules in the EDISON project. In: 2012 IEEE PES innovative smart grid technologies (ISGT); 2012. p. 1–7.
[64] San Román TG, Momber I, Abbad MR, Sánchez Miralles Á. Regulatory framework and business models for charging plug-in electric vehicles: infrastructure, agents, and commercial relationships. Energy Policy 2011;39(January (10)):6360–75.
[65] Lo Schiavo L, Delfanti M, Fumagalli E, Olivieri V. Changing the regulation for regulating the change: innovation-driven regulatory developments for smart grids, smart metering and e-mobility in italy. Energy Policy 2013;57 (January):506–17.
[66] Brown S, Pyke D, Steenhof P. Electric vehicles: the role and importance of standards in an emerging market. Energy Policy 2010;38(January (7)):3797–806.
[67] Foley AM, Winning IJ, O Gallachoir BP. State-of-the-art in electric vehicle charging infrastructure. In: 2010 IEEE vehicle power and propulsion conference; 2010. p. 1–6.
[68] Bruninga R, Sorensen JAT. Charging EVs efficiently NOW while waiting for the smart grid. In: 2013 IEEE green technologies conference (GreenTech); 2013. p. 1–7.
[69] Morcos MM, Gomez JC. Flicker sources and mitigation. IEEE Power Eng Rev 2002;22(January (11)):5.
[70] Mohsenian-Rad AH, Leon-Garcia A. Optimal residential load control with price prediction in real-time electricity pricing environments. IEEE Trans Smart Grid 2010;1(January (2)):120–33.
[71] Nijhuis M, Babar M, Gibescu M, Cobben JFG. Demand response: social welfare maximisation in an unbundled energy market case study for the low-voltage networks of a distribution network operator in the Netherlands. In: 15 IEEE international conference on environment and electrical engineering; 2015. p. 1–6.
[72] Azami R, Fard AF. Impact of demand response programs on system and nodal reliability of a deregulated power system. In: 2008 IEEE international conference on sustainable energy technologies; 2008. p. 1262–6.
[73] Fan Z, Kulkarni P, Gormus S, Efthymiou C, Kalogridis G, Sooriyabandara M, et al. smart grid communications: overview of research challenges, solutions, and standardization activities. IEEE Commun Surv Tutor 2013;15(January (1)):21–38.
[74] Palensky P, Dietrich D. Demand Side Management: demand response, intelligent energy systems, and smart loads. IEEE Trans Ind Inform 2011;7(January (3)):381–8.
[75] Poudineh R, Jamasb T. Distributed generation, storage, demand response and energy efficiency as alternatives to grid capacity enhancement. Energy Policy 2014;67(January):222–31.
[76] Strbac G. Demand side management: benefits and challenges. Energy Policy 2008;36(January (12)):4419–26.
[77] Verzijlbergh RA, Grond MOW, Luksz Z, Slootweg JG, Ilic MD. Network impacts and cost savings of controlled ev charging. IEEE Trans Smart Grid 2012;3(January (3)):1203–12.
[78] Kim JH, Shcherbakova A. Common failures of demand response. Energy 2011;36(January (2)):873–80.
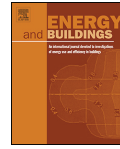
Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data

M. Nijhuis [a],[*], M. Gibescu [a], J.F.G. Cobben [a],[b]

[a] Electrical Energy Systems, Eindhoven University of Technology, Den Dolech 2, Eindhoven, The Netherlands
[b] Liander N.V., Utrechtseweg 86, Arnhem, The Netherlands

ARTICLE INFO

ABSTRACT

In the residential sector, with the introduction of electric vehicles and photovoltaics, developments are taking place which have an impact on residential load curves. In order to assess the integration of these new types of technologies on both the generation and load side, as well as to develop mitigation strategies like demand side management, detailed information is required about the load curve of a household. To gain knowledge about this load curve a residential load model is developed based on publicly available data. The model utilises a Markov Chain Monte Carlo method to model the occupancy in a household based on time use surveys, which together with weather variables, neighbourhood characteristics and behavioural data are used to model the switching pattern of appliances. The modelling approach described in this paper is applied for the situation in the Netherlands. The resulting load curve probability distributions are validated with smart meter measurements for 100 Dutch households for a week. The validation shows that the model presented in this paper can be employed for further studies on demand side management approaches and integration issues of new appliances in distribution grids.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The energy transition is changing the loading of the distribution grid especially in residential areas. With distributed generation consumers are becoming producers and the electrification of heat and transportation shifts the energy demand from fossil fuels to electricity. The current practices of grid development employed by network operators needs to be adjusted to ensure efficient integration of new loads and generation technologies and to reap the benefits of demand side management. However the current modelling of residential loads is not accurate enough to assess these problems and opportunities [1], as the energy transition and demand side management alter the diversity and stochastic characteristics of the household load curve. With the installation of smart meters more opportunities arise to validate a residential load modelling approach, while the need for modelling remains present from the perspective of forecasting, long-term planning and due to privacy concerns about smart meter data.

In the literature models have been presented for the estimation of the residential load curve. These models can be divided into roughly two basic categories: Top-down models; which focus

on the loading of MV/LV transformers and generate load curves based on this aggregation level [2] and bottom-up methods which employ statistical energy usage data or time use data to construct load profiles. There are many different approaches when it comes to building the bottom-up models. These bottom-up models can be combined with the top-down models through smart meter data [3,4]. Machine learning approaches are applied to model household load curves based on smart meter data measurements [5,6]. The residential load curve can also only be modelled at the peak times [7,8]. The modelling of household load over multiple decades requires an adjusted modelling approach, for instance based on how typiclose all cal households, behaviour and appliances change over scenarios [9]. More in-depth reviews of these different household load curve models have been performed [10,11].

The integration of distributed generation creates a voltage rise which is the main problem in distribution grids [12]. To assess the level of the voltage a more in depth assessment than just the peak load is required. The bottom-up time series approach is the most adequate for the modelling of residential load curves in order to assess the integration of distributed generation. To be able to assess the impacts of DSM, information on the appliance level needs to be known. The changes in the loading because of the shift in appliance usage depend on the actual appliances which is shifted. Demand side management approaches are generally based on information on the appliance level (e.g. [13,14]). A bottom-up household load

Nomenclature

$\mathbb{A}$	set of appliances used in model
$\mathbb{A}_H$	set of appliances present in household
A	appliance
$f_{t,A}$	fraction of time appliance A is on
f_A	fraction of households with appliance A
hr	hour
$I(t)$	irradiance
L_A	average lifetime of appliance A
l_A	remaining lifetime of appliance A
O	occupancy
$P_H(t)$	household power use
$P_A(t)$	power used by appliance A
$P_{R,A}(t_A)$	power of appliance A while running during time t_A
$P_{S,A}$	stand-by power of appliance A
r_h	household size adjustment factor
r_w	household wealth adjustment factor
t	time
t_0	initial time shift
t_A	time after which the appliance A became active
T_w	wind speed adjusted temperature
$t_{A,m}$	minimum run time of appliance A
$t_{A,r}$	average run time of appliance A
$T_{h/c}$	heating/cooling temperature
$n[\mu,\sigma]$	random variable from normal distribution $\mathcal{N}(\mu,\sigma)$
$u[a,b]$	random variable from uniform distribution over interval [a, b]
$x[\lambda]$	random variable from exponential distribution $exp(\lambda)$

Fig. 1. General schematic of a Markov Chain model with three states.

2. Modelling approach

The proposed modelling approach is based on the occupancy of the household, and to a lesser extent the behaviour of the members of the household. This approach is taken as occupancy is a driving factor for energy usage [22] and the changes in occupancy play an important role in the changes in energy consumption. The occupancy is modelled based on a Markov Chain Monte Carlo method (an explanation on Markov Chain Monte Carlo modelling is given in [23]). In Fig. 1 a schematic of a generic Markov Chain model is given with probabilities $\mathcal{P}$ of changing from one state to the other. At each time instance depending on the current state and the associated probabilities the model switches to another state or remains in the current state.

Next to the occupancy of a household the set of electrical appliances in a household and their energy use need to be modelled. This is done by creating a model for the distribution of appliances over the household based on statistical data on appliance ownership and the level of wealth in the neighbourhood and size of the dwelling. These were identified as key drivers for the degree of appliance ownership within a household [24,25].

After the occupancy of a household has been established and appliances assigned to the household the simulation of the electricity usage patterns for the appliances can be performed. This is done by simulating the switching on/off of each appliance individually using another Markov Chain based on the time of day and the weather conditions.

A flow chart of the model is presented in Fig. 2 to illustrate the computation of the load curves. The flow chart starts on the left with the inputs of the model, where the ellipses are national/state-wide inputs and the rounded rectangles are local inputs. In the following subsections the steps are explained in more detail.

2.1. Occupancy modelling

To get a better understanding of the occupancy, the occupancy profile as reported in a time use survey (2042 respondents, reporting their behaviour at a 15 min time interval, available from the Dutch national statistics agency: Statistics Netherlands) has been plotted in Fig. 3 for two consecutive days for three individual persons of different households. From this figure the large differences between the behaviour of the three persons becomes apparent. These differences will translate into differences in energy use, therefore it is important that the variation present in the occupancy is also incorporated into the model.

The first step in creating the occupancy model is determining the characteristics of a reported occupancy time series. As residents can be either active (at home and not sleeping) or inactive

modelling approach is therefore a logical choice for the assessment of demand side management (since the bottom-up approach allows for the shifting of individual appliances). Different value propositions exist for demand side management of residential loads, based on the self-consumption, electricity price or network loading. Therefore the residential load curve should be adaptable for many different scenarios, for the model to be usable when assessing future possibilities for demand side management. The bottom-up time series approach may be found in a number of references [15–20], however these approaches either focus on a limited aspect of the residential load, use private data or do not allow for the inclusion of scenarios.

The approach shown in [16] is based on a large database of measurements of appliances in a household and as this information is usually not publicly available, this approach cannot be taken. The approach described in [17] focuses on the heating load, which is appropriate for areas with electric heating, however as the Netherlands has a fairly small percentage of electric heating loads, these kind of approaches lack the required level of detail for the non-heating loads. The approach described in [19], is not designed for load modelling over multiple decades and is therefore hard to use in grid development. With the approach used in [18] assumptions have to be made on the switching behaviour of loads. This is not necessary in the approach described in this paper since statistical data from time use surveys are available in many countries [21].

The paper is organised as follows: Section 2 describes the modelling approach, followed by the modelling of the occupancy of a household and the modelling of the appliances in the household. In Section 3 the results of the model for a typical neighbourhood in the Netherlands are presented and the model is validated using smart meter measurements. Conclusions are given in Section 4.
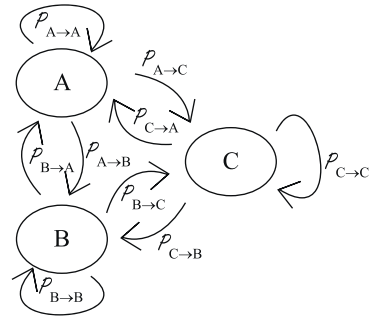
Fig. 2. Approach for the creation of the residential load curve.

Fig. 3. Occupancy from the time use survey for three persons of different households over two days with a 15 min resolution where active indicates the person is present in the household and awake.

(away from home or sleeping) and there is no clear ranking of the two values taken by the variable, the time series is of a categorical binary type. To gain more insight into the occupancy time series, its autocorrelation is computed in Fig. 4.

A high correlation can be observed at very short time lags, indicating a high persistence and at the 24 h time lag, indicating a daily pattern (between weekdays). Between 4 h ahead and 23.5 h ahead there is no sufficient correlation to assume any importance of these time lags in the creation of the model. The occupancy at highly correlated moments, i.e. less than 3 h and around 24 h lags should thus
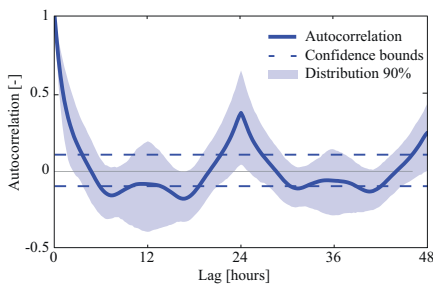
Fig. 4. Autocorrelation of the occupancy from the time use survey with the 5% confidence bounds indicated.

be utilised as input for the occupancy model. As modelling occupancy is a form of modelling behaviour, a closer look is taken at behavioural modelling as applied in the Social Sciences. The use of Markov Chain models with dynamic transition matrices show accurate results [26]. The model is also required to have an adaptive nature, this makes it harder to use a black box model like a neural network (which has also shown accurate results). The transition matrices of the Markov Chain model can easily be adjusted by for instance lowering the chance that people become active in the early afternoon to mimic an overall increase in employment, therefore the occupancy is modelled based on a Markov Chain. To ensure that effects of appliance usage times for different persons are implemented in the Markov Chain, the transition probability matrix is adjusted for the key drivers that determine appliance usage [27]: household size and age of occupants. This adjustment is done by generating the probability transition matrices from subsets of the time use survey data with respondents of certain age groups and household sizes.

The Markov Chain model is constructed to only have two states: at home and active or away/inactive. The transition probability matrix of the Markov model changes depending on the time of the day, whether it is a weekday, a Saturday or a Sunday and for weekdays the matrix is based on the occupancy of the previous day at the same time (see high auto-correlation in Fig. 4), the number and age of the members of the household. A single Markov Chain is used for each occupant of the household, only the probabilities within the Markov Chain changes over time. The occupancy Markov Chain model is defined as:

$$O(t) = \begin{cases} 1, & \text{if } u[0,1] \geq \mathcal{P}(O(t)|O(t-1)) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where the probability $\mathcal{P}(O(t)|O(t-1))$ comes from the transition probability matrix, which is constructed depending on the household characteristics as described above. This approach keeps the actual Markov Chain small and allows for quick alteration in occupancy patterns of certain groups. The probability $\mathcal{P}$ is only based on the previous value, however it is also possible to use multiple previous values $\mathcal{P}(O(t)|O(t-1)|O(t-2)|O(t-3))$. The use of multiple previous values in the model has been performed but this only show a marginal accuracy increase and is therefore not utilised. The transition matrix is constructed based on the time use surveys by determining the chance of a change in occupancy in a particular group (certain age and household size). As the correlation between two consecutive weekdays is significant a correlation is introduced
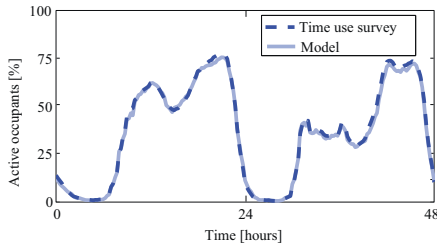
Fig. 5. Comparison of the average level of occupancy of 2042 occupants between the model and the time use survey, for a Sunday and a Monday.
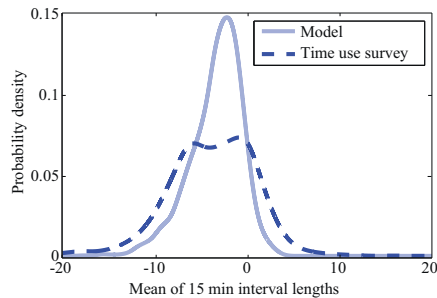
Fig. 6. Comparison of the distribution of the average number of consecutive 15 min intervals of continuous activity or inactivity level of 2042 occupants.

to adjust the occupancy based on correlation with previous day at same time by applying the following formula:

$$O(t) = \frac{O(t - 24\,hr) \times a + O^*(t)\sqrt{1 - a^2}}{a + \sqrt{1 - a^2}} \text{ if } t \text{ AND } t$$

$$- 24\,hr \text{ are both weekdays} \tag{2}$$

with $O^*(t)$ the initial occupancy as calculated with Eq. (1) and $a = u[0.4, 0.1]$ which is determined from the distribution of individual households of the autocorrelation data at the 24 h time lag. To test the quality of the Markov Chain model, the synthesised occupancy data is compared to the actual data from the time use survey.

In Fig. 5 the average level of occupancy over a set of 2042 occupants over two days for both the model and the time use survey is shown. From the figure it can be seen that there is a large difference in the occupancy during the morning and the middle of the day between the Sunday and the Monday. The modelled and the measured average occupancies are more or less similar. The aggregated modelled occupancy thus closely matches the measured one, however next to the aggregated occupancy the distribution of individual occupancy time series should also match. In order to test this, Figs. 6 and 7 have been created.

In Figs. 6 and 7 the occupancy time series is transformed from a binary series to a series indicating the consecutive number of 15 min intervals of being active as a positive number and the consecutive number of intervals of being inactive as a negative number. The distribution of the mean and standard deviation of the 2042 individual series are shown in the figures. From Fig. 6 it becomes clear that the mean of the distributions is equal for both the model and the time use survey (in accordance with Fig. 5), however the
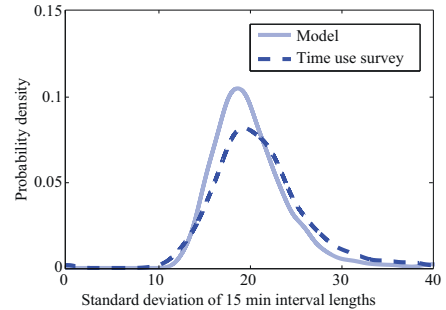
Fig. 7. Comparison of the distribution of the standard deviation of the number of consecutive 15 min intervals of activity or inactivity level of 2042 occupants.

distribution is flatter for the time use survey then for the model. This indicates that the model does not create enough extreme cases where people on average are staying active or inactive for the majority of a day. For single person households, this has a definite effect on the energy use, however for households consisting of multiple persons the extreme cases of inactivity and activity could be combined to have little effect on the energy use. The standard deviation of the consecutive 15 min intervals as shown in Fig. 7 shows a closer connection, indicating that the variability within the individual times series is similar for the modelled and survey data

2.2. Appliances present in the household

The appliances which are present in the household can be modelled based on statistical data on the level of penetration of a specific appliance in a given population group. As these data are usually only available on the national scale, adjustments have to be made to generate an estimate of the appliances present in a single household. The level of these appliances scales with income and dwelling size (number of rooms in the dwelling) as these are identified as the most important parameters determining the level of appliances [27]. Information on the penetration of different appliances in a household is estimated from the Statistics Netherlands, various commerce organisations and environmental NGO's. Statistics Netherlands has for a number of appliances the penetration grade differentiated with respect to neighbourhood wealth level and dwelling size. The income and dwelling size dependency are modelled through logistic regression analysis on these data and assumed equal for similar appliances for which no statistical data was available. The regression results for these two variables can be seen in Fig. 8 and the resulting formula for the allocation of appliances to a household can be seen in Eq. (3).

$$\text{for all } A \in \mathbb{A} \begin{cases} A \in \mathbb{A}_H, & \text{if } u[0, 1] \leq f_A \times r_h \times r_w \\ A \notin \mathbb{A}_H, & \text{otherwise} \end{cases} \tag{3}$$

2.3. Appliance model

The appliances which are present in the model can be divided into a number of categories. These categories require a different modelling approach. The main modelling of the appliances is done on a 15-min basis. There are however appliances which have an operating time of less than 15-min this requires a shorter time frame as well as to model the starting and stopping of an appliance random within the 15-min interval. The appliances are modelled with 1-min time frame t_A for their own energy usage P_A. For certain
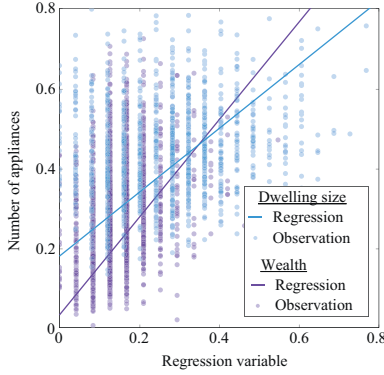
Fig. 8. Regression of the number of appliances versus the normalised dwelling size and income level.

appliance categories the usage is dependent on the time of day. For these appliances the notation includes *hr* which is use for reasons of clarity and interpreted as 60 min in all the calculations. The household load is calculated as the summation of the loads of the individual appliances:

$$P_H(t) = \sum_{A \in \mathbb{A}_H} P_A(t) \tag{4}$$

For each category the modelling approach is discussed below. For the appliances the information on the load profile is taken from [28–30].

- Baseload (e.g. fridge, modem) is assumed to be independent of occupancy in terms of energy use. The modelling of these appliances is simply a constant energy use or a switch source with constant switching times evaluated with a variable step size related to the switching of an average appliance of that kind.

$$P_A(t) = \begin{cases} P_{R,A}, & \text{if } (t - t_0) - hr \cdot \left\lfloor \dfrac{t - t_0}{hr} \right\rfloor < f_{t,A} \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{5}$$

with the initial time shift t_0 randomly chosen from $u[-30, 30]$ to create diversity between appliances. The switching behaviour is based on the percentage of time the appliance is on in one hour by taking the lower bound of the time divided by an hour.

- Night load (e.g. electric boilers, dishwashers) is switched on at an exponentially distributed time interval after 21 h (the beginning of the off-peak tariff).

$$P_A(t) = \begin{cases} P_{R,A}(t_A), & \text{if } x\left(\dfrac{1}{2\,hr}\right) < \dfrac{\left\lfloor \dfrac{t - t_0}{24\,hr} \right\rfloor - 21\,hr}{2\,hr} \\ P_{R,A}(t_A), & \text{if } P_A(t-1) \neq P_{S,A} \text{ AND } t_{A,r} \geq t_A \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{6}$$

- Heating and cooling loads (e.g. air-conditioner, electric heating) have a probability to switch on depending on the temperature and the wind speed [31] and are modelled according to a heating degree day approach as illustrated in [32]. The irradiance also has an influence on the required heating, but this effect is for a large part already included in the temperature. For the construction of the daily heating profile the data from the allocation process in the Dutch gas sector [33] is used. This approach is adjusted

for occupancy by the addition of a rule depending on the occupancy and the temperature. If there is an active occupant or the wind speed adjusted temperature $T_w(t)$ drops below zero degrees and at least two hours have passed since the last operation of the central heating system then heating will be used. For a more advanced heating model this rule can be replaced based on the specific proprieties of the dwelling (e.g. orientation, insulation, surface and window area, etc.).

$$P_A(t) = \begin{cases} P_{R,A}(t_A, T_w), & \text{if } T_w(t) \leq T_{h/c}(t) \text{ AND } \sum_{i=t-120}^{t} O(i) > 0 \\ P_{R,A}(t_A, T_w), & \text{if } t_A < t_{A,m} \text{ AND } P_A(t-1) > P_{S,A} \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{7}$$

- Lighting is modelled based on the solar irradiation and the occupancy, with lights turning on if the irradiance is below $60\,\text{W/m}^2$ [34]. The level of lighting usage is scaled with the number of occupants present in the dwelling [35].

$$P_A(t) = \begin{cases} P_{R,A}(t_A) \times \sqrt{O(t)}, & \text{if } O(t) > 0 \text{ AND } I(t) < 60 \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{8}$$

- Behavioural loads (e.g. TV, oven) are loads which can have a direct link to the activities reported in the time use survey, like watching TV or preparing food. As the energy related behaviour of the occupants depends on a large number of factors [36], the modelling is simply based on the statistical data. The power of the behavioural loads is modelled in a Markov Chain with a transition matrix which changes over time and these loads can only be activated if at least one person is active in the household. For the creation of these matrices, the same regression approach (based on household size and age) as applied in the modelling of the occupancy is employed.

$$P_A(t) = \begin{cases} P_{R,A}(t_A), & \text{if } u[0, 1] \geq \mathcal{P}(P_A(t)|P_A(t-1)) \text{ AND } O(t) > 0 \\ P_{R,A}(t_A), & \text{if } t_A < t_{A,m} \text{ AND } P_A(t-1) > P_{S,A} \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{9}$$

with $\mathcal{P}(P_A(t)|P_A(t-1))$ the chance of switching the appliance on, if the appliance was off in the previous time instance.

- General loads (e.g. water-cooker, hair dryer) can only be activated if at least one member of the household is active. This category include all loads which are not used frequently enough to link them to activities reported in the time use survey, or have an operating time which is smaller than 15 min.

$$P_A(t) = \begin{cases} P_{R,A}(t_A), & \text{if } u[0, 1] \leq f_{t,A} \text{ AND } O(t) > 0 \\ P_{R,A}(t_A), & \text{if } t_A < t_{A,m} \text{ AND } P_A(t-1) > P_{S,A} \\ P_{S,A}, & \text{otherwise} \end{cases} \tag{10}$$

2.4. Scenario implementation

In order to gain insight in the development of the residential load when the penetration of technologies like heat pumps or EV increases, scenarios can be combined with the residential load model. With the introduction of a certain scenario, the type and number of appliances in the household, the usage of the appliances and the occupancy of the household can change. For most scenarios only large changes in the presence of certain appliances in the household are assessed. Scenarios on the changes in behaviour or the occupancy are not as often discussed in literature and smaller changes are expected. For both types of scenarios the implementation in the model is elaborated upon below.
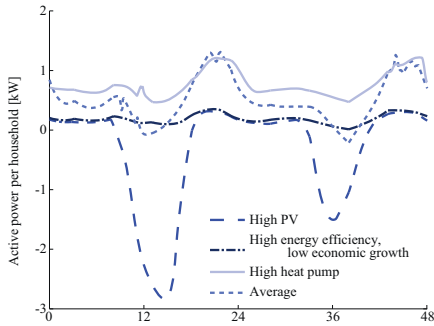
Fig. 9. Comparison of the active power consumption of a household (averaged over 100 households) for two weekdays, for four different scenario combinations.

As the probability matrices for the occupancy are already divided into groups depending on age and household size, implementing scenarios which change these parameters can be done by changing the utilised probability matrices to the matrices from another group. The changes in behaviour relate to changes in the occupancy or the activities which the persons of the household performs. Scenario implementation involves changing the probability matrices, this is done by updating the probability matrices each half year with ones adjusted to the changes introduced by the scenarios.

To implement the changes in appliances, every appliance has a lifetime l_A based on average reported lifetimes L_A [37]. At a predetermined number of intervals the appliances remaining lifetime is calculated and the available list of appliances is updated with (depending on the scenario) more efficient appliances. If an appliance remaining lifetime reaches zero, the appliance is removed from the household and based on the penetration level there is a chance that a new appliance of the same type is added to the household.

At each of these intervals there is also a chance that a household gains an additional appliance based on the household characteristics, penetration level and the lifetime of the appliance. If the penetration level of an appliance changes an additional chance of either adding or removing the appliance from the household is applied to each household. This is applied in the following manner:

$$l_A = L_A \times u[0.5, 1.5] \tag{11}$$

if l_A reaches zero, appliance A is removed from the set. After each week $\mathbb{A}_H$ is updated with formula (3) with an additional chance based on the penetration of the appliance in the model compared to the expected penetration with respect to the chosen household size and neighbourhood wealth level. The factor f_A changes based on the scenarios on the penetration of technologies, r_w changes with the economic growth scenario and the factors $P_{R,A}(t_A) \& P_{S,A}$ change with the efficiency scenarios.

The appliance efficiency scenario of the EU [38], the scenario on the Dutch economic growth [39], scenarios on the implementation of electric vehicles, PV, micro-CHP and heat pumps [40,41] are implemented in the model, to illustrate the implementation of the scenarios in the model. To generate a clearer implementation of these scenarios, the scenario studies are consolidated into three scenarios on the implementation of a certain technology; a high, medium or low penetration scenario. In Fig. 9 an example of changes in the aggregated household load for a selected quartet of scenario combinations is given.

From the figure a clear difference can be seen between the different scenarios. The high PV scenario creates a clear energy surplus in

Table 1
RMSE in Watt over one week for seven weeks of transformer data, smart meter data and modelled data.

	Transformer	Smart meter	Model
Transformer	86	78	91
Smart meter	78	72	79
Model	91	79	64
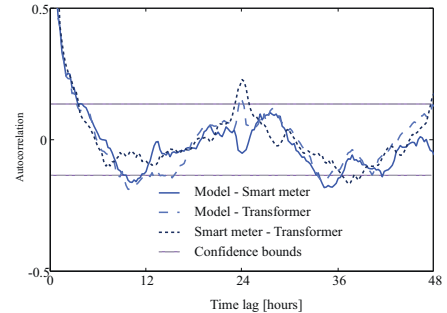
Fig. 10. Autocorrelation of the residuals of the modelled data, with the smart meter data and the transformer data and the residuals of the transformer data with the smart meter data.

the middle of the day, while the heat pump scenario generates more baseload, while the peak is not effected. The introduction of more energy efficient appliances with a low economic growth reduces the overall loading significantly.

3. Results and validation

To assess the performance of the model, a couple of simulations has been performed. First 100 households are simulated for 1 year on an Intel i7-3770 3.4 GHz processor, 8 GB ram computer. The simulation of these 100 households took 27s and 800 MB of RAM memory, expanding the simulation to 1000 households increased the computation time to 249 s and memory usage to 2 GB, while lengthening the simulation of the 100 households to 10 years instead of 1 increased the computational time to 214 s and the memory usage to 1 GB. Indicating the model can used to generate large sets of household load curves with no scalability issues. The 1 year of data of these 100 households are compared to smart meter data aggregated to 100 households and scaled (to account for the losses and the number of connected households) transformer data of a transformer of 107 households and one of 94 households. The root mean square error (RMSE) is calculated between the time series of these data sets and shown in Table 1.

The table shows that the RMSE of the model, the transformer data and the smart meter data all are in the same range. The RMSE between the sets of smart meter data and the two set transformer time series are in the same range as the RMSE of these data set with the model. This indicates that the differences between two datasets due to the stochastic nature of the residential load has the same order of magnitude as the differences between the modelled data and the measured transformer and smart meter data. To gain more insight into the errors between the modelled data and the measured data the residuals are computed and the autocorrelation of these residuals is calculated. The results are shown in Fig. 10.

The autocorrelation of the residuals gives an indication as to whether the model constantly generates a positive or negative error at certain time steps. The autocorrelation of the residuals shows a positive peak at the 24 h time lag and a negative peak at the 12 h
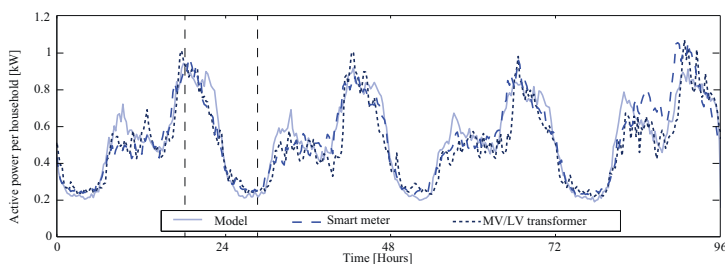
Fig. 11. Comparison of the average active power consumption, for the model and measured data for 100 households over four weekdays.
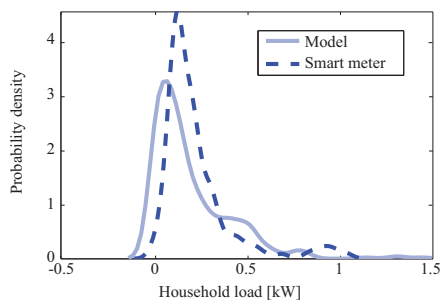
Fig. 12. Probability distribution of the electricity use of 100 households during a 15 min interval in the night (4 AM) for the model and smart meter data normalised to 1 kW base.
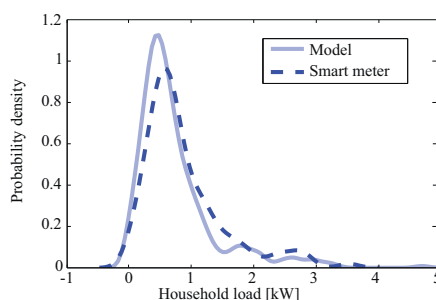
Fig. 13. Probability distribution of the electricity use of 100 households during a 15 min interval in the peak (7 PM) for the model and smart meter data normalised to 1 kW base.

Table 2
Mean and standard deviation of the night (4 AM) and peak load (7 PM) for both the model and the smart meter data.

		Smart meter	Model
Night	Mean [W]	228	237
	Std [W]	195	234
Peak	Mean [W]	820	729
	Std [W]	680	641

time lag. Both these peaks are larger in magnitude then the confidence bounds, indicating that there is a period biased present in the residuals. The residuals between the transformer and the smart meter data and the modelled and the measured data both all show the same effect. Though there is some bias within the model, this does not exceed the bias between the measured data.

Secondly the model was used to simulate the energy use in 100 households for four days (Wednesday to Saturday) in the winter. The resulting load profiles are compared with data from smart meters and from MV/LV transformer measurements of different neighbourhoods with similar characteristics as can be seen in Fig. 11. As the smart meter data was in 15 min and the transformer measurements in 10 min resolution, a time step of 15 min is employed for the comparison.

For the majority of time instances Fig. 11 shows that the three time series are close together, however the stochastic nature of the residential load, even at an aggregation level of 100 households, remains clearly visible. The modelled household load has on weekdays a morning peak which is slightly higher compared to the smart meter data and the MV/LV transformer data and a minimum load level which is lower compared to the measured data. The evening peak, and the night time energy use is comparable between the three time series, however the peak of the loading measured at the MV/LV transformer level is higher (due to the energy losses in the LV-cables) in comparison to the peak of the smart meter measurements and the model, with the model giving a relatively low peak on the fourth day.

The kernel smoothed probability density function of the set of the energy use of 100 individual households is plotted in Fig. 12 for a single 15 min interval during the night time (at 4 AM as

indicated in Fig. 11). In Fig. 13 the distribution is plotted for a 15 min interval during the peak (at 7 PM as indicated in Fig. 11). The mean and standard deviation for the model and the smart meter data are given in Table 2. The distributions of the model and the smart meter data have a similar shape, indicating that the model simulates individual household load curves with a similar stochastic properties as measured by the smart meters. The difference in the mean of the distribution of the model and the smart meter data can also be seen in Fig. 11 at the indicated time steps.

The smart meter measurements and the model both show a smaller variance in the night compared to the peak distribution, which is caused by the relatively high share of base load appliances running at night. The peak time shows more variability since at that moment many occupancy/behavioural driven appliances are operated.

4. Conclusions

A model for the creation of residential electrical load curves is presented. The model utilises a bottom-up time series approach which makes it applicable for the assessment of demand side

management through the use of home automation systems. The model is developed in a highly adaptive manner so it can be used to analyse future behavioural changes and the integration of emerging technologies, renewable generation and changes in the energy usage of appliances. The bottom-up approach is based on data publicly available and employed for the specific case of the Netherlands. The model is however general enough to be applied for any country where appliance ownership, time of use survey data is available and can be adjusted to the local conditions based on the specific information about the climate and the age, wealth and household size distribution. Scenario analysis has been introduced in the model so the effectiveness of different home energy management strategies can be evaluated over a range of scenarios. On an aggregated level the model shows behaviour similar to measurements at the MV/LV transformer and to that of smart meter data. The distribution of modelled individual load curves during the daily peak and the night time is comparable to the distribution obtained from smart meter data, validating the hypothesis that the model could be used to compute the loading of a set of aggregated households. The validation with smart meter data on the distribution of individual households shows also that the model generates a residential load profile which is close to the actual measured data.

4.1. Future work

The space heating demand in a household is at the moment simply modelled by using the approximate heat demand profile given by the gas usage, however this could be improved to a more sophisticated method which employs the thermal dynamics of the dwelling as done for instance in [17]. In addition to the heating demand, the tap water was also not modelled in great detail. Methods such as [42] could be implemented in the model to determine the effects of switching to electric instead of natural gas for the heating of tap-water. No scenarios for the human behaviour are currently implemented in the model, as these are at the moment researched to a far smaller extent than scenarios on the changes in appliances. Scenarios can be however be implemented by changing the transition probability matrices over time. As many statistical agencies have been taking time of use surveys for multiple years, trends from these surveys could be deduced and extrapolated as scenarios on the changes in behaviour of occupants (e.g. more prevalence of home office users).

References

[1] J. Dickert, P. Schegner, Residential load models for network planning purposes, MEPS (2010) 1–6.
[2] C.M. Colson, M.H. Nehrir, An alternative method to load modeling for obtaining end-use load profiles, in: NAPS, 2009, pp. 1–5, http://dx.doi.org/10.1109/NAPS.2009.5484036.
[3] K.E. Train, An assessment of the accuracy of statistically adjusted engineering (SAE) models of end-use load curves, Energy 17 (7) (1992) 713–723, http://dx.doi.org/10.1016/0360-5442(92)90079-F.
[4] F. Wallin, C. Bartusch, E. Thorin, E. Dahlquist, Important parameters for prediction of power loads – a bottom-up approach utilizing measurements from an automatic meter reading system, in: PowerAfrica, 2007, pp. 1–7, http://dx.doi.org/10.1109/PESAFR.2007.4498105.
[5] A. Azadeh, Z.S. Faiz, A meta-heuristic framework for forecasting household electricity consumption, Appl. Soft Comput. 11 (2011) 614–620, http://dx.doi.org/10.1016/j.asoc.2009.12.021.
[6] T. Zia, D. Bruckner, A. Zaidi, A hidden markov model based procedure for identifying household electric loads, in: IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society, 2011, pp. 3218–3223, http://dx.doi.org/10.1109/IECON.2011.6119826.
[7] M.A. López-Rodríguez, I. Santiago, D. Trillo-Montero, J. Torriti, A. Moreno-Munoz, Analysis and modeling of active occupancy of the residential sector in Spain: an indicator of residential electricity consumption, Energy Policy 62 (2013) 742–751, http://dx.doi.org/10.1016/j.enpol.2013.07.095.
[8] O. Ardakanian, S. Keshav, C. Rosenberg, Markovian models for home electricity consumption, in: Proceedings of the 2nd ACM SIGCOMM workshop on Green networking – GreenNets '11, 2011, p. 31, http://dx.doi.org/10.1145/2018536.2018544.
[9] A. Grandjean, G. Binet, J. Bieret, J. Adnot, B. Duplessis, A functional analysis of electrical load curve modelling for some households specific electricity end-uses, in: EEDAL, Vol. 6th, 2011, p. 1.
[10] A. Grandjean, J. Adnot, G. Binet, A review and an analysis of the residential electric load curve models, Renew. Sustain. Energy Rev. 16 (9) (2012) 6539–6565, http://dx.doi.org/10.1016/j.rser.2012.08.013.
[11] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.
[12] A. Canova, L. Giaccone, F. Spertino, M. Tartaglia, Electrical impact of photovoltaic plant in distributed network, IEEE Trans. Ind. Appl. 45 (1) (2009) 341–347, http://dx.doi.org/10.1109/TIA.2009.2009726.
[13] D.G. Infield, J. Short, C. Horne, L.L. Freris, Potential for domestic dynamic demand-side management in the UK, in: 2007 IEEE Power Engineering Society General Meeting, 2007, pp. 1–6, http://dx.doi.org/10.1109/PES.2007.385696.
[14] D. Caprino, M.L. Della Vedova, T. Facchinetti, Peak shaving through real-time scheduling of household appliances, Energy Build. 75 (2014) 133–148, http://dx.doi.org/10.1016/j.enbuild.2014.02.013.
[15] J.V. Paatero, P.D. Lund, A model for generating household electricity load profiles, Int. J. Energy Res. 30 (2006) 273–290, http://dx.doi.org/10.1002/er.1136.
[16] J. Widén, E. Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand, Appl. Energy 87 (6) (2010) 1880–1892, http://dx.doi.org/10.1016/j.apenergy.2009.11.006.
[17] K.P. Schneider, J.C. Fuller, Detailed end use load modeling for distribution system analysis, in: PES General Meeting, 2010, pp. 1–7, http://dx.doi.org/10.1109/PES.2010.5588151.
[18] J. Dickert, P. Schegner, A time series probabilistic synthetic load curve model for residential customers, in: PowerTech, 2011, pp. 1–6, http://dx.doi.org/10.1109/PTC.2011.6019365.
[19] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: a high-resolution energy demand model, Energy Build. 42 (10) (2010) 1878–1887, http://dx.doi.org/10.1016/j.enbuild.2010.05.023.
[20] R. Yao, K. Steemers, A method of formulating energy load profile for domestic buildings in the UK, Energy Build. 37 (6) (2005) 663–671, http://dx.doi.org/10.1016/j.enbuild.2004.09.007.
[21] Access time use data – centre for time use research, 2014 http://www.timeuse.org/information/access-data.
[22] S. Firth, K. Lomas, A. Wright, R. Wall, Identifying trends in the use of domestic appliances from household electricity consumption measurements, Energy Build. 40 (5) (2008) 926–936, http://dx.doi.org/10.1016/j.enbuild.2007.07.005.
[23] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Springer US, 1996, http://dx.doi.org/10.1007/978-1-4899-4485-6.
[24] Y.G. Yohanis, J.D. Mondol, A. Wright, B. Norton, Real-life energy use in the UK: how occupancy and dwelling characteristics affect domestic electricity use, Energy Build. 40 (6) (2008) 1053–1059, http://dx.doi.org/10.1016/j.enbuild.2007.09.001.
[25] A. Druckman, T. Jackson, Household energy consumption in the UK: a highly geographically and socio-economically disaggregated model, Energy Policy 36 (8) (2008) 3177–3192, http://dx.doi.org/10.1016/j.enpol.2008.03.021.
[26] A. Pentland, A. Liu, Modeling and prediction of human behavior, Neural Comput. 11 (1) (1999) 229–242, http://dx.doi.org/10.1162/089976699300016890.
[27] Nederlands instituut voor budgetvoorlichting, Energielastenbeschouwing: Verschillen in energielasten tussen huishoudens nader onderzocht, Tech. rep., NIBUD, Utrecht, 2009.
[28] R. Stamminger, C. Pakula, H. Jungbecker, M. Braun, I. Rüdenauer, C. Wendker, G. Broil, Synergy Potential of Smart Appliances, Tech. rep., Rheinische Friedrich-Wilhelms-Universität Bonn, 2008.
[29] ACS-F1: database of appliance consumption signatures, 2014, http://http://www.wattict.com/web/index.php/databases/acs-f1.
[30] Virginia Tech advanced research institute research data, 2014 http://www.ari.vt.edu/research-data.
[31] M. Hart, R. de Dear, Weather sensitivity in household appliance energy end-use, Energy Build. 36 (2) (2004) 161–174, http://dx.doi.org/10.1016/j.enbuild.2003.10.009.
[32] R.G. Quayle, H.F. Diaz, Heating degree day data applied to residential heating energy consumption, J. Clim. Appl. Meteorol. 19 (1980) 241–246, http://dx.doi.org/10.1175/1520-0450(1980)019<0241:HDDDAT>2.0.CO;2.
[33] Nederlandse mededingingsautoriteit, Allocatievoorwaarden gas, Tech. rep., NMA, 2012.
[34] I. Richardson, M. Thomson, D. Infield, A. Delahunty, Domestic lighting: a high-resolution energy demand model, Energy Build. 41 (7) (2009) 781–789, http://dx.doi.org/10.1016/j.enbuild.2009.02.010.
[35] M. Stokes, M. Rylatt, K. Lomas, A simple model of domestic lighting demand, Energy Build. 36 (2) (2004) 103–116, http://dx.doi.org/10.1016/j.enbuild.2003.10.007.
[36] J. Stephenson, B. Barton, G. Carrington, D. Gnoth, R. Lawson, P. Thorsnes, Energy cultures: a framework for understanding energy behaviours, Energy Policy 38 (10) (2010) 6120–6129, http://dx.doi.org/10.1016/j.enpol.2010.05.069.
[37] Average life expectancy of major household appliances, 2014 http://www.statista.com/statistics/220020/average-life-expectancy-of-major-household-appliances/.

[38] W. Eichhammer, T. Fleiter, B. Schlomann, S. Faberi, M. Fioretto, N. Piccioni, S. Lechtenböhmer, A. Schüring, G. Resch, Study on the energy savings potentials in EU member states, candidate countries and EEA countries, Tech. rep., Commission, European, 2009.

[39] L.H.J.M. Janssen, V.R. Okker, J. Schuur, Welvaart en leefomgeving: een scenariostudie voor Nederland in 2040, Tech. rep., Centraal Planbureau, Milieu- en Natuurplanbureau en Ruimtelijk Planbureau, 2006.

[40] J.J.C. Bruggink, The next 50 years: four European energy futures, Tech. rep., ECN, 2005.

[41] E. Veldman, Power play impacts of flexibility in future residential electricity demand on distribution network utilisation, Ph.D. thesis, Eindhoven University of Technology, 2013.

[42] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, E. Wäckelgård, Constructing load profiles for household electricity and hot water from time-use data – modelling approach and validation, Energy Build. 41 (2009) 753–768, http://dx.doi.org/10.1016/j.enbuild.2009.02.013.

CLUSTERING OF LOW VOLTAGE FEEDERS FROM A NETWORK PLANNING PERSPECTIVE

Michiel Nijhuis	Madeleine Gibescu	Sjef Cobben
Eindhoven University of Technology	Eindhoven University of Technology	Liander N.V. – The Netherlands
The Netherlands	The Netherlands	Eindhoven University of Technology
m.nijhuis@tue.nl	m.gibescu@tue.nl	sjef.cobben@alliander.com

ABSTRACT

The utilisation of low voltage (LV) networks is changing due to electrification of heating and transportation and the increase in distributed generation. To ensure that existing grids are reinforced in the most timely and cost-efficient manner, LV network planning is performed. The LV grid of the Netherlands alone consists of over 300.000 feeders, which makes accurately assessing the future loading and the required alterations in the grid for all feeders individually enormously time consuming. The clustering of feeders to a set of generic types, which can be studied in detail can provide a suitable alternative. A fuzzy k-medians clustering approach is proposed based on the data of the LV networks (88.000 feeders) of Liander (largest Dutch DSO). The main network parameters: impedances, cable length, number of branches and branch depth and the number and type of connected customers are used in combination with the graph theory concepts of degree distribution, sequence and the centrality of the power, impedance and length. By using these parameters feeders can be rebuilt from the cluster centres which would represent the structure and loading of the original feeder. Based on these generic feeders an initial assessment of the capacity and voltage deviations allows for the classification of the current LV-network of Liander as 92,55% low risk, 5,91% medium risk and 1,54% high risk feeders.

INTRODUCTION

The energy transition, from a fossil fuel based energy supply to more sustainable sources is changing the loading of the low voltage (LV) grids. The electrification of transportation and heating loads increases the electricity demand of households, while rooftop PV systems generate electricity at the households and can introduce bidirectional power flows. The LV-grids has been built over the past decades and was not designed to meet these load changes. Therefore LV network planning is becoming more and more important to deal with the reinforcing of the LV-grid in the most cost-effective way.

One of the main difficulties with the planning and assessment of the LV-grid is the large number of LV-feeders and MV/LV substations. In the Netherlands alone there are over 120.000 MV/LV substations with over 300.000 LV-feeders connected to them. The analysis of each substation and/or each feeder individually becomes a computationally intensive task. The application of mitigation measures, for a LV-feeder which is found inadequate to deal with the future loading, needs to be standardised in order to ensure a more cost-effective solution. The creation of a limited amount of generic LV-feeders can increase the effectiveness of the LV network planning. As the generic LV-feeders must have a close resemblance to the actual feeders in the field, a clustering approach on the data of the whole LV-grid is the most suitable method for creating these generic feeders.

The characterisation and clustering of electrical networks has previously been studied. For the evaluation of reliability and susceptibility to threats, clustering based on graph theory is already being used especially for the transmission network [1][2][3]. In [4] a small number of networks are defined, based on the length of the feeder, the number of connected customers and the number of branches. Though some analysis can be performed on these representative networks, they are not classified based on enough detail to be usable for network planning. A more extensive approach is required to be able to create generic grids with a strong relation to the existing low voltage grids.

In this paper a new clustering approach based on a number of LV-feeder characteristics and parameters derived from these characteristics is presented. First the feeder characterisation is described followed by a description of the clustering approach. Hereafter the clustering is applied to the LV-grid of Liander and an initial risk assessment of the LV-grid is performed to show a possible application of the clustering method.

FEEDER CHARACTERISATION

The LV-feeders within The Netherlands are constructed from the 1900's onwards mostly by municipality owned electric utilities. Depending on the local situation the most adequate LV-feeder topology was used. This created a large variety in the topology of the LV-feeders, which are not all equally capable of dealing with the future trends in electricity consumption. To give an indication of the differences in the LV-feeders the occurrence of feeder length, number of customers and cable types are shown inFigure 2 Figure 1.
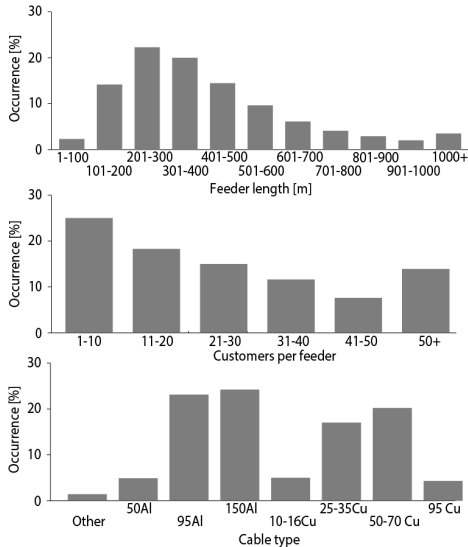
Figure 1: Distribution of the feeder length, customers per feeder and cable types in the LV-grid of a Dutch DSO.

These characteristics only give an indication of the LV-feeder type as it is not yet possible to assess the adequacy of a LV-feeder solely based on this information. Next to these parameters, the distribution of the loads over the cable, the branching within a feeder and the location and the lengths of the cables should all be known. To be able to gain enough insight to assess the LV-feeder adequacy the characterisation must be detailed enough to allow for load flow calculations. The parameterisation of a LV-feeder therefore should include more metrics on the LV-feeder then the ones depicted above, as explained below.

Clustering parameters

To characterise a LV-feeder a number of parameters are used, to illustrate these parameters they are described based on the example LV-feeder in Figure 2. To characterise the loading of a feeder, the number of household (houses in the figure) and non-household loads (arrows in the figure) connected to the feeder and the combined total and average yearly energy usage (indicated by the flows towards the transformer) of these loads are used. For the topology of the feeder the number of branches (4 from a-g, c-j, i-k and e-m) of the feeder and the maximum branch depth (3 between bus a and k from branch a-c, c-i and i-k) are used. For the characterisation of the cables the length (distance between the buses) of the average distance and the average impedance per meter (thickness of the lines between the buses) between buses and the total cable length and total impedance is utilised. To gain more insight in the distribution of the connections over the cables the average impedance at the point of connection is included as an additional parameter.

From a planning perspective the age of the cables in the LV-feeder is also an important parameter, as it gives an indication when the cable sections of a LV-feeder should be replaced, as an aging LV-cable is more prone to failure and the reliability becomes the main criteria for the cable replacement. The quality of the data on the age of the cable is however, not sufficient to be included in the clustering for the case of Liander.

Additionally, a number of parameters is derived from these characteristics, mainly to gain insight into the sequence and distribution of these characteristics, as it makes a large difference if for instance loads are mostly connected a short distance or a relatively long distance from the LV-busbar. To include this kind of parameters in the clustering, complex network analysis is applied to the LV-feeders [5]. The network is converted to a weighted graph by converting the buses to the vertices and the cables to the edges between the vertices. The edges are subsequently weighted according to one of the three following measures: The distance, the impedance and the power flow between two buses. The degree of vertex n, denoted as $d(n)$, is the sum of the weights of the edges $e_{i\text{-}j}$ which contain n as begin or end vertex. This is illustrated for vertex c in Figure 2 by the formula below:
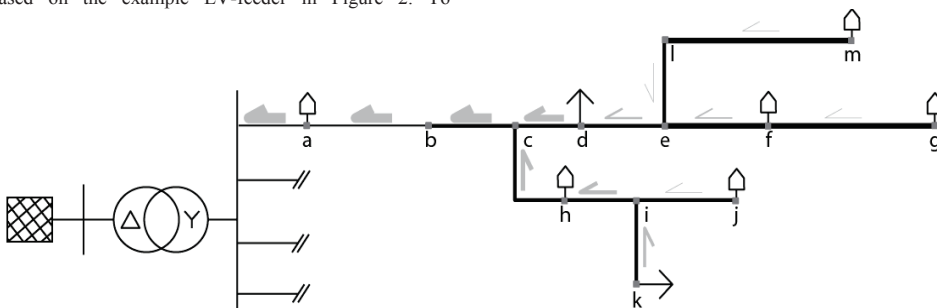
Figure 2: Overview of the parameterisation of the LV-feeders; thickness of line segments denote average impedance; gray arrows denote average yearly energy consumption.

$$d(c) = e_{b-c} + e_{c-d} + e_{c-h} \qquad (1)$$

As the number of vertices vary between LV-feeders the degree of the LV-feeder is expressed in four different ways, the mean degree (μ_d), the standard deviation of the degree (σ_d), the maximum closeness centrality (c_{c_d}):

$$\max_{i \in C}\left(c_{c_d}(i) = \frac{1}{\sum_{n \in C} d_c(i,n)} \right) \qquad (2)$$

With $d_c(i,n)$ defined as:

$$d_c(i,n) = e_{i-(i+1)} + \cdots + e_{(n-1)-n} \qquad (3)$$

and the maximum eigenvector centrality of the degree (c_{e_d}):

$$\max_{i \in C}\left(c_{e_d}(i) = \lambda \sum_{u \in C_i} c(u) \right) \qquad (4)$$

where $c(u)$ is the centrality of vertex u, C_i the set of all vertices connected by a single edge to vertex i and λ is the largest eigenvalue of the adjacency matrix (matrix of all vertices containing one if they are connected and zero if they are not connected) of i. By using these four measures for the impedance, distance and the power flow in combination with the characteristics described earlier, the feeder which is reconstructed from this data should have a close resemblance to the original feeder.

The majority of the resulting parameters have a distribution which resembles an exponential distribution with additional modes. Clustering methods generate the best results if the distribution of the data is not heavily skewed. To generate a more symmetric distribution while not affecting the modes in the data, a logarithmic transformation is applied to the non-discrete parameters (all but the number of (non)-household loads, branches and branch depth).

CLUSTERING

In order to generate the generic low voltage feeders, a clustering approach based on the data of the low voltage networks of Liander (largest Dutch DSO) is utilised. The two main approaches applicable to clustering of electrical network data are k-means and hierarchical clustering [6]. A hierarchical clustering approach would generate more accurate clustering results for the following three reasons: It can be expected that no clear independent clusters exist, outliers exist due to the power law distribution of certain network data [7] and the number of clusters is still unknown. The computational burden however makes hierarchical clustering infeasible if all the distribution

feeders are included in the clustering. Therefore an adjusted k-mean approach is chosen; fuzzy k-medians clustering. This approach utilises the median instead of the mean as cluster centre and has a soft cluster assignment [8] as is explained in more detail below.

Fuzzy k-medians clustering

In a standard k-medians clustering procedure, the clusters are generated by first selecting random data points as cluster centres. All the observations are subsequently assigned to the nearest cluster centre based on the squared Euclidean distance. For each cluster a new centre is computed by calculating the median of the observations in the cluster. This procedure is repeated until the cluster centres no longer change between iterations. If no convergence is reached within 30 iterations the randomly chosen initial cluster centres are rejected. The clustering is performed with different randomly selected observations until 5 sequential initial clusters centres no longer generate a better result. This is judged by an increase of the average distance between the cluster centres and a decrease of the average distance between points in the same cluster and its centre.

As many of the parameters which are used for the classification procedure are continuous, a fuzzy assignment of the observations to the cluster centres is introduced. This fuzzy assignment assigns each observation to each cluster centre with a certain weight based on the relative distance to the cluster centre. These weights are used for the computation of the new location of the cluster centres. The fuzzy allocation of the clusters to the cluster centres is computed with the following weights:

$$w_{i,j} = \frac{d_{i,j}}{\sum_{j=1}^{n_c} d_{i,j}} \qquad (5)$$

where $w_{i,j}$ is the weight associated with the distance between observation i and cluster centre j, $d_{i,j}$ is the distance between observation i and cluster centre j and n_c is the number of cluster centres.

Determination of the number of clusters

The number of clusters is an input variable for the clustering algorithm. To determine which number of clusters generates the best result, a number of metrics are computed to gain insight in the effects of increasing or decreasing the number of clusters. The silhouette of the clusters and cross validation are used to evaluate the adequacy of the number of clusters chosen [8]. With the silhouette, the difference between a certain observation and the other observations in the same cluster is compared to the difference between the observation and all of the observations in the nearest cluster. The silhouette is then averaged for each cluster and for all the

Table 1: Overview of the most common types of LV-feeders, after applying the clustering algorithm.

Cluster #	Length [m]	Branches [#]	Z_{POC} [mΩ]	Main cable type	Customers [#]	Occurrence [%]
1	184	1	24	150AL	17	6.4
2	270	2	35	70CU	24	4.5
3	266	1	43	95AL	39	4.5
4	218	1	26	50CU	19	4.4
5	362	3	56	150AL	32	4.1
6	290	2	74	50AL	26	3.4
7	386	2	57	95AL	49	3.4
8	633	5	107	150AL	70	3.3

clusters leading to the following formula:

$$S = \frac{1}{n_c} \sum_{c_a=1}^{n_c} \frac{1}{n_a} \sum_{i=1}^{n_a} \frac{\sum_{i=1}^{n_b} d_{i,c_b} - \sum_{i=1}^{n_a} d_{i,c_a}}{\max_{a,b \in n_c} \left(\sum_{i=1}^{n_b} d_{i,c_b} - \sum_{i=1}^{n_a} d_{i,c_a} \right)}$$

with S being the silhouette number, a being the cluster observation i belongs to with cluster centre c_a, and b being the nearest cluster to observation i. In addition to the silhouette, the mean square error of the cluster centres between two independently clustered subsets of the observations and the average distance between the observations and the clusters centres of these two subsets is computed:

$$MSE = \left(\frac{1}{n_c} \sum_{c_a=1}^{n_c} \frac{1}{n_a} \sum_{i=1}^{n_a} \hat{d}_{i,c_a} - \frac{1}{n_c} \sum_{c_a=1}^{n_c} \frac{1}{n_a} \sum_{i=1}^{n_a} d^*_{i,c_a} \right)^2$$
$$+ \left(\frac{1}{n_c^2} \sum_{c_a=1}^{n_c} \sum_{c_b=1}^{n_c} \hat{d}_{c_a,c_b} - \frac{1}{n_c^2} \sum_{c_a=1}^{n_c} \sum_{c_b=1}^{n_c} d^*_{c_a,c_b} \right)^2 \quad (6)$$

with d^* denoting the distance for the first clustered subset of the data and $\hat{d}$ the distance for the second clustered subset of the data.
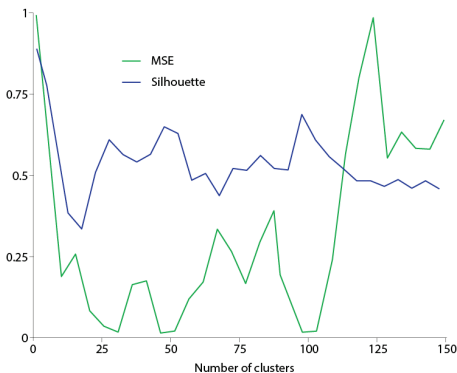
Figure 3: Cluster separation for various cluster sizes.

These measures are both computed for the number of clusters varying from 10 to 150 clusters. The resulting silhouette and normalised MSE values for the LV-feeder data set analysed are plotted in Figure 3. A silhouette close to one indicates an ideal number of clusters, while a MSE close to zero indicates the same. From the figure it can be seen that the MSE is lowest for a number 35 clusters, around 50 clusters and around 100 clusters. For the silhouette the ideal number of clusters lies around the 50 or the 100 value as can be seen from the spikes in the graph at these numbers of clusters. On closer inspection of both indicators around the 100 clusters mark, the number of clusters for the Liander network data is chosen to be 94 as it scores well on both indicators simultaneously.

RESULTS

The clustering approach described in the last section was applied to the network data of Liander. The 88.000 LV-feeders are parameterised and clustered, resulting in 94 classes. The main characteristics of the eight most common types of feeders, which account for about a third of the total LV-grid, are depicted in Table 1.

The table shows that the top seven LV-feeders have quite some similarities, while feeder 8 is significantly longer, with more connections and a higher average impedance. The most common feeder has a relatively low impedance, length and number of customers, indicating that no problems are expected for this feeder type.

Feeder assessment

To show a possible use of the clustered feeders an assessment of these feeders based on the voltage limits and overloading of the cables criteria is performed. A load flow is computed with generic 15-min load profiles [9] for the entire year 2014 for the resulting feeders. From this load flow, the minimum and maximum voltage and the maximum loading of the cables can be determined.. For all 94 generic LV-feeders these calculations have been performed. The results are converted to a high, medium or low risk based on the

risk levels given in Table 2.

Table 2: Risk levels used for feeder assessment

	High risk	Medium risk	Low risk
Voltage	<0.92 & >1.08	0.92-0.94 &1.06-1.08	0.94-1.06
Overloading	>1	0.9-1	<0.9

The risk level for the voltage deviations in the table are based on the voltage limits and do not account for a possible voltage drop at the MV side of the MV/LV transformer (the voltage at the LV side is assumed to be 1 p.u.). With these risk levels and the occurrence of the classes in the LV-grid of Liander an estimation of the currently present risks in the low voltage grid can be computed. The resulting risks for the entire LV-grid of Liander are shown in Table 3.

Table 3: Risks in the LV-network

	High risk	Medium risk	Low risk
Voltage	1.54%	5.91%	92.55%
Overloading	0.97%	2.88%	96.15%

The risks with respect to the voltage deviations and overloading is calculated for the current situation and therefore is low, as the current grid is well functioning. By using scenarios for future loads, the adequacy of the LV-grid can also be assessed with respect a certain future loading.

CONCLUSION

A clustering approach for the LV-feeder data has been proposed. This approach utilises a fuzzy k-medians clustering technique of parameterised LV-feeders. By parameterising the LV-feeders based on common characteristics such as cable length, loading, number of customers and impedance in combination with graph-theory concepts of degree and centrality, representative LV-feeders can be created from these parameters. With the use of the LV-grid of Liander, the clustering approach was tested and the LV-grid was clustered in 94 classes. An initial risk assessment on these clusters gives the indication that the current risks with respect to overloading and voltage deviations are not large

Future work

The classes of LV-feeders which have been determined from the clustering approach should be used in combination with a scenario assessment on the future loading of the grid in order to gain more insight on the risk the energy transition may be imposing on the grid. The initial assessment can be expanded with the possible safety risk within the LV-grid of Liander. A more

detailed analysis on the actual safety risks and the estimation of the 'hidden' earthing in the LV-grid should be performed to gain more insight into the acceptable risks levels for the earthing in the LV-grid.

The approach in this paper should be extended to include the distribution of feeders at a MV/LV substation as well as the upstream MV-grid to gain full insight into the characteristics and risks to the distribution grid.

REFERENCES

[1] A.J. Holmgren, 2006, "Using graph models to analyze the vulnerability of electric power networks", *Risk Analysis*, vol. 26, no. 4, pp 955-969.

[2] Y. Xu, A.J. Gurfinkel, P.A. Rikvold, "Architecture of the Florida power grid as a complex network", *Physica A: Statistical Mechanics and its Applications*, vol. 401, pp 130-140.

[3] P. Schulyz, 2014, "A random growth model for power grids and other spatially embedded infrastructure networks", *The European Physical Journal Special Topics*, pp 2593-2610.

[4] J. Dickert, , M. Domagk, , P. Schegner, , 2013, "Benchmark low voltage distribution networks based on cluster analysis of actual grid properties", *Proceedings PowerTech*.

[5] G.A. Pagani, M. Aiello, 2011, "Towards Decentralization: A Topological Investigation of the Medium and Low Voltage Grids", *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp 538-547.

[6] R.J. Sanchez-Garcia, , M. Fennelly, S. Norris, N. Wright, G. Niblo, J. Brodzki, J.W. Bialek, 2014, "Hierarchical Spectral Clustering of Power Grids", *IEEE Transactions on Power Systems*, vol. 29, pp 2229-2237.

[7] P. Crucitti,, V. Latora, M. Marchiori, 2004, "A topological analysis of the Italian electric power grid", *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1, pp 92-97.

[8] C.K. Reddy, C.C. Aggarwal, 2013, *Data Clustering: Algorithms and Applications,* CRC Press, Boca Raton, US

[9] Energy Data Services Netherlands, (online) Nov 2014, *verbruiksprofielen,* http://www.edsn.nl/wp-content/uploads/2012/08/Profielen-Elektriciteit-2013-versie-1_00.zip

Gaussian Mixture Based Probabilistic Load Flow For LV-Network Planning

Michiel Nijhuis, *Member, IEEE*, Madeleine Gibescu, *Member, IEEE*, and Sjef Cobben

Abstract—Due the many uncertainties present in the evolution of loads and distributed generation, the use of probabilistic load flow in low voltage (LV) networks is essential for the evaluation of the robustness of these networks from a planning perspective. The main challenge with the assessment of LV-networks is the sheer number of networks which need to be analyzed. Moreover, most loads in the LV-network have a volatile nature and are hard to approximate using conventional probability distributions. This can be overcome by the use of a Gaussian mixture distribution in load modeling. Taking advantage of its radial nature and high R/X ratios, the LV-network can be analyzed more efficiently from a computation viewpoint. By the application of simplifications defined in this paper, the backward–forward load flow can be solved analytically. This allows for the direct computation of the load flow equations with a Gaussian mixture distribution as load. When using this new approach, the required calculation time for small networks can be decreased to 3% of the time it takes to generate a similar accuracy with a Monte Carlo approach. The practical application of this load flow calculation method is illustrated with a case study on PV penetration.

Index Terms—Distribution network, gaussian mixture distribution, load flow analysis, power system analysis, probabilistic load flow.

I. INTRODUCTION

THE upcoming introduction of new technologies on the demand side, like electric vehicles and heat pumps, and the proliferation of distributed generation requires the distribution network operator to evaluate whether its network is still sufficiently strong. The low voltage (LV) network especially will play an important role in the transition towards a more sustainable energy system, where electricity plays a prominent role. To assess whether these LV-networks are robust enough to handle the energy transition, load flow calculations have to be performed. The loads on the LV-network are mainly residential, and detailed information about the load profile is rarely available. Assumptions have to be made regarding the magnitude of the individual residential loads. For LV-networks planning horizons are typically 30-40 years and longer, therefore introducing

a large load uncertainty [1]. To include these uncertainties about future residential loads a probabilistic load flow model should be employed. The sheer number of LV-networks and possible future scenarios in combination with an accurate assessment of the probability density function (PDF) for the residential loading would lead to infeasible computational times. Advanced probabilistic load flow methods are therefore necessary to determine the effects of the new load and generation technologies on the LV-network.

There have been many different probabilistic optimal power/load flow formulations [2]–[5] that are able to accurately and relatively quickly assess the distribution network. These methods generally reduce the required amount of samples in a Monte Carlo (MC) simulation or are based on two-point [6] or multiple point estimates [7]. Gaussian mixture models have already been used to improve the probabilistic load flow computations as well. Application to medium voltage networks with the amount of variable loads limited to mostly the distributed renewable energy resources has been performed [8], [9]. For the DC load flow, a probabilistic load flow based on cumulants and Gram-Charlier expansion has already been applied [10]. For the LV network, the DC simplifications will not hold. The LV-network has in most cases a radial topology combined with a high R/X ratio, hence other simplifications to the load flow can be made to further improve the computational speed. However, for a rough initial network assessment, the trade-off between speed and accuracy is different. With a quick initial assessment, the networks which truly require attention can be identified and more in-depth analysis can be performed on a limited subset of LV-networks.

In this paper, simplifications for the backwards-forwards load flow calculation are presented. By applying these simplifications the load flow calculation is solved analytically employing a Gaussian mixture distribution directly into the load flow formulation. In this way, the probabilistic load flow for LV-networks can be solved in more quickly. First, a Gaussian mixture distribution is applied to model the residential load. Subsequently, based on this probability distribution, the load flow calculation for a radial LV-network is altered to allow for a quick assessment of the adequacy of the network. A conventional load flow would be too computationally intensive, therefore some simplifications are made as discussed in Section III. The resulting Gaussian mixture based load flow calculation is shown in the next section. This is followed by a case study on the possible PV penetration levels in a sample network to show the advantages of the altered Gaussian mixture based probabilistic load flow.
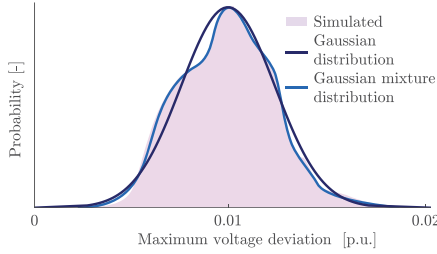
Fig. 1.   Probability of the maximum voltage deviation occurring for different combinations of smart meter data.

II. GAUSSIAN MIXTURE BASED LOAD MODELLING

The first step in assessing the adequacy of an LV-network is to model the residential load. For an efficient estimation of the admissible loading situations in a network, the load should be modelled as a PDF. The loading of an individual household is, however, a stochastic process highly dependent on the switching of high-powered appliances. To accurately model this type of load, a unimodal PDF is not optimal as it cannot always capture the exact distribution of the load which is often multi-modal because of the switching behaviour of household appliances. To illustrate this, the minimum voltage in a simple feeder with 24 households has been calculated based on all possible combinations of smart meter measurements over the households. This results in a voltage distribution as shown in Fig. 1.

In the figure also a Gaussian distribution and a five component Gaussian mixture distribution are plotted for reasons of comparison. Based on the central limit theorem, the distribution of the voltage drops should approximate a Gaussian distribution, for large enough sample size. However, from the figure, it becomes apparent that for this 24-household feeder the voltage drop in the feeder is still significantly different from the Gaussian distribution, as multiple modes are clearly present. The modelling through a combination of Gaussian distributions would generate a more realistic voltage drop profile.

A Gaussian mixture distribution is given by the following formula:

$$f(x) = \sum_{k=1}^{K} \omega_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

$$0 \leq \omega_k \leq 1 \text{ and } \sum \omega_k = 1 \tag{1}$$

where K is the number of components, ω_k is the weight, σ_k the standard deviation and μ_k the mean of the k-th component. A Gaussian mixture consisting of five components can estimate the empirical distribution of the voltage deviation with greater accuracy than the normal distribution, as shown by Fig. 1. Therefore it is useful to develop a LV load flow algorithm based on Gaussian mixture distributions, as explained in the next section.

A. Load Estimation Through Gaussian Mixture Distribution

A Gaussian mixture distribution has already been applied for the modelling of the load. In [11] and [12] methods are

shown to model the load through the use of the Expectation-Maximization (EM), either with the use of Akaikes Information Criterion (AIC) to determine the necessary number of components or with a predefined number. In this paper, the use of the approximation of the parameters of the components through the EM-algorithm is also chosen, however for the determination of the number of the components, the Bayesian information criterion (BIC) is applied [13]. This is done to give more emphasis on a lower number of components than prescribed by AIC. From a computational point of view this is favourable for the subsequent load flow calculation and to gain a better estimate of the required number of components with respect to non-machine choice.

The EM algorithm is applied, using a two-step iterative approach to find the maximum likelihood for the parameters of the Gaussian mixture model. Assuming a Gaussian mixture model with K components with the weights ω_k and component distributions $p_k(x_i|\theta_k)$, the likelihood is given by:

$$\mathcal{L}(\Theta|X) = \prod_{i=1}^{N} \sum_{k=1}^{K} \omega_k p_k(x_i|\theta_k) \tag{2}$$

with N being the number of observations of the empirical data X and θ_k consist of the parameters for the normal distribution $\theta_k = \{\mu_k, \sigma_k\}$. The maximum likelihood estimate $\arg\max_\Theta \mathcal{L}(\Theta|X)$ cannot be determined analytically. The EM algorithm interprets the data X as incomplete and adds a binary vector Y. This vector indicates which observation of X belongs to which component in the Gaussian mixture. This vector Y is determined by using a K-means clustering approach. Resulting in the following likelihood:

$$\mathcal{L}(\Theta|X,Y) = \prod_{i=1}^{N} \sum_{k=1}^{K} y_{k,i} \omega_k p_k(x_i|\theta_k) \tag{3}$$

The EM algorithm first step is to find the expected value of the complete likelihood given the parametrisation in the previous time step Θ^{p-1} through the Q-function:

$$Q(\Theta, \Theta^{p-1}) = E\left[\log p(\Theta|X,Y)|(Y, \Theta^{p-1})\right] \tag{4}$$

The next step is the maximisation of the expectation of the previous step $\Theta^p = \arg\max_\Theta Q(\Theta, \Theta^{p-1})$. This procedure is iterated until the difference between two iterations is less than 10^{-6}. The EM algorithm needs to be started with an initial value for the first iteration. This initialisation is done with the use of a k-means clustering of the data. For the value of K (the number of clusters) the BIC is calculated through:

$$BIC = K\ln(N) - 2\ln\arg\max_\Theta \mathcal{L}(\Theta|X) \tag{5}$$

The BIC is calculated for increasing values of K till the value of $BIC(K) - BIC(K+1)$ becomes smaller than a predefined stopping criterion. This procedure is applied to smart meter measurements of 197 households for the peak hours (from 18:00-20:00) to generate the probability distribution of load values during the peak hours. The loss of accuracy for using of the BIC instead of the AIC criterion can be determined by the integral squared difference [14], which is $3 \cdot 10^{-6}$. The resulting three
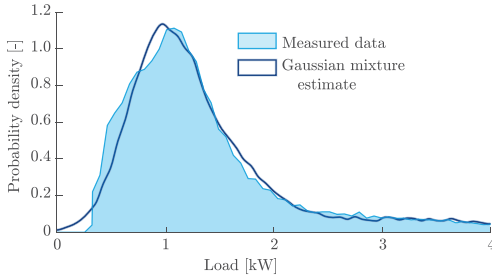
Fig. 2. Probability density function of smart meter data and the estimated Gaussian mixture distribution for the peak hours.

TABLE II
OVERVIEW OF THE THREE FEEDER CHARACTERISTICS

	Bus [#]	Load [#]	Load [kVA]	Voltage LN [V]	Phases [#]	R/X [−]
US Residential	14	10	37	120	1	3.14
US Commercial	12	8	160	120	3	2.17
EU Residential	906	55	55	240	3	9.71

Fig. 3. The European low voltage test feeder.

TABLE I
CORRELATION AND P-VALUE OF THE ELECTRICITY CONSUMPTION
FOR 120 HOUSEHOLDS WITHIN A SINGLE NEIGHBOURHOOD

Time	Correlation	p-value
00:00-02:00	0.085	0.27
12:00-14:00	0.028	0.36
18:00-20:00	0.012	0.37
00:00-24:00	0.258	0.01

component Gaussian mixture distribution and the original data are plotted in Fig. 2.

From the figure, it can be seen that for most parts the Gaussian mixture distribution and the empirical probability distribution are close together. One of the main differences is the existence of negative values in the Gaussian mixture distribution, as the Gaussian distribution ranges from $-\infty$ to ∞. Truncating the mixture distribution or individual components would result in a closer matched distribution. From the point of view of the final load flow outcome these negative values have limited effect due to their low probability and truncation is therefore not performed at this stage.

To determine how the household load can be modelled for network planning, the correlation between households within the same snapshot time is computed. This gives an indication of the level of dependence between the electricity consumption of two households. Smart meter data with a time resolution of 15-min from 120 households within the same neighbourhood are used for the analysis. As for network planning, the most important time periods are the hours of maximum and minimum load, and with the introduction of PV the household load at the middle of the day as well. In Table I the correlation coefficients and p-values are given for these 3 times of the day.

From the table, it can be seen that the correlation is low for all the evaluated time frames except when the correlation is calculated over the whole day. This indicates that the households follow a similar profile throughout the day, however, when looking at a shorter time frame this relation is no longer present. For the modelling of the household load through a Gaussian mixture distribution, this implies that the distributions of the different household loads can be considered independent for short time frames.

III. LOAD FLOW APPROXIMATIONS

For probabilistic load flow calculations in the LV-network, all the loads can be represented by PDF. As the load flow calculation needs to be solved in an iterative manner, the probabilistic approach is usually performed through MC simulations. Fortunately, LV-networks have characteristics which allow for some simplifications to generate a probabilistic load flow approach which no longer requires MC sampling. These assumptions are discussed in this section, preceded by a short overview of the characteristics of the feeders used to test the resulting load flow computation.

For testing of the Gaussian mixture based load flow approach two of the Cigré US LV test feeders [15] (with the voltage level of the US feeder reduced to the more common 120 V) and the IEEE European test feeder [16] are used. The main characteristics of these feeders are given in Table II. In the US test feeders the loads are combined at the buses to create a smaller network, therefore the US residential feeder consists of only 14 buses. For the European case the network modelling is much more detailed generating a 906 bus network. The US networks consist of a combination of overhead lines and underground cables, while the European network has only underground cable. For the EU test network the structure is also given in Fig. 3.

These LV-feeders will be used throughout to evaluate the assumptions made and to demonstrate the results of the proposed method. With the differences in the US and EU radial feeder characteristics, the suitability of the proposed method can be determined for different feeder types. The scalability can be assessed through the difference in the number of buses.

The first assumption is to assume balanced operation of the network, although the imbalance in LV-networks is generally

TABLE III
MAXIMUM ABSOLUTE VOLTAGE ERROR ASSUMED CONSTANT δ FOR
DIFFERENT LOADING SCENARIOS LEADING TO A CERTAIN MINIMUM VOLTAGE

Minimum Voltage [p.u.]	0.99	0.95	0.9	0.85	0.8
US Residential [p.u.] $\cdot 10^{-3}$	0.13	0.65	1.33	1.86	2.39
US Commercial [p.u.] $\cdot 10^{-3}$	0.04	0.25	0.53	0.84	1.21
EU Residential [p.u.] $\cdot 10^{-3}$	0.10	0.53	1.08	1.66	2.28

Fig. 4. The difference between the initial voltage drop and final voltage drop versus the voltage drop caused by the adjustment in losses and load current.

higher than the one on higher voltage levels due to the large presence of single phase loads. Through the implementation of new technologies as PV or electric vehicles, the expectancy is that the imbalance will further rise. In most cases at LV levels, there is either a single- or a three-phase network available and the mitigation of imbalance problems in a three-phase network can often be done by switching a customer from one phase to another. As this will not affect the long-term network planning results, the issue of imbalance in the LV-networks will be neglected.

The second assumption is to assume that the voltage angle is constant. As most LV-networks have high R/X ratios the change in angle throughout the network is usually limited. For the older underground cable networks in Europe, the R/X ratio is usually between 8 and 10. For newer networks the ratio is around 3, equal to the value for networks of overhead lines used in other parts of the world. To assess the difference in load flow outcome when assuming the voltage angle is constant the difference in minimum voltage has been calculated for different loading scenarios. The maximum absolute difference is voltage is calculated and displayed in Table III.

In the table, the maximum absolute error in the minimum voltage is depicted for different loading levels. The first row in the table indicates the minimum voltage which is present in the network. For all the feeders the error increases as the voltage level starts to drop, which is logical as the voltage angle differences in the network increase with an increasing current. The US commercial network has the lowest impedance and thus the smallest error. For all three networks, the voltage magnitude estimated when neglecting the phasor angle will result in an answer which is 99.7% accurate, which is more than sufficient for an initial analysis of the network from a planning perspective.

Most LV networks are radially designed or at least radially operated. This allows for the use of a backwards forwards sweep load flow approach to the probabilistic load flow calculation [17]. In a backwards-forwards sweep the currents are summed moving from the end of the network to the MV/LV substation. Hereafter, the voltages starting from the slack bus towards the end of the feeders are calculated, followed by an adjustment of the load current based on the new voltages and losses. To make a conservative estimate (estimate which leads to the highest currents and largest voltage deviations) of loading of the network a constant power load model is assumed, which leads to the following load flow expression:

$$U(i) = U_0(i) - \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=n}^{N_b} \left(\frac{S(m) + L(m)}{U(m)} \right) \right) \quad (6)$$
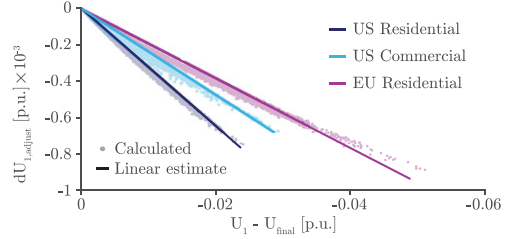
where $U(i)$ is the voltage at bus i, bus indexes i, n and m running from the busbar (index 0) to the end of the feeder (index N_b), $U_0(i)$ the base voltage, $Z(n)$ the impedance of the section between node n and $n - 1$, $S(m)$ the apparent power loading at bus m and $L(m)$ the loss $(Z \cdot I^2)$ between bus m and bus $m + 1$ as calculated by:

$$L(m) = (U(m) - U(m+1)) \sum_{k=m+1}^{N_b} \left(\frac{P(k)}{U(k)} \right) \quad (7)$$

Generally this iterative backward-forward procedure is repeated until the change in voltage between the iterations is lower than a certain threshold. To accelerate the computation of the load flow, the voltage drop difference between two iterations can be multiplied with a scalar value. If the appropriate scalar value is used the load flow can be computed with just a single iteration. To calculate this scalar value the relationship between the final voltage deviations and the voltage deviations of the losses and estimated adjusted load is defined in this paper. This relation is described as follows:

$$F_L(i) = \frac{U_1(i) - U_{\text{final}}(i)}{dU_1(i)^2 + \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=i}^{N_b} \left(\frac{L(m)}{U_0(m)} \right) \right)} \quad (8)$$

where $dU_1(i)$ is the change in voltage between iterations 0 and 1 at node i. For the three different feeders this relation is plotted in Fig. 4. As can be seen from the figure this relation is almost linear. As the relationship defined in (8) shows a linear relation it can be used to generate a linear estimate of the load flow problem for LV-networks. The load flow computation can be adjusted to include this relation, leading to the following function:

$$U_{\text{final}}(i) = U_1(i) - F_L(i) \cdot (U_1(i) - U_0(i))^2$$
$$- F_L(i) \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=n}^{N_b} \left(\frac{L(m)}{U_0(m)} \right) \right) \quad (9)$$

with U_1 calculated by using flat start voltages and no losses. This generates a non-iterative procedure for the computation of the load flow. The main difficulty with this load flow computation is the lack of knowledge about the value of $F_L(i)$ and the error introduced by this approximation. To evaluate the applicability of this load flow computation simplification, this method is applied to the three test feeders. For the load flow computation, the loading of the network is assumed to consist half of a constant

TABLE IV
AVERAGE MAXIMUM ABSOLUTE VOLTAGE ERROR FOR USING F_L

Minimum Voltage [p.u.]	0.99	0.95	0.9	0.85	0.8
US Residential [p.u.] $\cdot 10^{-3}$	0.25	1.86	5.16	6.62	5.10
US Commercial [p.u.] $\cdot 10^{-3}$	0.06	1.21	2.77	2.72	9.74
EU Residential [p.u.] $\cdot 10^{-3}$	0.06	1.06	2.78	2.96	2.21

TABLE V
AVERAGE MAXIMUM ABSOLUTE VOLTAGE ERROR
FOR COMBINED ASSUMPTIONS

Minimum Voltage [p.u.]	0.99	0.95	0.9	0.85	0.8
US Residential [p.u.] $\cdot 10^{-3}$	0.37	2.21	5.72	7.71	6.23
US Commercial [p.u.] $\cdot 10^{-3}$	0.22	1.93	4.10	3.43	7.46
EU Residential [p.u.] $\cdot 10^{-3}$	0.14	1.19	2.41	1.74	3.88

load and half of a uniformly distributed variable load. This in order to generate significant differences between the instances of the calculation of the load flow. The linear relation $F_L(i)$ is approximated based on a low loading and a high loading point estimate. Subsequently, with this value for $F_L(i)$ 10^4 load flow calculations have been performed to estimate the induced error.

In Table IV the resulting errors have been depicted. From the table, it becomes clear that the single iteration approximation induces a larger error than the previous assumption of the constant voltage angle. Note that the accuracy decreases to 98.8% for the commercial US feeder. The error also increases significantly as the minimum voltage in the system goes down. This is caused by the increasing influence of the part of the voltage reduction which is based on the factor F_L. The magnitude of the error is still within acceptable levels for an initial network adequacy assessment, especially if the voltage drop stays within the limit of ±10%.

In the final load flow calculations, both the assumption of the single iteration and the assumption of the constant voltage angle will be used. Therefore the combined error of these two assumptions is shown in Table V. When comparing the table to Tables IV and III it can be seen that the combined error of the two assumptions is less than the addition of the two individual errors. Also for some cases the error value is actually lower than the maximum error encountered when looking at the assumptions individually. The total introduced error is therefore low enough to allow the proposed load flow computation to be used for the initial adequacy assessment of a LV-feeder from a planning perspective.

A. Convolution of Gaussian Mixture Distributions

Let $f(x)$ be a Gaussian distribution with mean μ_f and variance σ_f^2 and $g(y)$ a Gaussian distribution with mean μ_g and variance σ_g^2 and the correlation between the two distributions is given by ρ. The resulting PDF of the addition of two random variables from these distributions $\mathbb{N}_x + \mathbb{N}_y$ can be computed through the use of the characteristic function of the Gaussian distribution [18].

The convolution of two independent Gaussian distributions (F_x and F_y) can be computed through their combined characteristic function $\varphi_{x+y}(t) = \varphi_x(t)\varphi_y(t)$

$$
\varphi_{x+y}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(\mu_x+\mu_y)-\frac{(\sigma_x^2+\sigma_y^2+2\sigma_{xy})t^2}{2}}\,dxdy
$$
$$
= \int_{-\infty}^{\infty} e^{it\mu_z - \frac{\sigma_z^2 t^2}{2}}\,dz \tag{10}
$$

with $\mu_z = \mu_x + \mu_y$ and $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}}$ and σ_{xy} is the covariance of x and y. For a mixture of independent ($\sigma_{xy} = 0$) Gaussian distributions this can be expanded to:

$$
f(x+y) = \sum_{k_x=1}^{K_x}\sum_{k_y=1}^{K_y} \frac{\omega_{k_x}\omega_{k_y}}{\sqrt{2\pi(\sigma_{k_x}^2+\sigma_{k_y}^2)}} e^{-\frac{\left(x+y-(\mu_{k_x}+\mu_{k_y})\right)^2}{2(\sigma_{k_x}^2+\sigma_{k_y}^2)}} \tag{11}
$$

This will increase the number of components in the mixture to $n_x \times n_y$ components. The number of components increases rapidly if all of the loads are represented as Gaussian mixture distributions. For a simple network with only 15 loads which all are represented by Gaussian mixture distributions with three components, the number of resulting components from summing these loads leads to over 14 million (3^{15}) components. To ensure this method remains computationally feasible, a component reduction needs to be applied. The most common way to reduce the number of components in the new Gaussian mixture distribution is to apply the EM algorithm with a reduced number of components. When assessing a LV-network many of the loads will have a similar probability distribution, this leads to equal or near equal terms in the Gaussian mixture. These equal terms can be compounded into one, reducing the number of components based on the following rules:

$$
\omega_{1,2} = \omega_1 + \omega_2
$$
$$
\mu_{1,2} = \frac{\mu_1\omega_1 + \mu_2\omega_2}{\omega_1 + \omega_2} \tag{12}
$$
$$
\sigma_{1,2}^2 = \frac{\sigma_1^2\omega_1 + \sigma_2^2\omega_2 + \frac{(\mu_1-\mu_2)^2\omega_1\omega_2}{\omega_1+\omega_2}}{\omega_1 + \omega_2}
$$

In the computation of the load flow, the multiplication of two dependent Gaussian mixture distributions is required. The computation of this multiplication can be done with the assumption that the standard deviation of the components is much lower than the mean. For the random variables x and y their multiplication can be defined as:

$$
f_{xy} = (\mu_x + \mathcal{N}[0,\sigma_x])(\mu_y + \mathcal{N}[0,\sigma_y])
$$
$$
= \mu_y\mu_x + \mu_y\mathcal{N}[0,\sigma_x] + \mu_x\mathcal{N}[0,\sigma_y] + \mathcal{N}[0,\sigma_x]\mathcal{N}[0,\sigma_y]) \tag{13}
$$

Since in our case, the mean is much higher than the standard deviation the term $\mathcal{N}[0,\sigma_y]\mathcal{N}[0,\sigma_y]$ can be neglected, yielding the following Gaussian distribution:

$$
\mathcal{N}\left[\mu_y\mu_x, \sqrt{(\mu_y\sigma_x)^2 + (\mu_x\sigma_y)^2 + \mu_y\mu_x\sigma_x\sigma_y}\right] \tag{14}
$$

B. Gaussian Mixture Based Load Flow Algorithm

As an illustration, consider a radial network with loads at each node except for the slack bus at node 1. The voltage can also be calculated by matrix multiplication through the use of the adjacency matrix extended with the shortest path between the node and the slack bus. For the network the extended adjacency matrix is given by:

$$\mathbf{A_j} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & & & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (15)$$

Based on the extended adjacency matrix, the extended ZBus matrix C_j can be defined:

$$\mathbf{C_j} = \left(\left(\mathbf{A_j}^T \left(\mathbf{1}^T Z \right) \circ \mathbf{I} \right) \mathbf{A_j} \right) \quad (16)$$

The initial iteration of the load flow computation becomes:

$$U_1 = \frac{S\mathbf{C_j}}{U_0^2} \quad (17)$$

where S is the vector of loads, which are represented by the σ, μ and ω of their respective Gaussian mixture distribution. The resulting voltage deviation U_1 can be squared to obtain the approximate voltage deviation through the change in load current as the in the first iteration the voltage is equal to 1[p.u.] so the additional load current can be estimated by the voltage deviation for a constant power load. The additional change in voltage caused by the losses can be calculated by using the following formula:

$$dV_L = \frac{Z\left(\frac{\mathbf{A_j}S}{U_0}\right)^2 \mathbf{C_j}}{U_0^2} \quad (18)$$

The losses and the estimate voltage deviation from the change in load current need to be multiplied by the factor F_L as defined in equation (8). F_L can be estimated by performing two ordinary load flow computations, one with a relatively low load and one with a high load. The voltage deviations calculated over each branch need to be added to obtain the total voltage deviation. The branch currents can sequentially be calculated from the voltage deviations. For most networks the use of a component reduction step, as described in equation (12), is needed after each of the steps to keep the memory usage in check.

IV. RESULTS

The accuracy of the Gaussian mixture based load flow is evaluated with respect to the conventional approach of using a Latin hypercube sampling based MC simulation [19] in combination with load flow calculations using a full backwards-forwards sweep. The load flows are calculated by using scaled versions of the distribution as shown in Fig. 2 for each load. In the Fig. 5 the resulting PDF of the minimum voltage for the US residential feeder is plotted for the Gaussian mixture approach as well as for the MC approach with increasing sample sizes.
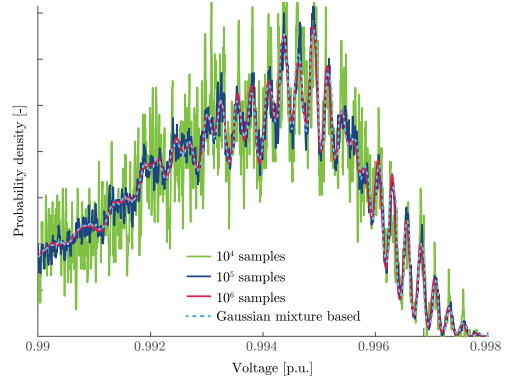
Fig. 5. Probability density function of the minimum voltage for the US residential feeder for the MC method and the proposed Gaussian mixture based load flow.

From the figure, it becomes clear that the PDF for the US residential case presents many modes. This can be explained by the low number of buses in the system, too low for the central limit theorem to apply. The MC sampling results with only 10^4 and to a lesser extent with 10^5 samples, show erratic behaviour. The MC method with 10^6 samples and the Gaussian mixture approach all are close together. A more detailed analysis is required to judge the applicability of the Gaussian mixture-based load flow in comparison with the established MC sampling.

The result of a probabilistic load flow is a histogram of the possible outcomes. To assess to what degree two histograms are derived from the same underlying distribution, a number of statistical tools can be used. The three tests used will generate an outcome of one if the histograms are exactly the same, while if there are discrepancies the result will be lower. To give a combined overview of the fit of histograms H and G the following compounded measure $D_{\text{tot}}(G, H)$ is defined for this paper, which is the product of three individual distance measures:

$$D_{\text{tot}}(G, H) = D_W(G, H) \cdot D_B(G, H) \cdot D_{\text{KL}}(G, H) \quad (19)$$

where $D_W(G, H)$ is the Wasserstein metric or earth movers distance [20], defined as follows:

$$D_W(G, H) = \frac{\min_{f_{ij}} \sum_{i,j} |f_{ij}| \cdot d_{ij}}{\sum_{i,j} f_{ij}} \quad (20)$$

where f_{ij} is a function for which the following holds: $G_i + f_{ij} = H_i$ and $G_j - f_{ij} = H_j$ the and d_{ij} the distance between i and j. $D_W(G, H)$ is the Bhattacharyya distance [21], defined as follows:

$$D_B(G, H) = -\ln \sum_i^n \sqrt{G_i \cdot H_i} \quad (21)$$

where G_i is the bin count of bin i in histogram G and H_i is the bin count of bin i in histogram H. $D_{\text{KL}}(G, H)$ is the

TABLE VI
EVALUATION OF THE MC AND GAUSSIAN MIXTURE-BASED (GM) LOAD FLOW METHODS FOR THREE FEEDERS BASED ON THE CALCULATION TIME AND THE DISCREPANCIES WITH A PROBABILITY DISTRIBUTION GENERATED WITH MC SIMULATION WITH 10^7 SAMPLES

US Residential [p.u.]	$MC10^4$	$MC10^5$	$MC5 \cdot 10^5$	GM
D_{tot} [–]	0.869	0.961	0.994	0.996
Calculation time [s]	6.17	49.3	221	4.58
US Commercial	**$MC10^4$**	**$MC10^5$**	**$MC5 \cdot 10^5$**	**GM**
D_{tot} [–]	0.879	0.965	0.995	0.992
Calculation time [s]	6.68	49.0	217	4.45
EU Residential	**$MC10^4$**	**$MC10^5$**	**$MC5 \cdot 10^5$**	**GM**
D_{tot} [–]	0.868	0.969	0.995	0.996
Calculation time [s]	7.98	64.8	307	51.5

Kullback-Leiber divergence [20] as defined by:

$$D_{KL}(G, H) = \sum_i^n G_i \ln \frac{G_i}{H_i} \qquad (22)$$

By combining these three metrics both the absolute and relative error in bin count as well as the total variation distance will be assessed. The results for the fit of the histogram with respect to a distribution generated with a MC simulation with 10^7 samples and the required computational times (using an Intel i7-3770 3.4 GHz processor, 8 GB ram computer running Matlab 2015a) are given in Table VI.

From the table, it can be seen that the Gaussian mixture based method performs comparable to an MC simulation with 10^7 samples. The computational time is however significantly lower with only 2.1% of the computational time needed for the US test feeders, while the Gaussian mixture approach required 16.8% of the time for the EU case. The difference in calculation time between US and EU test feeders indicates the scalability of the Gaussian mixture based is slightly lower, due to the need for additional component reduction steps. The computational time to obtain a similar level of accuracy is however still much lower compared to a conventional MC approach.

V. CASE STUDY

One of the applications of the probabilistic load flow is to assess the effects of PV generation on the distribution network [22]. The use of a Gaussian mixture based load flow has additional benefits compared to a normal probabilistic load flow. As the loads are modelled through the use of components within a Gaussian mixture distribution the increase or decrease in the probability of certain components can relate to an increase or decrease in penetration level of a certain technology. This is illustrated with a case study wherein the European LV test feeder is used with an increasing amount of PV generation. The PV generation is modelled based on a model for the creation of PV time series [23]. The generated PV time series for the times between 12:00 and 14:00 is added to the distribution of the household load between these two times to create the Gaussian mixture distribution of the load as illustrated in Fig. 6.

The European LV test feeder which is used in the analysis is a reasonably strong feeder, under normal load conditions the absolute voltage deviation is only 0.02. For the sake of the anal-
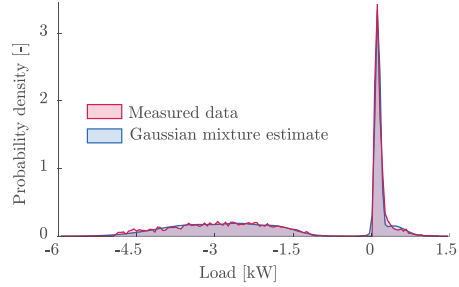
Fig. 6. Gaussian mixture distribution of residential load at midday with a PV penetration rate of 50%.
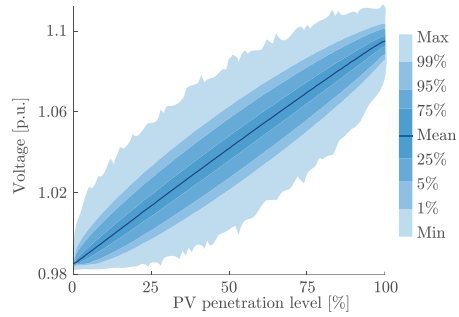
Fig. 7. Voltage distribution at the end of the feeder for different levels of PV penetration with a darker colour indicating a higher probability.

ysis, the loads on the feeder have been doubled to ensure some voltage violations will arise during the analysis. The maximum current of the cables within the network is based on their resistance with the main cable type 4c_70 having a maximum current rating of 160A per phase. The analysis is performed for the installation of rooftop PV at 0-100% of the households. By using the proposed approach the whole calculation for each percentile value took 272s. The resulting voltage at the end of the feeder is used as the main characteristic during the analysis as the expected voltage deviations are highest at that point and no data is available on the maximum loading of the different cable parts. In Fig. 7 the resulting voltage PDFs are plotted.

In the figure, a higher saturation indicates a higher probability of a certain voltage value occurring at the end of the feeder. From the figure, it can be seen that at a penetration level of 0% or 100% the PDF is narrow while around the 50% penetration rate the uncertainty about the actual voltage deviation is much larger.

For the planning of the reinforcements of the LV-network, it is important to know what the risk of overloading or the occurrence of voltage violations in the LV-network is. In order to gain insight into this statistic the chance overloading and of violating the voltage limits of the cables is plotted as a function of the PV penetration level for two different limits in Fig. 8.

In the figure the probability of exceeding the chosen maximum allowable current (0.7 or 1 [p.u.]) or the maximum allowable voltage deviation (±0.02 or ±0.05 [p.u.]) is shown. From
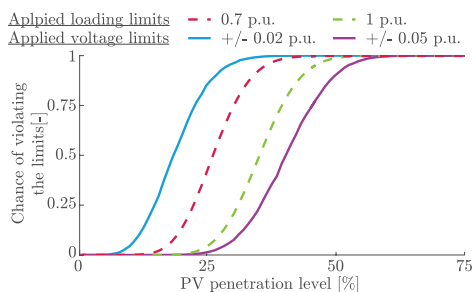
Fig. 8. The chance of violating different voltage and loading limits for different PV penetration levels.

the figure, it can be seen that all of these functions follow an S-curve. In the network the overloading of the cables will occur before the voltage limits are violated, only a very strict voltage limit of ± 0.02 would lead to an overvoltage before overloading problems start. The S-curves of these limits vary, with the ± 0.02 rising the fastest and ± 0.05 having the smallest slope for the voltage. The S-curve shows that the chance of an overloading or overvoltage can rise quickly with a relatively small increase in PV penetration level. Within these ranges of PV penetration levels, additional care must be taken when planning network re-inforcements as a small additional amount of PV can introduce a violation of the loading or voltage limits.

VI. CONCLUSION

By the application of simplifications to the backwards-forwards load flow applicable to most (radial) LV-networks it becomes possible to analytically solve the load flow equations with a Gaussian mixture distribution for the load as input. By utilising this approach the required calculation time for small networks can be decreased to 2.1% with respect to the time it takes to generate a similar accuracy with a Latin hypercube sampling MC approach.

The backwards forwards sweep for LV-networks has been simplified. The voltage angle can be neglected as the R/X ratios in LV-networks tends to be between 2 and 10. In addition, the acquired voltage difference from the initial iteration is shown to be almost linearly related to the difference between the initial and final iteration voltage deviation. The application of these two simplifications still allows for a load flow computation accuracy of 99.0%. These simplifications allow for to use the PDF of a Gaussian mixture distribution directly as an input for the load flow computation. Based on these computations the distribution of the voltage at each bus and power at each branch can be analytically determined. This allows for much faster computations of probabilistic load flow. The use of a Gaussian mixture distribution as input allows for a large variation in load distributions. To illustrate how this method could be applied, the case of assessing which level of PV-penetration a test feeder can handle has been performed. Within 5 minutes for each percentage point of PV penetration the associated risk has been

determined, showing that the method can analyse the risks due to load changes in LV-networks accurately and efficiently.

REFERENCES

[1] H. L. Willis, *Power Distribution Planning Reference Book*. Boca Raton, FL, USA: CRC Press, Mar. 2004, vol. 2.

[2] T. Williams and C. Crawford, "Probabilistic load flow modeling comparing maximum entropy and Gram-Charlier probability density function reconstructions," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 272–280, Feb. 2013.

[3] P. Siano and G. Mokryani, "Probabilistic assessment of the impact of wind energy integration into distribution networks," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4209–4217, Nov. 2013.

[4] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish–Fisher expansion," *Elect. Power Syst. Res.*, vol. 89, pp. 129–138, Apr. 2012.

[5] J. Liang, D. D. Molina, G. K. Venayagamoorthy, and R. G. Harley, "Two-level dynamic stochastic optimal power flow control for power systems with intermittent renewable generation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2670–2678, Aug. 2013.

[6] G. Verbic and C. A. Canizares, "Probabilistic optimal power flow in electricity markets based on a two-point estimate method," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1883–1893, Nov. 2006.

[7] J. M. Morales and J. Perez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1594–1601, Nov. 2007.

[8] G. Valverde, A. T. Saric, and V. Terzija, "Probabilistic load flow with non-Gaussian correlated random variables using gaussian mixture models," *IET Gener., Transm. Distrib.*, vol. 6, no. 7, pp. 701–709, Jan. 2012.

[9] F. J. Ruiz-Rodriguez, J. C. Hernandez, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *IET Renew. Power Gener.*, vol. 6, no. 2, pp. 110–121, Mar. 2012.

[10] D. Villanueva, A. E. Feijóo, and J. L. Pazos, "An analytical method to solve the probabilistic load flow considering load demand correlation using the dc load flow," *Elect. Power Syst. Res.*, vol. 110, pp. 1–8, Jan. 2014.

[11] C. Carmona-Delgado, E. Romero-Ramos, and J. Riquelme-Santos, "Probabilistic load flow with versatile non-gaussian power injections," *Elect. Power Syst. Res.*, vol. 119, pp. 266–277, Oct. 2014.

[12] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 29–37, Feb. 2010.

[13] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, Jan. 2000.

[14] J. L. Williams and P. S. Maybeck, "Cost-function-based Gaussian mixture reduction for target tracking," in *Proc. 6th Int. Conf. Inf. Fusion* 2003, pp. 1047–1054, vol. 2.

[15] K. Strunz, R. H. Fletcher, R. Campbell, and F. Gao, "Developing benchmark models for low-voltage distribution feeders," in *Proc. 2009 IEEE Power & Energy Soc. General Meeting*, Jul. 2009, pp. 1–3.

[16] IEEE PES Distribution Systems Analysis Subcommittee Radial Test Feeders European Low Voltage Test Feeder. (Feb. 26, 2016). [Online]. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders.html

[17] E. Janecek and D. Georgiev, "Probabilistic extension of the backward/forward load flow analysis method," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 695–704, May 2012.

[18] O. Knill, "Probability and stochastic processes with applications," Jun. 2011, pp. 7–14, 116–122. [Online]. Available: www.math.harvard.edu/~knill/teaching/math144 1994/probability .pdf

[19] H. u , C. . Chung, K. P. Wong, H. W. Lee, and J. H. hang, "Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 661–667, May 2009.

[20] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *Int. Statist. Rev.*, vol. 70, pp. 419–435, Dec. 2002. [Online]. Available: http://adsabs.harvard.edu/abs/2002ISRv...70..419G

[21] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhya: Indian J. Statist.*, vol. 7, no. 4, pp. 401–406, Jul. 1946.

[22] S. Conti and S. Raiti, "Probabilistic load flow using monte carlo techniques for distribution networks with photovoltaic generators," *Solar Energy*, vol. 81, pp. 1473–1481, Dec. 2007. [Online]. Available: http://adsabs.harvard.edu/abs/2007SoEn...81.1473C

[23] . Ren, W. a n, X. hao, W. Li, and J. u , "Chronological probability model of photovoltaic generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1077–1088, May 2014.

Variance-Based Global Sensitivity Analysis for Power Systems

F. Ni, M. Nijhuis, *Member*, *IEEE*, P. H. Nguyen, *Member, IEEE,* J. F. G. Cobben, *Member, IEEE*

Abstract—**Knowledge of the impact of uncertain inputs is valuable, especially in power systems with large amounts of stochastic renewable generations. A global sensitivity analysis (GSA) can determine the impact of input uncertainties on the output quantity of interest in a certain physical or mathematical model. The GSA has not been widely employed in power systems due to the prohibitively computational burden. In this paper it is demonstrated that, via the implementation of a basis-adaptive sparse polynomial chaos expansion, a GSA can be applied to the power system with numerous uncertain inputs. The performance of the proposed method is tested on both the IEEE 13-bus test feeder and the IEEE 123-node test system, in presence of a large amount of independent or correlated uncertain inputs. The possible application of a GSA on the basis of the basis-adaptive sparse polynomial chaos expansion in power systems are discussed in terms of various sensitivities. The findings cannot only be used to rank the most influential input uncertainties with respect to a specific output, such as variances of the nodal power, but also to identify the most sensitive or robust electrical variables such as the bus voltage with respect to input uncertainties.**

Index Terms— **Global sensitivity analysis, polynomial chaos expansion, variance analysis, correlation, distribution system.**

NOMENCLATURE

The primary notation used in this paper is listed below, while other symbols and abbreviations are given as needed.

A. Acronyms

ANCOVA	ANalysis of COVAriance
ANOVA	ANalysis Of Variance
BAP	Load bus active power
BASPC	Basis-adaptive sparse polynomial chaos
GSA	Global sensitivity analysis
LSA	Local sensitivity analysis
MC	Monte Carlo
PC	Polynomial chaos
PDF	Probability density function
QoI	Quantity of interest
SA	Sensitivity analysis
SC	Sensitivity coefficient

B. Constants and Indices

d	Truncation degree of universal polynomial
D_j	Partial variance of the j-th random variable
F	A σ-algebra on Ω
$\mathbf{H}$	Experimental matrix
i_j	The j-th degree of the i-th univariate polynomial
$K_q^{M,d} +1$	The total number of BASPC expansion terms
M	Number of random variables
N_{ED}	Number of the experimental design
N_{sam}	Number of samplers for MC-based simulation
Pr	A probability measure on (Ω, F)
P_i	Active power injection of the i-th load
q	Truncation norm
Q_i	Reactive power injection of the i-th load
S_j	First-order SC of the j-th random variable
ST_j	Total SC of the j-th random variable
$S_{j\text{U}}$	The SC due to physical contributions
$S_{j\text{|C}}$	The SC due to stochastically correlation
δ_{ij}	Kronecker delta function
λ	The positive penalty factor
Ω	A sample space

C. Variables and Sets

$\{a_i\}$	Coefficient of the i-th polynomial basis
$\mathbf{a}$	The vector of coefficients $\{a_j\}$
$\hat{\mathbf{a}}$	The estimated coefficient vector via regression
Cov(.)	Covariance of a random variable/ random vector
$g(.)$	Generic model with random inputs
$g_{PC}(.)$	PC expansion of the generic model $g(.)$
$I_q^{M,d}$	Hyperbolic index set of $\mathbf{i}$
$L^2(\Omega, F, P)$	Finite dimensional random space
$\mathbb{N}^M$	The M-dimensional sphere of natural numbers
$\mathbb{R}^M$	The M-dimensional sphere of real numbers
Var(.)	Variance of a random variable/random vector
$\text{Var}_{\mathbf{x}_{-j}}(.)$	Partial variance of $\mathbf{X}$ by fixing all variables except for the j-th random variable
X	Random variable
$\mathbf{X}$	Random vector
$\mathbf{X}_{-j}$	Sub-vector of $\mathbf{X}$, $(X_1,\dots,X_{j-1}, X_{j+1},\dots, X_M)$
$\mathbf{X}_\alpha$	Sub-vector of $\mathbf{X}$ whose components are with index in $\boldsymbol{\alpha}$
x	Sample of random variable X
$\mathbf{x}$	Sample of random vector $\mathbf{X}$
$\boldsymbol{\alpha}$	An integral index set
$\boldsymbol{\xi}$	A set of experimental designs of random vector
$\boldsymbol{\eta}$	The stochastic response vector corresponding to random input vector $\boldsymbol{\xi}$
$\{\Phi_i\}$	The i-th univariate polynomial basis
$\{\Psi_i\}$	The i-th multivariate polynomial basis
ω_X	Marginal PDF of random variable X
$\omega_{\mathbf{X}}$	Joint PDF of random vector $\mathbf{X}$
$\Pi_{\mathbf{X}}$	Support of random vector $\mathbf{X}$

F. Ni, M. Nijhuis, P. H. Nguyen, J. F. G. Cobben are with the Electrical Energy Systems Group, Eindhoven University of Technology, Eindhoven, 5612 AZ, The Netherlands (e-mail: [f.ni, m.nijhuis, p.nguyen.hong, j.f.g.cobben]@tue.nl). J. F. G. Cobben is also with the Alliander, Arnhem 6812 AH, The Netherlands.

∅   The empty set

I. INTRODUCTION

POWER system modelling, operating and planning are inevitably subjected to various sources of uncertainty. A high penetration of renewable energy generations such as photovoltaic and wind power brings more uncertainties to the power system due to their output variability, in addition to the already present uncertainty due to the stochastic behavior of end users. As a result, the uncertainty analysis has become a vital tool for the decision-making within the power system for all involved actors [1].

An uncertainty analysis contains three steps [2]: uncertainty modelling, uncertainty propagation and sensitivity analysis (SA). Due to the inherent randomness, measurement errors or natural variation, many power system parameters are subjected to uncertainty. Therefore, uncertainty modelling aims to depict a realistic picture of each critical source of uncertainty, based on available information. First of all, the uncertainty modelling intends to express target uncertainties by their corresponding possibilistic or probabilistic features in the form of a fuzzy set or a distribution function and their correlations. The uncertainty propagation addresses the issue of propagating uncertainties of the inputs to the outputs through a specific model. The SA is subsequently performed to quantify the individual or collective importance of each input variable on the model outputs.

SA is mainly classified into two categories: local SA (LSA) and global SA (GSA). An LSA analyses the effect of a single input on the output quantity of interest (QoI) at a time, while treating the other inputs as deterministic values. An LSA locally estimates each sensitivity measure in the vicinity of a nominal value of the selected input. Instead of exploring only a fraction of the design space and neglecting the joint effects among inputs, a GSA examines the sensitivity measure from the perspective of the entire range of each input's variation. Since many input variables in the power system are uncertain and can influence each other, a GSA is preferable over an LSA which gives an incomplete view of the sensitivities within a system. To perform a GSA, mainly four methods are used [2]: the non-parametric method, the regression-based method, the variance-based method, and the density-based method. Since the last decade, plenty of interpretations and discussions on the variance-based GSA have been presented [4], which highlights that the variance is a versatile and proper indicator to describe the output variability without any hypothesis on the linearity or monotonicity of the model. The variance-based GSA provides information on the contribution of each input variable to the variance of the selected output, either a single variable or in combination with others. In addition, the variance-based GSA conveys analysts and decision-makers with information on the model structure, which brings great advantages compared to the other three methods.

The variance-based GSA was first proposed by Cukier in [5], known as the Fourier Amplitude Sensitivity Test. In the FAST, only the effects of individual inputs were calculated. Later, the total effects were also included by Saltelli A. et al. [6]. Meanwhile, the original FAST was developed further by I. M. Sobol' [7]. In Sobol' method, the sensitivity is evaluated by Sobol' indices which are capable of measuring sensitivities for arbitrary groups of factors by Monte Carlo (MC) simulations. The extended FAST and Sobol' method are similar to Sobol' indices in terms of functionality, the only difference lies in the way of the multidimensional integration. Sobol' indices are traditionally computed by random simulations which require a large number of model evaluations. Hence, the Sobol' method can hardly be applied to computationally expensive models, e.g. power flow equations of a network in large size and of high complexity. In order to accelerate the computation of Sobol' indices, the surrogate model, also named meta-model, comes into play. Commonly-used surrogate models [8] include artificial neural network, polynomial chaos (PC) expansion, kriging, space mapping, etc., among which PC expansion has been applied to the GSA [2]-[9]. The PC expansion intends to build a reduced response model which is statistically equivalent to the original one but is less computation demanding. The advantage of the PC expansion for the variance-based GSA is two-fold: 1) analytical expressions for Sobol' indices can be obtained from the coefficients of PC expansion if the inputs are statistically independent; 2) numerical results can be estimated in an efficient manner if there is a correlation among inputs.

In power systems, the SA has a long history of being applied to a variety of problems. The SA has already been used to assist in designing measurement networks [10], ranking of generators [11] and loads [12],selecting the location for voltage control devices [13], [14], placing distributed generations [15], alleviating voltage problems [16], improving transmission capacity [17], reconfiguring radial networks [18], and also minimizing energy losses [19]. However, the state-of-the-art approaches to the SA in power systems are still adopting the idea of an LSA, which mainly include three methods. The first one uses the inverse of the power flow Jacobian matrix which can be directly obtained from the Newton-Raphson algorithm [20], [21]. This method has been widely applied, especially for the voltage-power sensitivity, due to the commonness of the Newton-Raphson algorithm and its rapid calculation of sensitivities. The second is the so-called 'perturb-and-observe' method which makes a small change to an input factor and records the effect [22]. This method is far more versatile as it does not require a specific algorithm for the power flow calculation. The third one applies Tellegen's theorem and the concept of adjoint networks [23]. Since the sensitivity can be obtained by analyzing the results of a single run of the power flow calculation, the computation of coefficients is thus more efficiently compared with the second method.

The conventional LSA methods, as mentioned above, lay the foundation of the SA for power systems, as they have been used for model simplification, importance rank, and risk reduction. Nevertheless, an LSA is limited to measure the importance of an input around its local conditions. Moreover, the impact of uncertain inputs of the power flow equations including power generations, power consumptions, and network parameters cannot be credibly assessed by an LSA. Hence, it is important to make use of a GSA rather than an LSA in the power system, in view of the growth of complexity as a result of the connection of power electronics interfaces as well as advanced control architectures, and the increase of uncertainty due to the penetration of renewable energy generations. GSA methods have been applied to power system analyses in a limited number of studies. Both the GSA and LSA methods have been

deployed to assess the importance of uncertainty with respect to the system frequency excursion in [24]. It is proved that an LSA is incapable of dealing with group inputs whereas a GSA is able to uncover complex sensitivities. Various types of SA technique including the LSA methods, semi-global methods and global methods were presented in two test networks in [25], [25], with an application of assessing the sensitivity of uncertain inputs to the system small-disturbance stability.

Generally, global methods are more thorough and rigorous for identifying the input-output relationships of medium/high complexity but computationally expensive by using traditional MC-based approaches. In the Kundur two-area network with 35 uncertain parameters, the time required to obtain Sobol' indices even with a small MC simulation numbers 5000 comes up to 9h 25m [25]. In the computationally intensive power system with a lot of uncertain inputs, new approaches are desired to enable and facilitate the implementation of GSA methods. To this end, this paper intends to show how a GSA can be applied to the power system efficiently and result in practical applications. The main contributions of the paper are:

- A novel approach based on the basis-adaptive sparse polynomial chaos (BASPC) expansion is proposed to make a GSA available in complex power systems.
- The variance-based GSA method, Sobol' method, is introduced to the power system analysis systematically;
- The comparison between a GSA based on MC and a GSA with the BASPC in terms of accuracy and efficiency is made using two test networks;
- By means of a GSA with the BASPC, influences of the density functions and correlation conditions of the uncertain inputs on the SCs with respect to different QoIs are discussed.

The paper is organized as follows: Section II introduces the mathematical formulation of the variance-based GSA mthod and Sobol' indices. Section III presents the procedure of how the variance-based GSA and the BASPC expansion are integrated and applied to the power system. The proposed method is verified on both the IEEE 13-bus test feeder and the IEEE 123-node test system in Section IV while conclusions and outlooks are summarized in Section V.

II. THEORETICAL BACKGROUND

Prior to applying a GSA to the power system, the related theoretical background of the variance-based GSA needs to be established. Consider a generic model defined in $\mathbb{R}^M$ with M random inputs $(X_1,\ldots,X_M)$, which is described by the following function:

$$Y = g(\mathbf{X}) = g(X_1, X_2, \cdots, X_M) \qquad (1)$$

where X_j is the j-th input variable and Y is the model response. In addition, the input vector $\mathbf{X}$ has a PDF $\omega_{\mathbf{X}}(x_1,\ldots,x_M)$ and a finite variance over the support $\Pi_{\mathbf{X}}$. The variance-based GSA aims to understand how the variation of a single input or a group of inputs affects the variance of the output. Depending on whether or not the random inputs are correlated, the ANOVA or ANCOVA can be applied to the variance-based GSA.

A. Independent input uncertainties

The ANOVA was introduced by Efron and Stein in [27] to quantify the contribution of each variable X_j to the variance of

the output. The influence of X_j on $\mathrm{Var}[g(\mathbf{X})]$ can be measured using the reduction in $\mathrm{Var}[g(\mathbf{X})]$ if the variation of X_j is known. The resulting variance of $g(\mathbf{X})$ is denoted by $\mathrm{Var}[g(\mathbf{X})|X_j = x_j^*]$ when X_j is fixed to a real value x_j^*. Nevertheless, X_j is a random variable, all of the possible realizations of X_j should be taken into account if one wants to obtain the conditional variance of $g(\mathbf{X})$, by taking all factors apart from X_j. Hence, it is sensible to take an expectation operator on X_j, that is, $\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)]$. Note that a smaller value of $\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)]$ indicates a greater influence of X_j on $\mathrm{Var}[g(\mathbf{X})]$.

Moreover, $\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)]$ plays a critical role in $\mathrm{Var}[g(\mathbf{X})]$ according to the law of total variance which is stated as follows:

$$\mathrm{Var}[g(\mathbf{X})] = \mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)] + \mathrm{Var}[\mathrm{E}(g(\mathbf{X})|X_j)] \qquad (2)$$

where the unconditional variance is the sum of an expectation of the conditional variance $\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)]$ and a variance of the conditional expectation $\mathrm{Var}[\mathrm{E}(g(\mathbf{X})|X_j)]$. The highest value of $\mathrm{Var}[\mathrm{E}(g(\mathbf{X})|X_j)]$ reveals that X_j is the most influential single factor among all the inputs. Accordingly, it is natural to propose the first-order sensitivity coefficient (SC) [28] as what follows:

$$S_j = 1 - \frac{\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|X_j)]}{\mathrm{Var}[g(\mathbf{X})]} = \frac{\mathrm{Var}[\mathrm{E}(g(\mathbf{X})|X_j)]}{\mathrm{Var}[g(\mathbf{X})]}, \ S_j \in [0,1]. \qquad (3)$$

The first-order SC measures the contribution of a single input on the variance of the output, but it is infeasible to explore joint effects among different input factors. Generally, the joint effect between two input factors consists of two aspects: the physical interaction and the statistical correlation.

Attention will be given to the physical interaction in this subsection, in the presence of independent input uncertainties. If physical interactions among input factors exist in a model, the second-order or even the higher-order effects should be analyzed to fully understand the sensitivity pattern of inputs to the outputs. In order to treat the physical interactions, the high-dimensional model representation (HDMR), a generic approach for expanding the response of a model as a series of hierarchical functions of the input variables, is adopted herein and is given by [2]:

$$g(\mathbf{X}) = g_0(\mathbf{X}) + \sum_{j=1}^{M} g_j(X_j) + \sum_{1 \le j < i \le M} g_{j,i}(X_j, X_i) + \cdots + \sum_{1 \le j_1 < \cdots < j_p \le M} g_{j_1,\cdots,j_p}(X_{j_1}, \cdots, X_{j_p}) + \cdots + g_{1,\cdots,M}(X_1, \cdots, X_M) \qquad (4)$$

in which $g_0(\mathbf{X})$ is the mean value of $g(\mathbf{X})$ over the entire domain of $\mathbf{X}$; the first-order function $g_j(X_j)$ denotes an independent effect of input X_j; the higher-order functions represent the cooperative effects of an increasing number of input variables. However, in a given model $g(.)$ there are plenty of choices for component functions in (4). Afterwards, Sobol' proved that $g(\mathbf{X})$ can be decomposed uniquely if $g_0(\mathbf{X})$ is a constant and the integral of each component function over the other variables is zero [4], that is,

$$\int_{\Pi_{X_l}} g_{j_1,\cdots,j_p}(x_{j_1},\cdots,x_{j_p}) \, d_{x_l} = 0,$$
$$\text{for } 1 \le j_1 < j_2 < \cdots < j_p \le M, \ l \in \{j_1, \cdots, j_p\}. \qquad (5)$$

in which Π_{X_l} is the support of random variable X_l. Equation (5), it implies that every first-order function has a zero mean and all the terms are orthogonal in pairs, namely,

$$\int_{\Pi_{\mathbf{X}}} g_{j_1,\cdots,j_p}(x_{j_1},\cdots,x_{j_p}) g_{i_1,\cdots,i_l}(x_{i_1},\cdots,x_{i_l}) \, d_{\mathbf{X}} = 0,$$
$$\text{for } \{i_1,\cdots,i_l\} \ne \{j_1,\cdots,j_p\}. \qquad (6)$$

By integrating the square of (4) and (6), the variance of $g(\mathbf{X})$ can be decomposed by a unique expression as follows:

$$\mathrm{Var}[g(\mathbf{X})] = \sum_{j=1}^{M} D_j + \sum_{1 \le j < i \le M} D_{ji} + \cdots + \\ \sum_{1 \le j_1 < \cdots < j_p \le M} D_{j_1,\cdots,j_p} + \cdots + D_{1,\cdots,M} \tag{7}$$

where the partial variances are defined as:

$$D_{j_1,\cdots,j_p} = \int_{\Pi_{\mathbf{x}_p}} g_{j_1,\cdots,j_p}^2 (x_{j_1},\cdots,x_{j_p}) \, d_{x_{j_1}} \cdots d_{x_{j_p}}, \\ \text{for } 1 \le j_1 < j_2 < \cdots < j_p \le M, \ p = 1,\cdots,M. \tag{8}$$

in which $\Pi_{\mathbf{x}_p}$ is the support of the p-dimensional sub-vector of $\mathbf{X}$. Dividing both sides of (7) by $\mathrm{Var}[g(\mathbf{X})]$, the Sobol' indices of a group of input variables read,

$$S_{j_1,\cdots,j_p} = \frac{D_{j_1,\cdots,j_p}}{\mathrm{Var}[g(\mathbf{X})]}, 1 \le j_1 < j_2 < \cdots < j_p \le M \tag{9}$$

by referring to (7), the Sobol' indices defined in (9) also satisfy,

$$\sum_{j=1}^{M} S_j + \sum_{1 \le j < i \le M} S_{ji} + \cdots + \sum_{1 \le j_1 < \cdots < j_p \le M} S_{j_1,\cdots,j_p} \\ + \cdots + S_{1,\cdots,M} = 1 \tag{10}$$

in which S_j is the first-order SC of X_j to $\mathrm{Var}[g(\mathbf{X})]$, S_{ji} is the second-order SC of the interacted pair (X_j, X_i), and so on for the higher-order SCs. On the basis of (9), the total contribution of X_j to $\mathrm{Var}[g(\mathbf{X})]$ can be calculated by summing up its first-order SC and all of the associated higher-order SCs, as shown in (11).

$$ST_j = \sum_{\Omega_j} S_{i_1,\cdots,i_l}, \\ \Omega_j = \{(i_1,\cdots,i_l): \exists k, 1 \le k \le l, i_k = j\}. \tag{11}$$

However, it is difficult to calculate the total SC of each single input variable via this intuitive approach. The obstacle herein is two-fold: 1) it is cumbersome to ascertain how many physical interactions among input variables exist unless the structure of the model is quite simple; 2) it is infeasible to compute all interaction effects for large systems, since there are up to 2^M-1 terms in (10). For ease of measuring the total effect of X_j on $\mathrm{Var}[g(\mathbf{X})]$, the law of total variance is recalled in another way:

$$\mathrm{Var}[g(\mathbf{X})] = \mathrm{E}[\mathrm{Var}(g(\mathbf{X})|\mathbf{X}_{-j})] + \mathrm{Var}[\mathrm{E}(g(\mathbf{X})|\mathbf{X}_{-j})] \tag{12}$$

where $\mathbf{X}_{-j}$ stands for fixing all variables in the vector $\mathbf{X}$ except for X_j. Similar to the definition of the first-order effect with respect to a single variable in (3), the quantity $\mathrm{Var}[\mathrm{E}(g(\mathbf{X})|\mathbf{X}_{-j})]$ is thus the first-order effect of $\mathbf{X}_{-j}$ on $\mathrm{Var}[g(\mathbf{X})]$, in terms of a group of input variables $(X_1,\ldots, X_{j-1}, X_{j+1},\ldots, X_M)$. On the contrary, $\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|\mathbf{X}_{-j})]$ indicates the expected remaining variance of $g(\mathbf{X})$ that would be left if all possible values of $\mathbf{X}_{-j}$ are known (deterministic). Considering X_j is the only relative component of $\mathbf{X}_{-j}$ in $\mathbf{X}$, the total effect of X_j reads:

$$ST_j = 1 - \frac{\mathrm{Var}[\mathrm{E}(g(\mathbf{X})|\mathbf{X}_{-j})]}{\mathrm{Var}[g(\mathbf{X})]} = \frac{\mathrm{E}[\mathrm{Var}(g(\mathbf{X})|\mathbf{X}_{-j})]}{\mathrm{Var}[g(\mathbf{X})]}, ST_j \in [0,1]. \tag{13}$$

B. Correlated input uncertainties

By means of the total indices in (13), it is able to measure the joint effect due to the physical interactions. Hereafter, it is time to explore the statistical correlation among the input variables. If the input variables of a model $g(.)$ are no longer independent, Sobol' sensitivity indices introduced in Section II.A can still be computed but the interpretation of the total SC becomes difficult, since the correlation makes a contribution to the total effect, in addition to any physical interactions. One possible

approach to handle this problem is to de-correlate the correlated input variables by applying an appropriate transformation. However, the resulting SCs of de-correlated inputs may lose their physical meanings. In this situation, the ANCOVA which generalizes the variance decomposition without hypothesis on the correlation condition of $\mathbf{X}$ will be of benefit to the analysis. Being an extension of ANOVA, the ANCOVA is based on the law of total covariance which is given by:

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}[\mathrm{Cov}(X_2, X_3|X_1)] + \\ \mathrm{Cov}[\mathrm{E}(X_2|X_1), \mathrm{E}(X_3|X_1)] \tag{14}$$

in which X_1, X_2, X_3 are random variables defined on the same probability space, and X_2, X_3 have the finite variances. This changes the decomposition of the $\mathrm{Var}[g(\mathbf{X})]$ which is originally represented in (7) as follows:

$$\begin{aligned} \mathrm{Var}[g(\mathbf{X})] &= \mathrm{Cov}[g_0 + \sum_{\boldsymbol{a} \subseteq \{1,\cdots,M\}} g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}}), \ g(\mathbf{X})] \\ &= \sum_{\boldsymbol{a} \subseteq \{1,\cdots,M\}} \mathrm{Cov}[g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}}), \sum_{\boldsymbol{a} \subseteq \{1,\cdots,M\}} g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}})] \\ &= \sum_{\boldsymbol{a} \subseteq \{1,\cdots,M\}} \{\mathrm{Var}[g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}})] + \mathrm{Cov}[g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}}), \\ &\qquad \sum_{\sim \boldsymbol{a} \subseteq \{1,\cdots,M\}} g_{-\boldsymbol{a}}(\mathbf{X}_{-\boldsymbol{a}})]\} \end{aligned} \tag{15}$$

where the HDMR decomposition of $g(\mathbf{X})$ in (4) is reformulated in a more compact way, i.e.,

$$g(\mathbf{X}) = g_0 + \sum_{\boldsymbol{a} \subseteq \{1,\cdots,M\}} g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}}) \tag{16}$$

with

$$\boldsymbol{a} = (j_1,\cdots,j_p), \mathbf{X}_{\boldsymbol{a}} = (X_{j_1},\cdots,X_{j_p}), \forall 1 \le j_1 < \cdots < j_p \le M, \\ \boldsymbol{a} \bigcup \sim \boldsymbol{a} = \{1,\cdots,M\} \text{ and } \boldsymbol{a} \bigcap \sim \boldsymbol{a} = \varnothing. \tag{17}$$

Based on the general decomposition of the variance in (15), the total effect of $\mathbf{X}_{\boldsymbol{a}}$ on $\mathrm{Var}[g(\mathbf{X})]$ is composed of two parts, namely, $ST_{\boldsymbol{a}} = S_{\boldsymbol{a}|\mathrm{U}} + S_{\boldsymbol{a}|\mathrm{C}}$. The $S_{\boldsymbol{a}|\mathrm{U}}$ stands for the whole physical contribution of $\mathbf{X}_{\boldsymbol{a}}$ on $\mathrm{Var}[g(\mathbf{X})]$ whereas the $S_{\boldsymbol{a}|\mathrm{C}}$ stands for the statistical correlation, as given by:

$$S_{\boldsymbol{a}|\mathrm{U}} = \frac{\mathrm{Var}[g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}})]}{\mathrm{Var}[g(\mathbf{X})]}, S_{\boldsymbol{a}|\mathrm{C}} = \frac{\mathrm{Cov}[g_{\boldsymbol{a}}(\mathbf{X}_{\boldsymbol{a}}), \sum_{\sim \boldsymbol{a} \subseteq \{1,\cdots,M\}} g_{-\boldsymbol{a}}(\mathbf{X}_{-\boldsymbol{a}})]}{\mathrm{Var}[g(\mathbf{X})]}. \tag{18}$$

In the case of a single variable X_j is of concern, it is easy to obtain its SCs based on (18), as shown in (19).

$$S_{j|\mathrm{U}} = \frac{\mathrm{Var}[g_j(X_j)]}{\mathrm{Var}[g(\mathbf{X})]}, S_{j|\mathrm{C}} = \frac{\mathrm{Cov}[g_j(X_j), \sum_{\sim j \subseteq \{1,\cdots,M\}} g_{-j}(\mathbf{X}_{-j})]}{\mathrm{Var}[g(\mathbf{X})]}. \tag{19}$$

C. Computational strategy

In order to obtain the indices in (3), (9) or (18), calculations of the mean values, total and partial variances, and covariance are required, which could be computed using MC simulation. Taking the first-order SC of an input variable X_j for instance, one can estimate the inner conditional expectation first, and then estimate the outer variance, by means of MC trials. In this method, N_{sam}^2 rounds of model evaluation are demanded where N_{sam} is the number of base samples [4]. The approach proposed by Sobol' and then adapted by Saltelli is able to reduce the cost to $(M+2)N_{sam}$ runs. As might be expected, the computational burden of the entire process is fairly heavy, which makes the variance-based GSA infeasible in computationally expensive models in the presence of a large number of random input variables. Since the last two decades, many researchers have been actively improving the efficiency of the computation procedure [28]. However, the computational burden of the

variance-based GSA by using conventional approaches remains a critical issue that impedes its wide application.

III. THE PROPOSED METHOD

The implementation of a variance-based GSA in the power system by means of MC simulations or MC-like methods is a computationally demanding task. For example, the number of nodes might be over one thousand in a distribution network, which gives rise to the computation time for a single run of the power flow calculation up to a few seconds. To enable the utilization of a GSA in complex systems such as the power system, the variance-based GSA can be carried out with the BASPC expansion.

A. Construction of the BASPC

As mentioned above, numerous evaluations of the model are required in the variance-based GSA, thus the computational burden depends, to a large degree, on the complexity (e.g. structure and size) of the model in question. In this context, the PC expansion can be applied, since it is able to resemble any output of the original system in the form of a mathematical model with reduced complexity. The PC expansion was first introduced by Wiener [29], which exclusively addressed standard normal variables. Xiu et al. extended this method to handle several more distribution types within a family of orthogonal polynomials [30]. By means of an iso-probabilistic transform other distribution types can be applied as well.

Given a finite dimensional random space $L^2(\Omega, F, Pr)$ where Ω is a sample space, F is the σ-algebra on Ω, and Pr stands for a probability measure on (Ω, F). According to the Cameron-Martin theorem, any stochastic response with a finite second moment can be expanded in a convergent series of orthogonal polynomials of the inputs. Let X be a random input variable, the stochastic response $g(X)$ can be represented by [30]:

$$g(X) = \sum_{i=0}^{\infty} a_i \Phi_i(X) \qquad (20)$$

where $\{a_i\}$ is the coefficient of the i-th polynomial, $\{\Phi_i\}$ is the i-th univariate polynomial basis. The orthogonality of the basis $\{\Phi_i\}$ with respect to a probability measure can be expressed as:

$$\mathrm{E}[\Phi_i \Phi_j] = \int_\Omega \Phi_i(x) \Phi_j(x) \, \omega_x(x) \, dx = \delta_{ij} \, \mathrm{E}[\Phi_i^2] \qquad (21)$$

in which δ_{ij} is the Kronecker delta function, and ω_X is the PDF of X. For a group of independent variables $\mathbf{X} = \{X_1, X_2, \ldots, X_M\}$ the corresponding model response is represented by,

$$g(\mathbf{X}) = \sum_{i=0}^{\infty} a_i \Psi_i(\mathbf{X}) \qquad (22)$$

where $\{\Psi_i\}$ is the multivariate orthogonal polynomial basis that can be constructed as a tensor product of univariate orthogonal polynomials as follows:

$$\Psi_i(\mathbf{X}) = \Phi_{i_1}(X_1) \otimes \Phi_{i_2}(X_2) \otimes \cdots \otimes \Phi_{i_M}(X_M) \qquad (23)$$

in which the subscript i_s ($s=1,\ldots,M$) refers to the s-th degree of the i-th univariate polynomial basis. Further, the orthogonality of $\{\Psi_i\}$ still holds with regard to the joint PDF of the random vector $\mathbf{X}$. For practical reasons, the infinite expression has to be truncated to a finite number of terms according to a preset criterion. Most often, the PC expansion is truncated in such a way that only those polynomial bases in (23) with a total degree lower or equal to d are maintained. This truncation strategy leads to the number of expansion terms as given by,

$$K^{M,d} = \frac{(M+d)!}{M! \, d!} - 1 \qquad (24)$$

When the number of random inputs M increases linearly, the conventional PC expansion encounters a polynomial growth of the number of expansion terms, which is so-called the curse of dimensionality [31]. In this context, the sparse-adaptive scheme is proposed to overcome this problem [32]. In the conventional truncation strategy, each univariate polynomial of the tensor product is treated as having the same importance. However, according to the sparsity-of-effects principle the main effect and lower order interactions dominate a stochastic response. That is, the interactions among univariate polynomial bases with lower degrees are of more statistical significance, compared with the interactions among higher-degree terms. As a result, one way to reduce the number of expansion terms $K^{M,d}$ +1 is to use a sparse truncation strategy that prefers low-degree interactions [32],

$$I_q^{M,d} \equiv \{ \mathbf{i} \in \mathbb{N}^M : \| \mathbf{i} \|_q = \sqrt[q]{\sum_{j=1}^M (i_j)^q} \le d, \ q \in (0,1] \} \qquad (25)$$

where $\mathbf{i}$ is the hyperbolic index, $I_q^{M,D}$ is the hyperbolic index set, and q is the truncation norm that is used to tune the sparsity of expansion. The total number of terms in the set $I_q^{M,d}$ is denoted by $K_q^{M,d}$, which corresponds to the classical $K^{M,d}$ at $q=1$. If q becomes closer to zero, $K_q^{M,d}$ will be significantly reduced since the high-degree interactions in (23) are penalized, and almost only additive terms are retained.

A value has to be assigned to d, but no general conclusions have been drawn on how to select the best value since d is problem-dependent. Instead of selecting the value by trial and error, an adaptive scheme can be employed, which sets up a range of candidates first, and then selects the best one through accuracy evaluations on the resultant PC expansions. By taking advantage of both the adaptive scheme and sparse truncation strategy, an improved PC expansion that is capable of tackling the problem of high dimensionality is defined, namely, the basis-adaptive sparse polynomial chaos (BASPC) [33].

After building up the truncation strategy, the coefficients $\{a_i\}$ should be determined at the final step. In recent literature, the linear regression method, also known as the stochastic response surface method, is advocated [34]. This method evaluates the coefficients by directly casting an optimization problem. Let $\boldsymbol{\xi} = \{\mathbf{x}_1,\ldots,\mathbf{x}_{N_{ED}}\}$ be a small set of representatives of the random input vector, also named as the experimental design (ED), and $\boldsymbol{\eta} = \{g(\mathbf{x}_1),\ldots,g(\mathbf{x}_{N_{ED}})\}^{\mathrm{T}}$ be the corresponding stochastic response vector, then the coefficients can be calculated by minimizing the mean-square of the residuals as follows:

$$\min_{\mathbf{a}} \sum_{l=1}^{N_{ED}} [g(\mathbf{x}_l) - \mathbf{a}^{\mathrm{T}} \mathbf{H}(\mathbf{x}_l)]^2 \qquad (26)$$

where $\mathbf{a}$ is the vector of coefficients $\{a_i\}$ ($i=0,\ldots,K_q^{M,d}$), and the experimental matrix $\mathbf{H}$ writes:

$$\mathbf{H} = \begin{bmatrix} \Psi_0(\mathbf{x}_1) & \cdots & \Psi_{K_q^{M,d}}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(\mathbf{x}_{N_{ED}}) & \cdots & \Psi_{K_q^{M,d}}(\mathbf{x}_{N_{ED}}) \end{bmatrix} \qquad (27)$$

Through the ordinary least-squares method, the coefficient vector $\mathbf{a}$ can be computed and expressed as $\hat{\mathbf{a}} = (\mathbf{H}^{\mathrm{T}} \mathbf{H})^{-1} \mathbf{H}^{\mathrm{T}} \boldsymbol{\eta}$. In the case of ill-conditioned matrix $\mathbf{H}^{\mathrm{T}} \mathbf{H}$, a special treatment such as the singular value decomposition should be employed. Apart from the revised truncation strategy in (25), the high

degree interactions in (22) can be further reduced by adding a penalty term on the right-side of (26) as:

$$\min_{\mathbf{a}}\{\sum_{l=1}^{N_{ED}}[g(\mathbf{x}_l) - \mathbf{a}^{\mathrm{T}}\mathbf{H}(\mathbf{x}_l)]^2 + \lambda \sum_{l=1}^{N_{ED}} a_l\} \quad (28)$$

where λ is a positive penalty factor. In order to solve (28), least angle regression (LAR) can be employed to deal with this task. LAR is a regression tool for fitting a linear model to high-dimensional data, which is effective when the number of regressors is much greater than the size of available ED. In the context of PC expansion, LAR has already been applied to build the sparse PC expansion, with the detailed steps explained in [32]. In particular, LAR suits for the case in which the number of predictors is of the similar size as or even greater than the number of observations. In this context, it is possible that the number of N_{ED} is smaller than the number of polynomial coefficients to be estimated/predicted, which in turn reduces the number of calls of the original model $g(.)$.

As for the variance-based GSA, the main attractive feature of BASPC expansion is the orthogonality of the basis $\{\Psi_i\}$. Due to this merit, it is possible to obtain statistical moments (such as the mean and variance) of the analytical stochastic response $g(\mathbf{X})$ by means of the coefficients $\{a_i\}$, as shown in (29). Since the calculation of SCs in the GSA largely demands statistical moments of $g(\mathbf{X})$, the BASPC expansion is therefore able to considerably ease the computational burden.

$$\mathrm{E}[g_{PC}(\mathbf{X})] = \mathrm{E}[\sum_{i=0}^{K_q^{M,d}} a_i \Psi_i(\mathbf{X})] = a_0,$$
$$\mathrm{Var}[g_{PC}(\mathbf{X})] = \sum_{i=1}^{K_q^{M,d}} a_i^2 \, \mathrm{E}[\Psi_i^2(\mathbf{X})]. \quad (29)$$

B. Variance-based GSA with the BASPC

In the presence of independent inputs, the computation of the variance-based GSA SCs with the BASPC expansion requires a combination of (3), (13) and (22). Therefore, the PC expansion of $g(\mathbf{X})$ is rewritten to its Sobol' decomposition as:

$$g_{PC}(\mathbf{X}) = a_0 + \sum_{j=1}^{M} \sum_{\beta \in \Omega_j} a_\beta \Psi_\beta(X_j)$$
$$+ \cdots + \sum_{1 \le j_1 < \cdots < j_p \le M} \sum_{\beta \in \Omega_{j_1,\cdots,j_p}} a_\beta \Psi_\beta(X_{j_1}, \cdots, X_{j_p}) \quad (30)$$
$$+ \cdots + \sum_{\beta \in \Omega_{1,\cdots,M}} a_\beta \Psi_\beta(X_1, \cdots, X_M)$$

in which

$$\Omega_{j_1,\cdots,j_p} = \left\{ \beta: \begin{array}{ll} a_l > 0 & \forall l \in \{j_1,\cdots,j_p\} \subseteq \{1,\cdots,M\} \\ a_l = 0 & \forall l \in \{1,\cdots,M\} \setminus \{j_1,\cdots,j_p\} \end{array} \right\}, \quad (31)$$

$$\beta = \{a_l, l = 1,\cdots,M, a_l \ge 0, K_q^{M,d} \ge \sum_{l=1}^{d} a_l\}.$$

Compared to the original Sobol' decomposition of $g(\mathbf{X})$, the summands in (4) are reorganized as:

$$g_{j_1,\cdots,j_p}(X_{j_1},\cdots,X_{j_p}) = \sum_{\beta \in \Omega_{j_1,\cdots,j_p}} a_\beta \Psi_\beta(X_{j_1},\cdots,X_{j_p}) \quad (32)$$

Based on (29), the first-order and total sensitivity indices of a group of input variables read:

$$S_{j_1,\cdots,j_p} = \frac{\sum_{\beta \in \Omega_{j_1,\cdots,j_p}} a_\beta^2 \Psi_\beta^2(X_{j_1},\cdots,X_{j_p})}{\sum_{i=1}^{K_q^{M,d}} a_i^2 \, \mathrm{E}[\Psi_i^2(\mathbf{X})]}, \quad (33)$$

$$ST_{j_1,\cdots,j_p} = \sum_{\{i_1,\cdots,i_l\} \in \Omega_{j_1,\cdots,j_p}} S_{i_1,\cdots,i_l}.$$

Since the summands in (33) can be directly derived from the polynomial coefficients in (29), the SCs shall be obtained as a byproduct of the BASPC expansion. Once $g_{PC}(\mathbf{X})$ is obtained, analytical solutions of all the Sobol' indices can be calculated with negligible computational efforts.

In the case of correlated inputs, the bases of PC expansion are no longer orthogonal with respect to the joint PDF, thus (33) cannot be applied directly. Nevertheless, the BASPC expansion of $g(\mathbf{X})$ can be used as a surrogate model to accelerate the calculation of the GSA SCs.

C. Procedure of variance-based GSA with the BASPC

In previous sections, the proposed method for the GSA has been discussed in the form of a generic model $g(.)$, M random inputs $(X_1,\ldots,X_M)$ and a scalar output Y, as shown in (1). With regard to the implementation of the proposed method in power systems, the power flow equations correspond to model $g(.)$, for the purpose of a steady-state analysis.

The power flow calculation enables engineers to diagnose and prognose whether the state variables are within the normal operating range for various grid situations. Without the loss of generality, a mathematical model of the single-phase equivalent power flow in polar form can be introduced by the following equations [35],

$$\begin{cases} P_i = V_i \sum_{j=1}^{n} V_j(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}), & i = 1,\cdots,n \\ Q_i = V_i \sum_{j=1}^{n} V_j(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}), & i = 1,\cdots,n \end{cases} \quad (34)$$

where P_i and Q_i are the active and reactive power injections at Bus i; V_i and θ_i are the magnitude and angle of the voltage at Bus i; G_{ij} and B_{ij} denote the conductance and susceptance of the branch connecting Bus i and Bus j, respectively, while the angle difference between the two involved buses is represented in the form of $\theta_{ij} = \theta_i - \theta_j$. In addition, n is the total number of buses.

When performing an SA on the power flow equations as shown in (34), the input variables are usually the power injections into each bus, namely, P_i and Q_i. Nonetheless, G_{ij} and B_{ij} might be also regarded as inputs for some studies, e.g., the network planning. The desired outputs are typically the voltage magnitudes and angles at buses, V_i and θ_i. In addition, some derived quantities, such as the power losses or the voltage unbalance factor (VUF), could be considered as outputs as well. With these inputs and outputs in mind, the following procedures are proposed for the application of a GSA in the power system:

Procedure 1: Independent Case
1. Determine the type and parameters of the PDFs of the input variables;
2. Construct the BASPC representation of each output:
1) Generate an $M \times N_{ED}$ matrix $\mathbf{H}_1$ of inputs as the ED;
2) Execute a batch of deterministic power flows (DPF), with system parameters given by the columns of $\mathbf{H}_1$;
3) Build up the BASPC expansion;
3. Derive the Sobol' decomposition of the output;
4. Calculate SCs via coefficients of the BASPC expansion.

Procedure 2: Correlated Case
1. Determine the type and parameters of the PDFs of the input variables;
2. Construct the BASPC representation of each output:
1) Generate an $M \times N_{ED}$ matrix $\mathbf{H}_1$ of inputs as the ED;
2) Execute a batch of DPFs, with system parameters given by the columns of $\mathbf{H}_1$;
3) Build up the BASPC expansion;
3. Generate an $M \times N_{sam}$ matrix $\mathbf{H}_2$ of random inputs;

4. Calculate SCs via the surrogate model and matrix $\mathbf{H}_2$.

In these two procedures, steps *1), 2), 4)* are the same whereas the step *3)* is distinguishable. After constructing the BASPC expansion for the outputs, the usage of the BASPC expansion is different. In **Procedure 1**, the SCs are directly calculated from the BASPC expansion as shown in (33), while in **Procedure 2**, the BASPC expansion is used to reduce the computational effort of each run of the DPF calculation. For clarity, the differences are clearly illustrated by the flowcharts of the two procedures as depicted in Fig. 2.

IV. TEST RESULTS AND DISCUSSIONS

In this section the method explained in Section III is applied on two IEEE test systems, in which the power of the loads and generators are the random inputs and the considered outputs are the bus voltages, branch currents, power losses and so on. The computations are performed in MATLAB R2015a on an Intel Core 4 Quad CPU at 3.40 GHz with 8 GB RAM. For the calculation of the DPF, the OpenDSS [36] is used and for the construction of the BASPC expansion, the UQLab toolbox [37] is adopted.

A. The IEEE 13-Bus Test Feeder

First, the IEEE 13-bus test feeder [38], as shown in Fig. 1, is used to test the performance of the proposed method. The width of arrows are altered to illustrate the differences in the load magnitude.

1) Comparison between GSA and LSA

To compare the results of the GSA, the perturb-and-observe method is also applied to this network to obtain the SCs from an LSA. In the perturb-and-observe method, small perturbations in the inputs are required. The injected value is generally a certain percent of the average bus load P_{ave} and/or Q_{ave}, e.g., between 0.5% and 2% [20]. If the variation of a certain bus power itself is of interest, a percentage of the magnitude of the load P_i and Q_i ($i=1,…,n$) can be considered as the injection.
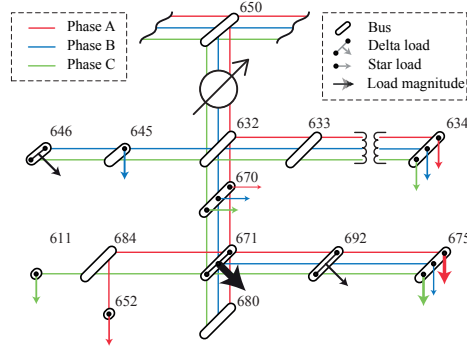
Fig. 1. The IEEE 13-bus test feeder.

In a radial distribution network, the voltage magnitude at the remote buses are one of the main areas of concern. Therefore, the voltage magnitude of phase A at Bus 675, $V_{675\text{-}A}$, is selected for the comparison. The subscript A, B and C represent the associated phase of an input or output in the remainder of this paper. Four different perturbations of the load bus active power (BAP) have been applied. The resulting SCs of $V_{675\text{-}A}$ for the perturb-and-observe method are shown in Fig. 3.

It can be seen that the LSA SCs of $V_{675\text{-}A}$ to the four injections are very close. In addition, the active power injection into phase A of Bus 675 is detected as the most influential factor whereas the active power injection into Bus 646 has the least impact on $V_{675\text{-}A}$. The SCs obtained from an LSA are able to provide applicable results in the model with deterministic inputs and a monotonic input/output relationship. In other cases, the use of an LSA might not be appropriate since the computed SCs cannot reveal the cause-and-effect relationships precisely. To illustrate differences in results between an LSA and a GSA, the proposed method is implemented for four scenarios that are comparable to the LSA scenarios, as listed in. Table I. Instead of considering deterministic/known perturbations on the input variables in an LSA, the injections in a GSA are treated as random variables.
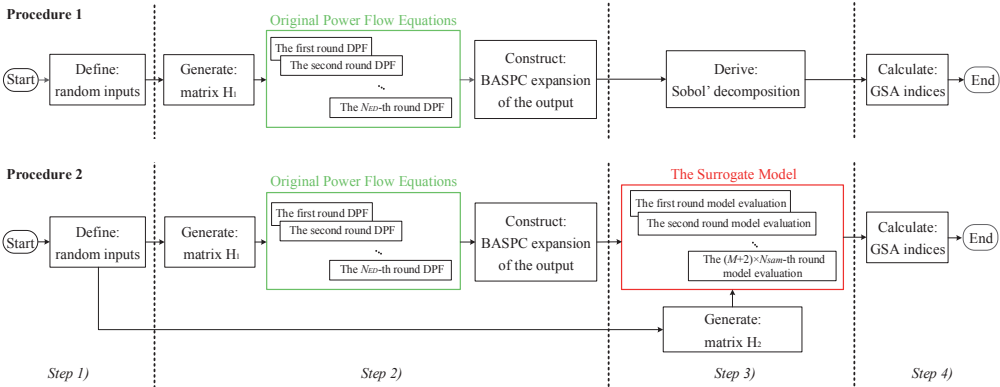
Fig. 2. Flowcharts of the two proposed procedures of GSA with BASPC
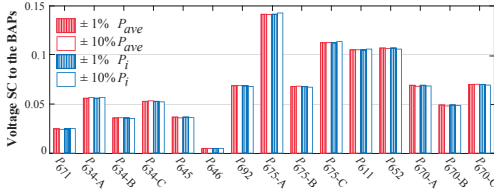
Fig. 3. LSA SCs of V_{675-A} to BAPs using the perturb-and-observe method.

TABLE I
INPUT UNCERTAINTIES IN SCENARIOS 1, 2, 3, 4

Scenario	BAP	Scenario	BAP
1	$U(P_i -0.01P_{ave}, P_i +0.01P_{ave})$	2	$N(P_i +0.01P_{ave}, 1^2)$
3	$U(P_i -0.03P_i, P_i +0.03P_i)$	4	$N(P_i, (0.01P_i)^2)$

*$N()$ stands for the Normal distribution and $U()$ for the Uniform distribution; P_{ave}=231.07 kW.

To implement a GSA in the IEEE 13-bus test feeder, the BASPC is configured with the truncation norm, q, 0.8, the range of truncation criteria, d, 1-6, and the number of ED, N_{ED}, 500. In Fig. 4, the GSA SCs of V_{675-A} to the uncertain BAPs are depicted. The four scenarios have been divided into two groups by the resulting SCs. It is noted that similar SCs are obtained for Scenarios 1-2, with the most influential factor being the variation of P_{652}. The SCs in Scenarios 3-4 show more variation, but are still close to each other, with the most influential factor being the variation of P_{675-A}. The main difference in the results between the two groups lies in which input factor is dominant. In Scenarios 3-4, the factor with a larger base load is more influential. This is essentially caused by the distinct variances of inputs, that is, the absolute magnitudes of the BAPs are taken into account in the associated variances in Scenarios 3-4 while the variances are constants in Scenarios 1-2. Compared to the LSA SCs in Fig. 3, the GSA SCs of a certain input on V_{675-A} do display a change if there are differences in the magnitude of the variation of an input. The results show that the GSA is more viable since the diversity among different scenarios can be revealed, whereas this diversity does not appear if an LSA is performed.
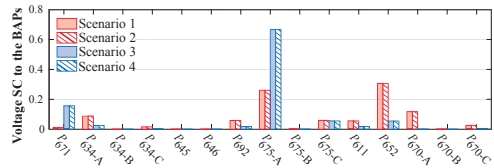

Fig. 4. GSA SCs of V_{675-A} to the BAPs in Scenarios 1, 2, 3, 4.

2) Influence of the modelling of the input uncertainties

As shown in Fig. 4, the variances of the inputs impact the GSA SCs. Comparative investigations on different variances of input uncertainties are therefore performed, based on Scenarios 3, 4 and two additional scenarios as listed in Table II.

TABLE II
INPUT UNCERTAINTIES IN SCENARIOS 5, 6

Scenario	5	6
BAP	$U(-P_i, P_i)$	$N(P_i, (0.5P_i)^2)$

In Fig. 5, the GSA SCs of V_{675-A} to the variation of BAPs are illustrated. One can find that the distribution type of the inputs has a minor influence on the GSA SCs if the magnitudes of

their variances are close. By comparing Fig. 4 and Fig. 5, it is found that the relative magnitude of the variance of a certain input determines the SCs more than the absolute magnitude of the variance employed in the GSA.
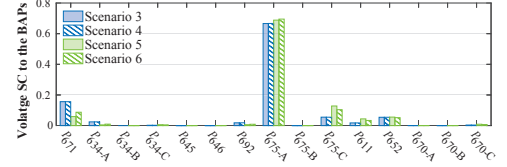
Fig. 5. GSA SCs of V_{675-A} to the BAPs in Scenarios 3, 4, 5, 6.

In view of above results, the determination of the individual input variances should take into account for the application of the GSA SCs. For instance, Scenarios 1-2 which consider the same variation of all BAPs can be used for the allocation of voltage control devices [14]; while Scenarios 3-6 that treat the variation as a relative quantity of each BAP, thus suit for the active power curtailment [16].

3) Influence of the correlation among input uncertainties

In Scenarios 1-6, the input uncertainties are assumed to be independent. For the purpose of investigating the influence of stochastic correlation, Scenarios 7-10 are designed as listed in Table III. For clarity, only one pair of input uncertainties are assumed to be statistically correlated in each scenario, and all other settings are kept the same as in Scenario 4.

TABLE III
CORRELATION BETWEEN INPUT UNCERTAINTIES IN SCENARIOS 7, 8, 9, 10

Scenario	7	8	9	10
BAP	(P_{692}, P_{675-A})	(P_{692}, P_{675-A})	(P_{652}, P_{675-A})	(P_{692}, P_{675-A})
Coefficient	r=0.2	r=0.6	r=0.6	r_s=0.6

*r: Pearson product-moment correlation; r_s: Spearman's rank correlation.

In (18), the GSA SC of an input that is statistically correlated to other inputs consists of a correlated part and an uncorrelated part. In the presence of positive correlations between two inputs, the SCs of V_{675-A} to the BAPs are illustrated in Fig. 6.
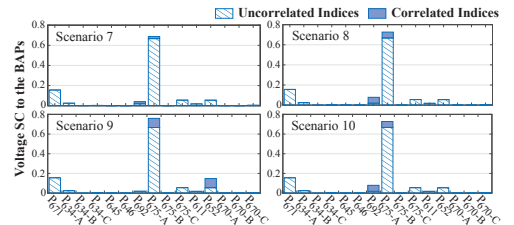
Fig. 6. Influence of the correlation on GSA SCs of V_{675-A} to the BAPs.

It shows that only the two correlated inputs get correlated indices while the rest of inputs have zero correlated indices. As a result, only the total SCs of the two corresponding inputs increase due to this correlation, which is detectable in the four scenarios. By comparing Scenario 7 to 8, it is noticed that a higher positive correlation leads to a greater correlated index due to the fact that a greater Pearson correlation coefficient leads to a higher covariance in (19). To be more specific, the covariance between the random inputs P_{692} and P_{675-A} is the product of r, standard deviations of P_{692} and P_{675-A}. Moreover, the correlated index is related to its uncorrelated counterpart

(physical contribution), by comparing Scenario 8 to Scenario 9. The correlated indices are larger in Scenario 9, since the uncorrelated index of P_{652} indeed exceeds that of P_{692}.

The statistical correlation between two random variables indicates the strength of their linkage. Different considerations of the linkage lead to different definitions of the correlation coefficient. To test the feasibility of the GSA to deal with other types of correlation in addition to the Pearson product-moment correlation, the Spearman's rank correlation is investigated in Scenario 10. The Spearman's rank correlation has been widely used in the power system analysis, especially for the correlation modelling of the wind power [39].

4) GSA for other electrical quantities

In subsections *1-3*, only the sensitivity of a single voltage magnitude with respect to the active power at the load buses has been discussed. However, there are many other QoIs which can be investigated by using a GSA. Hence, the GSA SCs of current magnitudes, VUFs, and active power losses to the BAPs are discussed as well. To start off the sensitivities of the voltage and current magnitudes to the fifteen base loads for Scenario 4, GSA SCs are ranked in a descending order in Table IV. With this table, the relative importance of each input uncertainty to a certain QoI can be easily distinguished. One can quickly see that the same input uncertainty might have quite a distinctive influence on different QoIs, such as P_{675-A} on V_{675-B} and V_{675-C}. Input P_{671}, however, has a large influence on each of the QoIs.

TABLE IV
RANKING OF THE BAPs IN TERMS OF VOLTAGE AND CURRENT SCs

BAP	Voltage Magnitude at Bus 675			Current Magnitude in Line 650-632		
	A	B	C	A	B	C
P_{671}	2nd	2nd	2nd	2nd	1st	1st
P_{634-A}	6rd	7th	13th	3rd	8th	12th
P_{634-B}	12th	6rd	10th	14th	4th	9th
P_{634-C}	9th	15th	5th	10th	15th	4th
P_{645}	11th	5th	8th	11th	2nd	8th
P_{646}	13th	10th	6th	13th	3rd	7th
P_{692}	7th	11th	7th	5th	11th	6th
P_{675-A}	1st	1st	12th	1st	7th	11th
P_{675-B}	14th	4th	9th	12th	5th	13th
P_{675-C}	3rd	9th	1st	7th	10th	2nd
P_{611}	5th	12th	3rd	8th	14th	3rd
P_{652}	4th	3rd	14th	4th	9th	14th
P_{670-A}	10th	14th	15th	6th	12th	15th
P_{670-B}	15th	8th	11th	15th	6th	10th
P_{670-C}	8th	13th	4th	9th	13th	5th

Regarding the VUF at Bus 675, Fig. 7(a) shows the results via a pie chart of six sectors in line with their SCs. The chart shows that only five individual inputs account for 84% of the total SC of VUF$_{675}$, with which the actions on improving the voltage unbalance at Bus 675 can be focused. Similarly, in Fig. 7(b), it can be seen that three of fifteen inputs have 88% of the total SC of the active power loss. This indicates that the input P_{671} should be considered first if one wants to reduce the active power loss. However, the importance of an input on a certain QoI might be different when the variances of all inputs are identical, as in Scenario 1-2. Fig. 8 shows the SCs of the active power loss to the BAPs in Scenario 2. In Fig. 8 (a), it shows that P_{671} is no longer the predominant factor. Furthermore, the uncertain inputs whose effects are negligible can be identified as well, as illustrated in Fig. 8 (b).
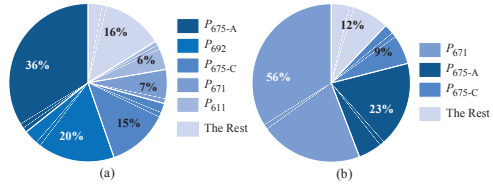
Fig. 7. GSA in Scenario 4: (a) SCs of VUF at Bus 675 to the BAPs; (b) SCs of the total active power loss to the BAPs.
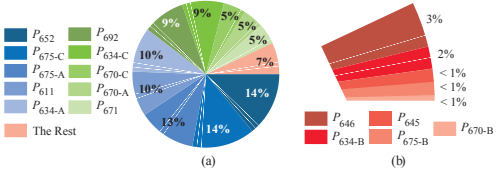
Fig. 8. GSA in Scenario 2: (a) SCs of the total active power loss to the BAPs; (b) Zoom-in on the 7% portion of (a).

5) GSA for the evaluation on a group of outputs

In the previous subsections, attention has been paid to the GSA of a single output with respect to all the input uncertainties. However, the overall effect of the inputs on a specific class of outputs is beneficial for the decision-making, especially in complex networks with high-dimensional input uncertainties. To illustrate this, the GSA SCs with respect to the eight VUFs are shown in Fig. 9.
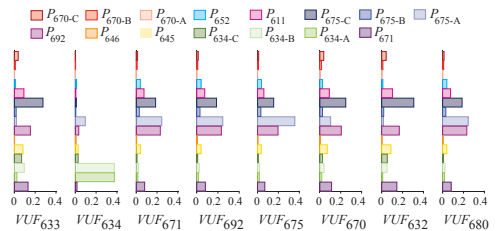
Fig. 9. Overall GSA SCs of VUFs to the BAPs.

By a horizontal comparison, it becomes clear that P_{692} and P_{675-C} have large cumulative effects on the voltage unbalance in this system while P_{646} and P_{670-A} have minor influences. It can also be found that P_{634-A} and P_{634-B} have predominantly impacts on VUF$_{634}$ if one vertically compares the results.

B. The IEEE 123-Node Test System

In order to assess the proposed method on a more complex power system, the variance-based GSA has also been applied to the IEEE 123-node test system. For the identification of the buses in this network, the original bus numbering [38] is used herein. The active power at each load bus is treated as a normal random variable, $N(P_i, (0.01P_i)^2)$ ($i=1,\dots,95$). For the GSA in this test system, the BASPC expansion is configured with q, 0.6; d, 1-6; and N_{ED}, 1000.

In this section, the current magnitude in a line adjacent to the transformer, Line 1-149, and the total active power loss are focused upon. Fig. 10 shows the GSA SCs of the three-phase current magnitudes in Line 1-149 to the BAPs. With this figure, the most significant inputs on the three QoIs are uncovered,

which will provide useful information when congestions occur. Specifically, inputs #53, #57 and #65, which correspond to the three largest current SCs in phase A, B and C, respectively, should be given more attentions in order to perform congestion management effectively.
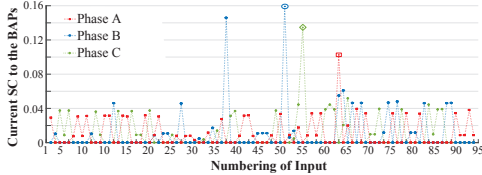
Fig. 10. Overall GSA SCs of current magnitudes in Line 1-149 to the BAPs.

Fig. 11 illustrates the GSA SCs of the total active power loss to the BAPs. It can be found that inputs #65 and #94 have the largest SCs among the 95 inputs. The global sensitivity of the total active power loss with respect to a certain BAP depends on many factors, which enables one to gain more insight into the system. In Fig. 11, the ratio of each load to the total amount of active power, r_p, and the ratio of the electrical distance between the load bus and the transformer to the total distance, r_d, are illustrated in combination with the SC of the active power loss. One can notice that the total active power loss is more sensitive to inputs which have higher values of r_p or r_d. For instance, input #65 has the largest value of r_p, and input #94 corresponds to the second largest value of r_d. In a similar fashion to improve the mitigation of network congestion, inputs #94 and #65 are of prime importance for the total active power loss reduction in the IEEE 123-node test system.
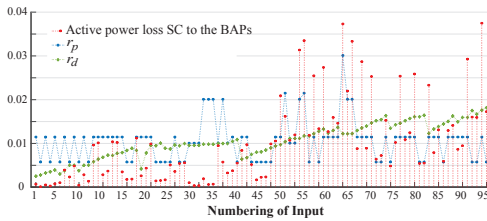
Fig. 11. GSA SCs of the total active power loss to the BAPs.

C. Comparison between GSA with BASPC and MC

As mentioned in Sections II and III, the variance-based GSA with the BASPC is capable of improving the computational efficiency to a great extent while not sacrificing on precision, compared to the traditional approaches. In order to quantify the accuracy of the proposed method relative to a reference method, the hamming distance (HD) and earth mover's distance (EMD) are adopted as the metrics [40]. For the importance ranking of input uncertainties, the HD is used to measure the sorted list of inputs in accordance with their SCs. The HD is an indication of the percentage difference between two vectors of equal length. A zero value of HD means that the two vectors are identical. If the numerical values of SCs are of interest, the EMD is employed to measure the dissimilarity between two groups of SCs obtained from the two methods. The EMD is the minimum cost needed to transform one vector into the other. Apart from the HD and EMD, the simulation time (SimT) is

presented in order to look into the computational cost of each of the methods.

Let the input uncertainties of both the test systems be normal variables $N(P_i, (0.01P_i)^2)$, where P_i is the nominal value of the active power at each load bus. Then, by taking results of the GSA based on a MC simulation with 10^6 trials as a reference, the performance of the GSA with the BASPC is displayed in Table V. Note that the HD of 0.133 in the IEEE 13-bus test feeder and 0.021 in the IEEE 123-node test system indicate that the importance of only two random inputs is incorrectly ranked. Moreover, in terms of the value of the EMD, the BASPC with N_{ED}, 200 outperforms the MC with N_{sam}, 10^5 in the IEEE 13-bus test feeder; and the performance of the BASPC with N_{ED}, 500 is comparable to that of the MC with N_{sam}, 10^5 in the IEEE 123-bus test feeder, giving an indication of the required size of the EMD.

TABLE V
OVERALL PERFORMANCE OF THE VARAINCE-BASED GSA WITH BASPC

QoI	Method	$N_{sam/ED}$	SimT	HD	EMD
IEEE 13-Bus Test Feeder: $V_{675\text{-}A}$	MC	10^6	**8.05(h)**	-	-
		10^5	2.38(h)	0.266	0.096
	BASPC	1000	8.65(s)	0.133	0.077
		500	6.96(s)	0.133	0.078
		200	1.96(s)	0.266	0.084
IEEE 123-Node Test System: $V_{450\text{-}A}$	MC	10^6	**152.07(h)**	-	-
		10^5	31.38(h)	0.063	0.358
	BASPC	2000	32.61(s)	0.021	0.199
		1000	14.66(s)	0.021	0.214
		500	8.14(s)	0.084	0.367

* $N_{sam/ED}$: number of samplers/the experimental design; SimT: simulation time; HD: hamming distance; EMD: earth mover's distance.

A variance-based GSA with the BASPC in both test systems achieves a high level of accuracy in terms of both the HD and EMD. The proposed method significantly reduces the computational burden in terms of the SimT. With regard to the complexity of a power system, it does not give rise to the difficulty of executing the variance-based GSA; but it increases the computation time of both methods. Nevertheless, the more complex a power system is, the more computational effort will be saved by applying the BASPC expansion. Furthermore, the performances of the GSA with the BASPC in both test systems vary with the size of the ED. Along with a growing size of N_{ED}, the result of the GSA with the BASPC gets more accurate whereas the calculation time slightly increases.

V. CONCLUSIONS

This paper builds the bridge between the variance-based GSA and the power system, and presents a novel approach to implement the GSA by applying the BASPC expansion for the surrogate modelling.

In contrary to the LSA, the GSA can include the uncertainties present within a power system in the SA. The proposed method is able to efficiently analyze the importance of a large number of independent or correlated input uncertainties in the power system. The proposed method was implemented on the IEEE 13-bus test feeder and the IEEE 123-node test system, where the voltage-power, current-power, VUF-power and loss-power sensitivities have been investigated. The results show the merit of the proposed method for the analysis of complex power systems in the presence of numerous uncertainties, especially for online usages. Additional analyses on the level of accuracy

and the computational burden of show that the proposed method is able to greatly improve the computational efficiency relative to the conventional MC-based GSA. The adoption of the BASPC within the GSA makes the GSA available for large complex networks such as the power system. It gives rise to a possible broad area of applications for the GSA in power systems, such as network planning and load management.

VI. References

[1] W. Li, Risk assessment of power systems: models, methods, and applications. John Wiley & Sons; 2014.
[2] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions." *Reliability Engineering & System Safety*, vol. 93, no. 7, pp. 964-979, 2008.
[3] M.M. Rajabi, et al., "Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations." *Journal of Hydrology*, vol.520, pp.101-122, 2015.
[4] A. Saltelli, et al. Global sensitivity analysis: the primer. John Wiley & Sons, 2008.
[5] R.I. Cukier, et al., "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory." *The Journal of chemical physics*, vol. 59, no. 8, pp.3873-3878, 1973.
[6] A. Saltelli, S. Tarantola and K.S. Chan, "A quantitative model-independent method for global sensitivity analysis of model output." *Technometrics*, vol. 41, no. 1, pp. 39-56, 1999.
[7] I.Y.M. Sobol', "On sensitivity estimation for nonlinear mathematical models." *Matematicheskoe Modelirovanie*, vol. 2, vol. 1, pp. 112-118, 1990.
[8] N.V. Queipo, et al., "Surrogate-based analysis and optimization." *Progress in aerospace sciences*, vol. 41, no. 1, pp.1-28, 2005.
[9] G. Blatman, and B. Sudret, "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions." *Reliability Engineering & System Safety*, vol. 95, no. 11, pp.1216-1229, 2010.
[10] D.H. Choi, L. Xie. "Sensitivity Analysis of Real-Time Locational Marginal Price to SCADA Sensor Data Corruption." *IEEE Trans. on Power Systems*, vol. 29, no. 3, pp. 1110-1120, 2014.
[11] F.B. Alhasawi and J.V. Milanovic, "Ranking the Importance of Synchronous Generators for Renewable Energy Integration." *IEEE Trans. Power Systems*, vol. 27, pp. 416-423, 2012.
[12] A. M. Leite da Silva, J. L. Jardim, L. R. de Lima, and Z. S. Machado, "A Method for Ranking Critical Nodes in Power Networks Including Load Uncertainties." *IEEE Trans. Power Systems*, vol. 31, pp. 1341-1349, 2016.
[13] F. Tamp, P. Ciufo, "A Sensitivity Analysis Toolkit for the Simplification of MV Distribution Network Voltage Management." *IEEE Trans. on Smart Grid*, vol. 5, no. 2, pp. 559-568, 2014.
[14] H. N. Ng, M. M. A. Salama and A. Y. Chikhani, "Classification of capacitor allocation techniques," *IEEE Trans. on Power Delivery*, vol. 15, no. 1, pp. 387-392, Jan 2000.
[15] D.K. Khatod, V. Pant and J. Sharma, "Evolutionary programming based optimal placement of renewable distributed generators." *IEEE Trans. on Power Systems*, vol. 28, no. 2, pp. 683-695, 2013.
[16] R. Tonkoski, L. A. C. Lopes and T. H. M. El-Fouly, "Coordinated Active Power Curtailment of Grid Connected PV Inverters for Overvoltage Prevention," *IEEE Trans. on Sustainable Energy*, vol. 2, no. 2, pp. 139-147, April 2011.
[17] V. Calderaro, G. Conio, et al., "Optimal Decentralized Voltage Control for Distribution Systems with Inverter-Based Distributed Generators." *IEEE Trans. on Power Systems*, vol. 29, no. 1, pp. 230-241, 2014.
[18] A. Gonzalez, F. M. Echavarren, L. Rouco and T. Gomez, "A Sensitivities Computation Method for Reconfiguration of Radial Networks." *IEEE Trans. on Power Systems*, vol. 27, no. 3, pp. 1294-1301, 2012.
[19] R. S. Rao, et al, "Power Loss Minimization in Distribution System Using Network Reconfiguration in the Presence of Distributed Generation." *IEEE Trans. on Power Systems*, vol. 28, no. 1, pp. 317-325, 2013.
[20] R. Kasturi, P. Dorraju. "Sensitivity analysis of power systems." *IEEE Trans. on Power Apparatus and Systems*, vol. 10, pp. 1521-1529, 1969.
[21] W. R. Barcelo and W. W. Lemmon, "Standardized sensitivity coefficients for power system networks." *IEEE Trans. on Power Systems*, vol. 3, no. 4, pp. 1591-1599, 1988.
[22] K. Christakou, J.Y. LeBoudec, M. Paolone and D.C. Tomozei, "Efficient Computation of Sensitivity Coefficients of Node Voltages and Line Currents in Unbalanced Radial Electrical Distribution Networks." *IEEE Trans. on Smart Grid*, vol. 4, no. 2, pp. 741-750, 2013.
[23] R. Gurram, B. Subramanyam, "Sensitivity analysis of radial distribution network-adjoint network method," *International Journal of Electrical Power & Energy Systems*, vol. 21, no. 5, pp. 323–326, 1999.
[24] R. Preece and J. V. Milanović, "Assessing the Applicability of Uncertainty Importance Measures for Power System Studies," *IEEE Trans. on Power Systems*, vol. 31, no. 3, pp. 2076-2084, May 2016.
[25] K. N. Hasan, R. Preece, and J. V. Milanović, "Efficient Identification of Critical Parameters Affecting the Small-Disturbance Stability of Power Systems with Variable Uncertainty," *in IEEE PES General Meeting, Boston, MA, USA, 2016*.
[26] K. N. Hasan; R. Preece; J. V. Milanovic, "Priority Ranking of Critical Uncertainties Affecting Small-Disturbance Stability Using Sensitivity Analysis Techniques," *IEEE Trans. on Power Systems*, vol. PP, no.99, pp.1-1.
[27] B. Efron, and C. Stein, "The jackknife estimate of variance." *The Annals of Statistics*, pp. 586-596, 1981.
[28] A. Saltelli, et al. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications*, vol. 181, no.2, pp. 259-270, 2010.
[29] N. Wiener, "The Homogeneous Chaos," *American Journal of Mathematics*, vol. 60, no.4, pp. 897–936, 1938.
[30] D.B. Xiu, G.E. Karniadakis, "The Wiener--Askey polynomial chaos for stochastic differential equations," *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619-644, 2002.
[31] E. Keogh, A. Mueen. "Curse of dimensionality." In Encyclopedia of Machine Learning, Springer US, 2011, pp. 257-258.
[32] G. Blatman, "Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis," Doctoral dissertation, Clermont-Ferrand 2, 2009.
[33] F. Ni; P. Nguyen; J. F. G. Cobben, "Basis-Adaptive Sparse Polynomial Chaos Expansion for Probabilistic Power Flow." *IEEE Trans. on Power Systems*, vol. 32, no. 1, pp.694-704, 2017.
[34] S.S. Isukapalli, et al., "Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems." *Risk analysis*, vol. 18, no. 3, pp. 351-363, 1998.
[35] A.G. Expósito, A. Gomez-Exposito, A.J. Conejo and C. Canizares, eds., Electric energy systems: analysis and operation. CRC Press. 2016.
[36] Electric Power Research Institute, OpenDSS, Distribution System Simulator [Online], Available: http://sourceforge.net/projects/electricdss.
[37] S. Marelli, and B. Sudret, "UQLab: A framework for uncertainty quantification in Matlab," Vulnerability, Risk Analysis and Management, Proc. 2nd Int. Conf. on, Liverpool, 2014, 2554-2563.
[38] IEEE PES. (2016, Feb.) Distribution test feeders. [Online]. Available: http://www.ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html
[39] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," IEEE Trans. Power Syst., vol. 24, no. 1, pp. 40–49, Feb. 2009.
[40] A. Bhattacharya, Fundamentals of database indexing and searching. CRC Press, 2014.

Biographies

Fei Ni received her Ph.D. degree in natural science at the Institute for Automation of Complex Power Systems of E.ON Energy Research Center at RWTH Aachen University, Aachen, Germany, in 2015. In 2015, she joined the Electrical Energy Systems group at Eindhoven University of Technology, Eindhoven, the Netherlands, as a postdoc researcher. Her research interests include uncertainty modelling and simulation, sensitivity and reliability analysis, data mining and predictive control of electrical systems.

Michiel Nijhuis received his BSc degree in aerospace engineering in 2009 and in 2011 his MSc degree in sustainable energy technology both from the Delft University of Technology. Hereafter he joined the Dutch distribution network operator Alliander as a trainee and work here for two and a half years at various departments. In November 2013 he left Alliander to pursue a PhD degree with the Electrical Energy Systems Group at Eindhoven University of Technology. His research focuses on the distribution

network topology and the influence of the energy transition on the distribution network requirements.

Phuong H. Nguyen (M'06) received his Ph.D. at the Eindhoven University of Technology, the Netherlands, in 2010. He is an Assistant Professor in the Electrical Energy Systems group at the Eindhoven University of Technology, the Netherlands. He was a Visiting Researcher with the Real-Time Power and Intelligent Systems (RTPIS) Laboratory, Clemson University, USA, in 2012 and 2013. His research interests include distributed state estimation, control and operation of power systems and multi-agent systems, and their applications in the future power delivery system.

Joseph F.G. Cobben (M'11) received the Ph.D. degree in Electrical Engineering from Eindhoven University of Technology (TU/e), Netherlands, in 2007. He works as a research scientist in Alliander and specialized in safety of the LV- and MV-networks, power quality and safety of installations connected to the networks. He is member of several national and international standardization committees about requirements for low and high voltage installations and characteristics of the supply voltage. Dr. Cobben is a part-time professor at TU/e with research area of smart grids and power quality, and the author of several technical books.
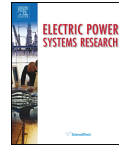
Valuation of measurement data for low voltage network expansion planning

M. Nijhuis [a],[*], M. Gibescu [a], J.F.G. Cobben [a],[b]

[a] *Electrical Energy Systems, Eindhoven University of Technology, Den Dolech 2, Eindhoven, The Netherlands*
[b] *Liander N.V., Utrechtseweg 86, Arnhem, The Netherlands*

ARTICLE INFO

ABSTRACT

The introduction of electric vehicles and photovoltaics is changing the residential electricity consumption. Distribution network operators (DNO) are investing in an advanced metering infrastructure (AMI) to enable cost reduction through smart grid applications. The DNO also benefits from the additional measurement data the AMI gives for the network planning process. The availability of AMI data can be limited by the cost of communication and by privacy concerns. To determine the social welfare of an AMI, the economic gains should be estimated. For the planning of the low voltage (LV) network, a method for determining the value of an AMI still needs to be developed. Therefore, a planning methodology which allows various levels of measurement data availability has been developed. By applying this approach the value of different levels of AMI from an LV-network planning perspective can be determined. To illustrated the application of this approach a case study for the LV-network of a Dutch DNO is performed. The results show that an increase in measurement data can lead to €49-254 lower LV-network reinforcement costs. A detailed analysis of the results shows that already 50% of the possible cost reduction can be achieved if only 65% of the households have AMI data available.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the electrification of heating and transportation loads and the introduction of distributed generation, the residential energy use is changing. The loading of the low voltage (LV) network can increase due to additional loads like heart pumps. Bi-directional power flows become possible through rooftop PV [1]. On the local level, the pace of and extent to which these developments take place can vary. Errors in the spatial load forecasting can have a significant effect on the required investments in the LV-network [2]. Currently, in most LV-networks, almost no measurement data is available, which generates uncertainty about the current loading of the LV-network [3]. This makes an adequate planning of the LV-network more complicated. The introduction of an advanced metering infrastructure (AMI) can, however, present the distribution network operator (DNO) with an unprecedented amount of data about the residential load.

The AMI data can be used for a multitude of applications [4], from demand response to automating the meter reading process

and more accurate LV-network expansion planning [5]. AMI data offers many opportunities for real time operation of the distribution network [6–8], however for the planning of the LV-network the available opportunities are not well documented. AMI data can be used to improve the load modelling [9] and load models based on AMI data can be implemented in the network planning process [10]. The question of how much value these measurements give to the DNO remains open. Privacy concerns about the use of AMI data require anonymisation. Depending on the implemented AMI communication infrastructure a limited read-out frequency and/or measurement frequency of the AMI data may be required [11]. This limits the DNO in the amount of information which can be extracted from the AMI measurements or requires additional communication infrastructure investments [12].

To be able to make adequate choices about the extent of the implementation of an AMI, the benefits of the AMI data should be assessed. The valuation of the AMI data from a demand response perspective has already been performed [13]. The assessment of the automation of a business process like the surveying of conventional electricity meters can easily be assessed from the expenses of the DNO. The use of AMI data to accurately assess transformer loading has also been studied [14], as well as the creation of load forecasts based on AMI data [15]. The usefulness of the additional measurements to increase observability from an operational perspective,

---

* Corresponding author.
E-mail addresses: m.nijhuis@tue.nl (M. Nijhuis), m.gibescu@tue.nl (M. Gibescu), j.f.g.cobben@tue.nl (J.F.G. Cobben).

e.g. in state estimation, is a common research topic [16], however, the data requirements from a long-term planning perspective are different. Preliminary research in using AMI data for the long-term network planning problem has been conducted [17]. The valuation of AMI data for the planning and reinforcement of LV-networks has not been covered. It is, therefore, hard to give an adequate estimation of the value of AMI data from an LV-network expansion planning perspective. This estimate is needed when determining the required characteristics of an AMI. From the point of view of generation adequacy studies into the value of additional measurement data have already been performed [18,19], however for the LV-network expansion planning the effects of the additional measurement data still need to be quantified. A uniform network planning methodology which incorporates different types of LV-measurement data is therefore needed to assess the value of AMI measurements for the LV-network planning process. Like standard network planning approaches, the methodology to provide a valuation for the AMI data from a network planning perspective should compute the LV-network cost, however, unlike the currently applied approaches, the methodology for the valuation should be able to handle load data with different levels of uncertainty in a similar way.

This paper proposes an LV-network expansion planning approach capable of handling these different levels of load uncertainty. How this approach subsequently can be applied to determine the value of AMI data from a network planning perspective is also shown in this paper. In Section 2 the effect of measurement data on the planning of the LV-network is discussed. In Section 3, the different levels of LV-measurement data availability are discussed, in combination with the methodology for incorporating this data in the LV-network expansion planning process. In Section 4, the application of this methodology is presented through a case study for the largest DNO in the Netherlands. Conclusions on the value of AMI data from a network expansion planning perspective are shown in Section 5.

2. Use of measurement data in LV-network expansion planning

The use of additional measurement data in LV-network planning reduces the uncertainty about the current loading of the LV-network. This reduction in uncertainty can lead to better investment decisions and thus to lower LV-network cost. If more measurement data is available the number of possible loading states of the LV-network reduces. This results in a tighter probability distribution of the load, which reduces the uncertainty and allows for more efficient investment decisions. Assume the loading of the network is normally distributed with mean μ and standard deviation σ. The network can be planned for the 98th percentile value of the peak load, or roughly the value of $\mu + 2\sigma$. This is illustrated in Fig. 1a where a normal distribution of the loading and the ideal planning level are shown. In this illustration, the network loading can be seen as the level of the peak load in the network. Where the distribution of this loading comes from the uncertainty about the actual level of peak load as there are no measurements available. A reduction of uncertainty would result in a lower variance, so $\sigma_{new} < \sigma_{old}$. The mean of the new probability density function can also change. Through the reduction of uncertainty, the mean of the probability density function moves closer to the actual value of the loading. Two situations are likely to occur $\mu_{new} + 2\sigma_{new} < \mu_{old} + 2\sigma_{old}$ or $\mu_{new} + 2\sigma_{new} > \mu_{old} + 2\sigma_{old}$. Both situations are shown in Fig. 1b and c, respectively.

$$\mu_{new} + 2\sigma_{new} < \mu_{old} + 2\sigma_{old}$$

a) Probability of network loading without measurements

b) Probability of network loading with measurements and equal mean

c) Probability of network loading with measurements and higher average loading
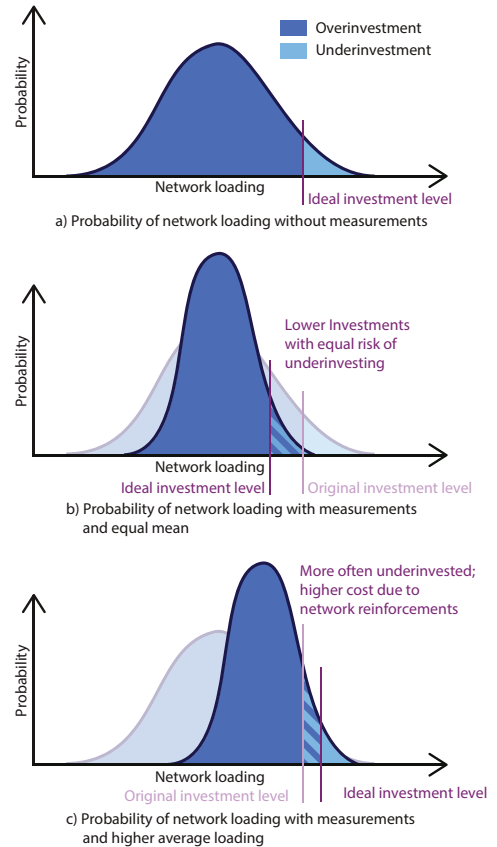
Fig. 1. Value of measurement data in the LV-planning process.

If the network is still planned with the same level of risk, the 98th percentile value of the peak load, the network investments will decrease. The value of the 98th percentile of the load is lower thus a less strong and cheaper network can be constructed with the same expected chance of overloading. This difference is the cost saving for the DNO from a network planning perspective.

$$\mu_{new} + 2\sigma_{new} > \mu_{old} + 2\sigma_{old}$$

If the network is still planned with the same 98th percentile value of the load, the DNO will construct a stronger and more expensive network. However, as the old network was planned with a percentile value for the risk which turned out to be lower than the 98th percentile, more cases in which the network has to be reinforced at a later stage will occur (more than the planned 2% of the cases). In the long run, this reduction in the need for future reinforcements of the network generates a lower total investment cost for the DNO, as unforeseen reinforcements can assumed to be suboptimal.

3. Methodology

To assess the value of data in the LV-network from a planning perspective and to consequently calculate the break-even point at which investing in additional measurements becomes economically infeasible, an LV-network expansion planning approach needs to be defined. Based on this planning approach the expected cost with different levels of data availability can be calculated. As the current approach to network planning is based on a simple peak load calculation, this approach first needs to be adjusted to allow for the inclusion of additional measurement data. Hereafter, a generic LV-network expansion planning approach capable of handling different levels of measurement data availability is defined. The implantation of different levels of measurement data availability is discussed hereafter. Finally, the metrics for the calculation of the added value of the additional measurements are determined.

The LV-network planning process is traditionally based on the assessment of the network over a time horizon of 40 years. From scenarios about future changes in the residential load, a single load growth factor is distilled. This yearly load growth factor is combined with an adjusted peak load per household. By taking into account the coincidence factor of the loads, the peak load in the whole network can be determined. The network is designed for this future peak load, by estimating the voltage levels and component loadings during the assumed peak load conditions [20]. In most LV-networks, only a single measurement at the MV/LV transformer is available. In addition, only the maximum loading of the transformer is recorded. This measurement data often needs to be collected manually. Therefore most of the MV/LV transformers only have one measurement point per year. For most residential customers only data on the yearly energy consumption exists, as it is needed for billing purposes.

3.1. LV-network planning with AMI data

In the current network planning approach, the emphasis is on creating a network that is strong enough to handle most load scenarios for the coming 40 years. This generally creates a more robust network than strictly necessary. As data about the current loading of the LV-network is usually unavailable, reinforcing the LV-network incrementally depending on the load growth is not the most cost effective option. With the introduction of an AMI or additional measurements in the LV-network, a more incremental approach for LV-network planning becomes feasible. The state of the LV-network can be better monitored, leading to a better planning and evaluation of possible network reinforcements. In this section, an approach is defined with respect to the planning of the LV-network for different levels of measurement data availability. An overview of the proposed planning approach can be seen in Fig. 2.

The step I in the adjusted planning process the future scenarios of the household load for the coming 40 years are generated based on the scenarios discussed in Appendix A. In step II these scenarios are used in a bottom-up household load modelling approach [21] to develop a set of load curves for the household load. These individual profiles need to be aggregated in order to be utilised in the LV-network planning process. Aggregated load profiles can be created from the individual load curves by clustering them to a limited number of load profiles per scenario and per time step [22,23], this is done in step III. Through a Monte Carlo approach random combinations of load curves are assigned to the households. In step IV the load curves are assigned to the households in a random manner. This assigning of the load curves is repeated multiple times to be able to assess different loading situations. In step V a single combination of load curves is transformed into different load levels based on the measurement data availability as discussed in Section 3.2.
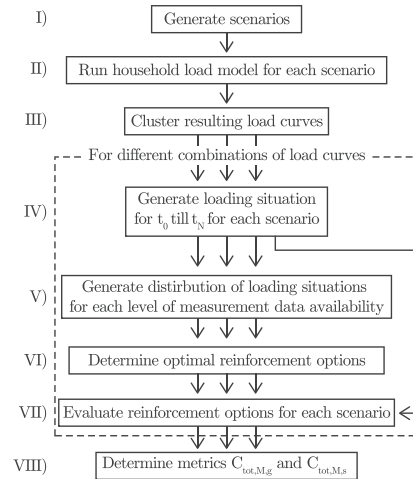
Fig. 2. Adjusted planning process.

Based on these loading situations the network planning optimisation is performed to determine the best reinforcement options in step VI. In step VII these reinforcement options are evaluated based on the original load curves as determined in step IV. By evaluating them with the original load curves a difference in cost is found with respect to the optimal reinforcements as determined in step VI. This difference in cost is the basis for the metrics with respect to the value of additional measurement data determined in step VIII.

The network loading is evaluated once every half year (to account for seasonal variation). Depending on the observability of the network, the loading can be determined to a certain accuracy level at each evaluation interval. Based on this information a new decision can be made on whether or not to invest in network reinforcements. If reinforcements are required, but not anticipated, the DNO has to make an unplanned network reinforcement, resulting in a higher cost. This cost is implemented in the method by using a penalty for any unanticipated reinforcements.

3.2. Measurement data integration in LV-network planning

Depending on the LV-network measurement approach, data with different levels of detail may be available. Based on the most logical locations of measurement devices, a number of different levels of data availability can be distinguished. These different levels are depicted in Fig. 3.

Fig. 3 shows the possible locations of the measurement devices in the LV-network. The purple indicates the situation now with just maximum current measurements at the MV/LV transformer. The blue indicates the next step in implementing measurements, with the maximum current also measured at each of the outgoing LV-feeders. In addition to just measuring the maximum current at the LV-feeder, time-series measurements at the beginning of the feeder can also be employed, this is indicated by the yellow in the figure. The DNO can also choose to collect AMI data every two months for planning purposes. This option is evaluated at two different participation rates for consumers. The red in the figure indicates the situation if 50% of the consumers are willing to let the DNO use their energy usage data for improved network planning, and the green indicates the situation if 100% of the consumers give their energy usage data to the DNO. The way these different levels of availability

Fig. 3. Different levels of data availability in the low voltage network.

Fig. 4. Distribution of the minimum voltage for a LV-feeder within a range of ±0.1% variation in feeder peak loading.

affect the planning of the LV-network is discussed for each of the five options below.

3.2.1. MV/LV transformer maximum

If only the maximum MV/LV transformer load is known, the loading of a feeder needs to be estimated. To be able to assess the loading of a single LV-feeder, a two-step approach is taken. First of all the variation of the maximum feeder load with respect to the maximum MV/LV transformer loading is estimated. Based on measurements of the maximum feeder current and the maximum MV/LV transformer current, the possible range of LV-feeder currents can be estimated. This is done by using the following formula:

$$P_{fdr} = (1 \pm 0.15) \frac{P_{trf} \cdot N_{HH,fdr}}{N_{HH,trf}} \tag{1}$$

with P_{fdr} the feeder loading, P_{trf} the transformer loading, $N_{HH,fdr}$ the number of customers connected to the f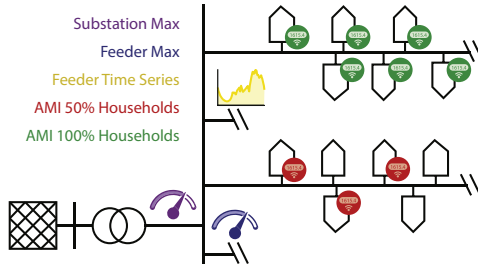eeder and $N_{HH,trf}$ the number of customers connected to the entire MV/LV substation. For a measured maximum MV/LV transformer power there is a 30% range (±0.15) for the maximum feeder loading. As there are no detailed measurements available the actual state of the feeder is considered to be unknown. For each of the maximum feeder loadings within this 30% range, a number of loading situations exist. These loading situations are assessed in the same way as the situation with only a maximum LV-feeder measurement available. This assessment is discussed in the next section. Only the feeder loading is estimated through (1), instead of measured.

3.2.2. LV-feeder maximum

Measurements of the peak current at each of the outgoing feeders can be performed. The maximum current at the beginning of each of the feeders is known in this case. By measuring the current at each feeder, a distinction can be made between feeders with a high and low load. As only the maximum loading of the feeder is recorded, variation in the load profile of the feeder is still possible. The households contributing to the peak load can be, for instance, all in the beginning of the feeder, uniformly distributed over the feeder or at the end of the feeder. The actual loading situation is taken from an assumption on the distribution of possible loading situations. This is illustrated in Fig. 4.

Fig. 4 shows the possible minimum voltage which occurs in the feeder for different combinations of household loads. The maximum current at the beginning of the feeder only varies by ±0.1%, while the minimum voltage lies in the range of 0.935–0.955 (p.u.). To construct the figure, load data from 120 smart meters (15-min data for 3 months) from a single neighbourhood are randomly distributed over a feeder with 67 connected consumers. This distribution of the voltage is generated for a range of feeder loadings. The 98*th* percentile of the voltage distribution is subsequently chosen as acceptable risk level. The load current I_m at bus m is assumed to be
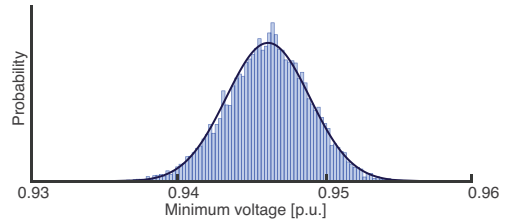
equal to the load current at the other buses $I_0 = I_1, \ldots, = I_m, \ldots, = I_{N_b}$, with N_b the number of buses. The load current I_m which results in a voltage deviation equal to the 98*th* percentile value is chosen as the current that is applied in the optimisation procedure as explained in Section 3.3.

3.2.3. LV-feeder profile

In addition to measuring the peak loading, the entire load profile at the beginning of the feeder can also be measured. It is assumed that the profile is measured with a 15-min time step and the data is sent every half year. The main difference from a practical point of view is not in the measurements of the LV-feeder but in the amount of data which needs to be transmitted from the MV/LV substation to the servers of the DNO. In this situation, the investments are assessed for each single set of load curves. This set of load curves $I_{L,0}, \ldots, I_{L,n}$ is used to generate a probability density function of the possible voltage deviations by changing which load curve is assigned to which household (e.g. $I_0 = I_{L,1}, I_1 = I_{L,0}$ or $I_0 = I_{L,0}, I_1 = I_{L,1}$). The assignment of load curves to the households that represent the 98*th* percentile value from the probability density function of the voltage deviations is taken and used in the optimisation procedure.

3.2.4. AMI 50% participation

If an AMI is introduced at the household level, it remains possible that consumers will opt-out of AMI data retrieval by the DNO. In this case, it is assumed that a randomly chosen 50% of the households chose to opt out of the data-sharing. This situation is assessed by giving 50% of the households the same load curve over multiple iterations, while for the other 50% of the households the load is chosen based on the 98*th* percentile value of the peak load distribution of the smart meter measurements. For the households without smart meter measurements, the loads are assumed to be equal $I_0 = I_1, \ldots, = I_{\frac{n}{2}}$ while for the households which are measured, individual load curves are assigned $I_0 = I_{L,0}, \ldots I_{\frac{n}{2}} = I_{L,\frac{n}{2}}$. The investment options are subsequently assessed for different sets of households with the same load curve over multiple iterations.

3.2.5. AMI 100% participation

If all the households allow the DNO to retrieve their load profile data, the complete loading situation is observable for the DNO. This allows the DNO to accurately assess the required network investments. Though the future network loading is still uncertain, the DNO can determine the chance of voltage or current violations with a high accuracy. For this situation, each load current $I_0, \ldots, I_n$ is assigned an individual value from the available set of load curves $I_{L,0}, \ldots, I_{L,n}$ (e.g. $I_0 = I_{L,0}, \ldots, I_n = I_{L,n}$).

3.3. Planning implementation

The goal of the network planning process is to find the lowest expected cost at which the network can function adequately.

The network planning is often limited to find the lowest investment and operational cost (energy losses) under the long term load uncertainty [24]. The availability is often also included in the LV-network expansion planning, however the majority of the LV-networks within Europe are radial networks constructed with underground cables. This lead to low failure rates and little network reconfiguration options, which in turn generates only small differences in availability of the different network expansion options. For these reasons availability is not taken into account in this work. The approach applied in this paper consist of taking binary reinforcement decisions d from the set of possible options $\mathbb{D}$ at time t, evaluated over the set of discrete times $\mathbb{T}$, to obtain the lowest possible cost C_{tot}. To account for the time value of money the total cost $C_{tot}(t)$ are discounted by the deprecation rate r. The solution needs the voltages at all the nodes n, U_n to be within the limits U_{min} and U_{max} and the branch currents I_{br} in the network to be within the current limits (1 (p.u.)) at all times. This can be expressed as:

$$\min \sum_{t \in \mathbb{T}} \frac{C_{tot}(t)}{(1+r)^t}$$

$$\begin{aligned} \text{s.t.} \quad & d \in \{0, 1\} && \forall d \in \mathbb{D} \\ & I_{br,n}(t) \leq 1 \,(\text{p.u.}) && \forall n = 0, \ldots, N_b \\ & U_{\min} \leq U_n(t) \leq U_{\max} && \forall n = 0, \ldots, N_b \end{aligned} \quad (2)$$

In reality, it can happen that the voltage and maximum current constraints are violated. These constraints are therefore modelled as a penalty function instead of a hard constraint. The value of the penalty function represents the cost for the DNO when either a violation of the voltage or the maximum current occurs. The penalty function is assumed to be a constant C_{pen}, independent of the severity of the violation. The penalty cost C_{pen} consists of a part of direct damages, for instance, due to a violation of the maximum current limit, the fuse at the beginning of the LV-feeder will trip, resulting in an outage and the need for a replacement fuse. The other part of the penalty function is the additional cost associated with an urgent network reinforcement instead of a planned network reinforcement. Therefore the penalty cost C_{pen} is a DNO specific value.

The total cost C_{tot} consists of multiple parts: The fixed investment cost C_{fix}, the penalty cost C_{pen}, the cost to reinforce the branches of the feeders $d_{cbl} p_{cbl} \cdot l_{branch}$, the cost of adding new feeders $d_{cbl} p_{fdr} + p_{cbl} l_{0 \to mid}$, the cost of an additional MV/LV substation $d_{MV/LV} p_{MV/LV} + p_{cbl,MV} l_{0 \to end}$ and the cost associated with energy losses $p_e \sum_{m=0}^{N_b} L_m$.

$$C(t) = C_{fix} + C_{pen} + d_{cbl} p_{cbl} l_{br,n} + d_{fdr} p_{fdr} + d_{fdr} p_{cbl} l_{0 \to mid}$$
$$+ d_{MV/LV} p_{MV/LV} + d_{MV/LV} p_{cbl,MV} l_{0 \to end} + p_e \sum_{m=0}^{N_b} L_m \quad (3)$$

with p_{cbl}, p_{fdr}, $p_{MV/LV}$, $p_{cbl,MV}$ and p_e being the prices of one meter LV-cable, an additional feeder at the LV busbar, an additional MV/LV-substation, one meter of MV cable and one kWh of electricity respectively, $l_{br,n}$ the length of branch n, $l_{0 \to mid}$ and $l_{0 \to mid}$ the distance between the MV/LV-substation and the middle and end of the feeder, d_{cbl}, d_{fdr} and $d_{MV/LV}$ the vectors of binary decision variables corresponding to whether or not to reinforce a branch, add a new feeder or add an additional MV/LV substation respectively.

The branch currents I_{br}, bus voltages U_n and branch losses L_m are computed from a backwards forwards sweep load flow calculation:

$$U_i = U_{i,0} - \sum_{n=i}^{N_b} \left(Z_n I_{br,n} \right) \quad (4)$$

where U_i is the voltage at bus i, bus indexes i, n and m running from the busbar (index 0) to the end of the feeder (index N_b), $U_{i,0}$ the base voltage (400 V L-L), Z_n the impedance of the section between node n and $n - 1$, $I_{br,n}$ the branch current as calculated by:

$$I_{br,n} = \sum_{m=n}^{N_b} \left(\frac{U_{m,0} I_m + L_m}{U_m} \right) \quad (5)$$

with I_m the current of the load at bus m. The loss L_m between bus m and bus $m + 1$ is calculated by:

$$L_m = \alpha Z_n I_{br,n}^2 \quad (6)$$

with α a factor to account for the fact that the loss calculation is now only based on a single peak load moment instead of the entire load curve. The factor α is calculated at the beginning of the optimisation by calculating the loss by applying the entire load curve and dividing it by the loss calculated during peak load times.

The voltage limits that are used during the assessment of the LV-feeders are chosen as ± 0.05 (p.u.). This more stringent limit is applied instead of the generally applied limits of ± 0.1 (p.u.), as the assessment is based only on the LV-feeders. The voltage fluctuations in the MV-network are not taken into account. The residential load currents I_m are constructed based on the amount of available measurement data as discussed in the previous section. In order to have an idea of the time evolution of the load currents, the load current needs to be constructed for all the scenarios s in the scenario space $\mathbb{S}$.

For the assessment of the possible network reinforcement alternatives, genetic algorithms [25] and particle swarm optimisation [26] are often used. Computation wise there is little difference in the performance of these methods [27,28]. A genetic algorithm is applied in this paper to determine the best reinforcement options. It is assumed that the best reinforcement options are equal for loading situations where $\sum_{n=1}^{N_B} Z_{poc,n} \times I_n$ differs less than 1% of the maximum value, with N_B the number of buses, $Z_{poc,n}$ the impedance at the points of connection and I_n the current injected at bus n. This allows for a lower number of computations while determining the best reinforcements option for each scenario and for each level of measurement data availability.

3.4. Metrics to valuate network planning data

To valuate the available measurement data for the network planning process the difference in total cost with different levels of data availability is calculated by (2). Not all networks might have the same need for measurement data from a network planning perspective. Networks for which no problems are expected in the coming years, even under extreme load growth scenarios, obviously do not benefit from having measurement data available. The question remains for which type of networks the additional measurement data holds more value. As the LV-network consists of many different types of feeders, whether or not the value of additional measurement data differs for these feeders should be calculated. This value can be calculated by adjusting the minimisation of the cost as given in (2) with an objective function which assumes an equal probability of each scenario s in the scenario space $\mathbb{S}$.

$$C_{tot,M,g} = \frac{1}{|\mathbb{S}|} \min_{d \in \mathbb{D}} \sum_{s \in \mathbb{S}} C_{tot,M,s,g} \quad (7)$$

where $C_{tot,M,s,g}$ is the total cost of LV-feeder g under scenario s and with the level of measurement data availability M. By comparing these total cost $C_{tot,M,g}$ for a specific feeder g, but different levels of data availability M, the value of the measurement data for a specific feeder g can be determined. The optimisation is done based on the evaluation of the cost under all scenarios, as it is assumed that the scenarios are indistinguishable at the time that the investments are made. In this case, the scenarios are assumed to have an equal probability. If more information on the individual probabilities of specific scenarios is present the sum in Eq. (7) can be changed into a weighted sum.

The expected load growth also should have a significant effect on the value of the additional measurement data, therefore a metric to quantify the effects of using different load growth scenarios is presented. If the loading is expected to remain more or less constant in the coming years, additional measurement data will not add any value, while scenarios with a fluctuating load growth should give more value to the measurement data. To assess the effects of the scenarios on the value of measurement data, the objective function as given in (2) is changed to generated the expected cost for each scenario s, while taking into account all the LV-feeders:

$$C_{tot,M,s} = \frac{1}{|\mathbb{G}|} \sum_{g \in \mathbb{G}} \min_{d \in \mathbb{D}} C_{tot,M,s,g} \qquad (8)$$

where $\mathbb{G}$ is the set of generic LV-feeders g. The objective function (8) is very similar to the objective function in (7), where the objective in (7) is averaged based on the scenarios s, (8) is averaged based on the feeders g. The main difference is that in (7) the minimisation is performed based on the performance under all the scenarios, while in (8) the minimisation is performed for each scenario individually.

Another aspect which can be determined by the use of this methodology is the incremental value of an increase in the penetration of AMI. If more and more consumers connected to the same feeder allow for the use of their AMI data for network planning purposes the expected total investment cost will decrease. Also with an increasing AMI usage, there will be a point when the additional measurements at the feeder will no longer generate a lower expected cost. It is interesting to determine at which point it is no longer useful for the DNO to increase the measurement set from a network planning perspective. This is calculated by determining the expected cost at different levels of AMI penetration in the same manner as done for the 50% AMI penetration.

4. Case study

To illustrate the application of the proposed approach for the valuation of measurement data for the planning of an LV-network, a case study on the network of Liander (Dutch DNO) is performed. For this case study, first the network which is considered is discussed. This is followed by an elaboration on the creation of the household load for the coming years. Next to this, the reinforcement options and costs are given. Based on this information the analysis of the value of measurement data from a network planning perspective can be performed. The results are given in Section 4.1.

To gain an idea of the value of measurement data for the DNO, a representative subset of the network of the DNO needs to be selected for the case study as the complete network is too large for the analysis (>80,000 LV-feeders). The selection of the representative LV-feeders is done by the application of a clustering approach to the entire LV-network of Liander. By applying a clustering approach to the entire network generic feeders can be extracted from the network. These generic feeders are representative for the whole network. A fuzzy k-medians clustering approach is used to generate a set of generic LV-feeders based on the network data. The following network parameters are used in the clustering approach:

Table 1
Cost of the components used in the case study.

| Symbol	Component	Cost (€)
p_{cbl}	LV-cable (\km)	45,000
$p_{cbl,MV}$	MV-cable (\km)	70,000
p_{msr}	Additional MV/LV substation	25,000
p_{fdr}	LV-busbar expansion for an additional feeder	7000
p_e	Electricity price (\MWh)	40
C_{pen}	Penalty cost	8500
C_{fix}	Fixed investment cost	1300

impedances, cable length, number of branches, branch depth and the number and type of connected customers. Next to these network parameters, the graph theory concepts of degree distribution, sequence and the centrality of the power, impedance and length are also used for the clustering. By using these parameters feeders can be rebuilt from the cluster centres which would represent the structure and loading of the original feeder. For a more detailed description of the applied clustering approach [29] can be consulted. By applying this clustering approach a subset of 96 feeders is extracted from the entire LV-network. These feeders are subsequently analysed with a peak load of 125% of the current peak load, to determine which feeders would require reinforcements within the next 5 years. This creates a set of 26 generic feeders that are used to valuate the measurement data from an LV-network planning perspective. The other feeders are not taken into account as reinforcements are not required independent of the amount of available measurement data.

For the creation of the scenarios which are used to assess the adequacy of the LV-feeders, a scenario space has been constructed based on all the combinations of future scenarios for electric vehicles (EV), photovoltaics (PV), heat pumps, economic growth and appliance efficiency. Appendix A gives an overview of these scenarios. The household load is generated by using a bottom-up Monte Carlo Markov Chain approach [21]. In this method, the behaviour of each household member is modelled using Markov Chains, based on data from time use surveys. The appliances are modelled individually and switch on or off based on the modelled behaviour of the users. This generates a load curve for each individual appliance within a household. The appliances are subsequently aggregated to generate the load curve of the household. For each of the individual households within a network, this approach generates a unique load curve. In order to create the load curve of a household for the coming years, scenarios on how the appliances within the household will develop, need to be implemented. These scenarios are implemented by updating the appliances in the households based on the penetration rate of new appliances (EV, PV, etc.) for a certain scenario. In the same way, trends with respect to the changes in appliance ratings (e.g. efficiency and size of appliances) are also taken into account. With this model, the load curve for 2000 individual households for one day each half year for the coming 40 years has been created for the scenarios as described in Appendix A. To make the analysis of possible loading situations computationally feasible the household load curves have been clustered by using a k-means clustering approach 10 representative load curves for each half year of each scenario are generated [23]. More advanced approaches exist to reduce the dimensionality of load profiles [30,31], however the accuracy of k-means approach was found to give an acceptably small error (a Wasserstein distance of 0.984 was found when comparing the voltages calculated with the original load profiles to the voltages created with clustered load profiles).

The reinforcement options are assessed based on their expected value. For the calculation of the expected value, the data in Table 1 is used. The depreciation rate r is chosen as 3% per year and for simplicity, a single cable type is used for replacements: a 4×150AL

Table 2

Expected discounted cost for the coming forty years for LV-network reinforcements with different levels of data availability per LV-feeder.

	Cost (€)	Percentage
100% AMI	520	67.2%
50% AMI	680	87.8%
Feeder time series	674	87.1%
Feeder maximum	725	93.7%
Substation maximum	774	100%
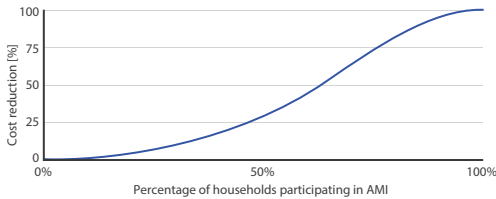
Fig. 5. Mean expected cost with decreasing AMI participation.

LV-cable. With this data, the expected cost of the reinforcements can be computed for the different levels of measurement data availability.

4.1. Results

The approach described in the previous section has been applied to valuate the measurement data for the LV-network of Liander. The results are shown in this section. First, the cost of the different levels of data availability is given. Secondly, the results are analysed on the different LV-feeders characteristics, followed by an analysis based on the chosen scenarios.

4.1.1. Valuation of measurement data

The cost of the network per LV-feeder are calculated for different levels of data availability and the results are shown in Table 2.

As expected, the table shows an overall reduction in expected cost as the availability of the measurement data increases. The use of feeder data instead of substation measurements can reduce the expected cost for feeder reinforcement by about 10%, compared to the current situation. This translates to a cost saving of €49 over 40 years per LV-feeder. For these savings to be achieved, additional measurement devices will have to be installed, maintained and the measurement data has to be collected. Most likely this results in an uneconomical business case if only the benefits from the network expansion planning perspective are taken into account. The possibilities of having time series feeder measurements and acquiring AMI data from 50% of the households generate a more or less equal expected cost that is slightly lower than the expected cost using a measured feeder maximum. Acquiring all the data from the AMI infrastructure can lead to an additional 20% lower expected cost. The majority of the analysed feeders hardly need any reinforcements, even in the most extreme scenario. The added value of measurement data for these feeders is zero from a network planning perspective.

The expected cost between the 100%, 50% and 0% AMI data availability cases differ significantly. To further illustrate this the expected costs for intermediate levels of AMI availability have been computed.

In Fig. 5 the expected cost for increasing levels of AMI data availability is shown. It can be seen that the expected cost as a function of the level of AMI data availability follows an S-curve. The inclusion or exclusion of just a couple of households from the AMI data does

Table 3

Characteristics of the feeders for which the results are shown in Fig. 6.

	A	B	C	D	E
Length (m)	900	450	500	890	820
Connections	9	36	45	15	95
$\bar{Z}_{PoC}$ (mΩ)	133	110	105	162	140
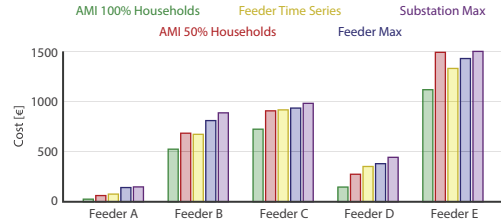
Fig. 6. Expected cost network reinforcements for five feeders with different levels of data availability.

not have a large effect on the expected cost. If the level of available AMI data can be increased it pays off most if a level of data availability of at least 40% is already present. Increasing the level of data availability past the 90% no longer generates any significant additional returns. For a level of AMI data availability of 65% of all the households, a 50% cost reduction is achieved.

4.1.2. Feeder characteristics

Next to the results for the combined LV-feeders, also the results for an individual feeder can be given. From all the feeders used in the calculations, five feeders have been selected. The characteristics of these feeders are shown in Table 3.

The table shows the length, number of connections and average impedance at the point of connection (PoC) for each of the feeders. Feeders A and D are rural feeders with a large length and a small number of customers. Feeders B and C are suburban feeders with an intermediate length and amount of customers and feeder E is an older urban feeder with a long length, high number of customers and high average impedance at the PoC. For these five feeders, the results for the calculation of the network cost with different levels of data availability are plotted in Fig. 6.

The trend of a decrease in expected cost with an increase in available measurement data is visible in the figure of the individual feeders as well. There are however large differences in the cost savings which can be achieved for the individual feeders. For feeder A the availability of all the AMI data for the LV-network planning process can lead to a cost saving of about 80%, while for feeder E the cost saving is only 20%. This difference can be explained by the fact that feeder A is, for a large percentage of the analysed scenarios, strong enough to not require any reinforcements, while Feeder E needs reinforcements reasonably early. The use of additional measurement data can in the case of feeder A avoid any unnecessary reinforcements. For feeder E the large differences between the loading of the individual households occur early on, as the penetration rates are small in the beginning. If one knows the sum of all the loads on the feeder, the small number of households which use a certain technology is taken into account, while if a random 50% of the household have a known loading, the small number of households might be excluded. This leads to relatively high expected cost for feeder E. A similar picture to feeder A is present at feeder D. For feeder E the use of AMI data for half of the households does not assist much in the grid planning process, while for the other feeders it gives an additional benefit. For the suburban feeders B and C, the reduction in expected cost is between the reduction of

Table 4
Characteristics of the scenarios for which the results are shown in Fig. 7.

	Low	Low PV, EV	Medium	High PV, EV	High
PV	Low	Low	Medium	High	High
EV	Low	Low	Medium	High	High
Heat pump	Low	Medium	Medium	Medium	High
GDP	Low	Medium	Medium	Medium	High
Efficiency	High	Medium	Medium	Medium	Low
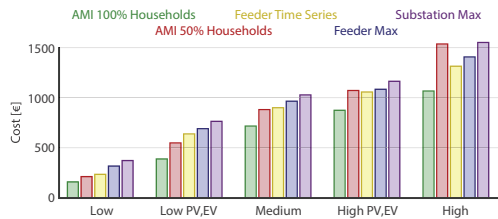
Fig. 7. Expected cost of network reinforcements for the five scenarios from Table 4 with different levels of data availability.

feeders A and E. The additional measurement data has more influence on feeder B in comparison to feeder C. This can be explained by the slightly higher number of customers connected to feeder C. The error induced by the use of average values is, therefore, lower and the benefits of the additional measurement data as well. Overall feeders with a lower need for reinforcements can have more benefits from AMI data in comparison to the feeder reinforcement cost, however the absolute benefits are higher for the feeders that require more reinforcements.

4.1.3. Scenario influence

Next to the feeder characteristics, the uncertainty of the future loading of the network is also expected to have a large influence when it comes to the value of the measurement data. To gain more insight in the effect of the scenarios on the value of measurement data, the results have also been analysed for different scenarios. From the scenarios which have been used in the computation, the subset depicted in Table 4 is used for a more in-depth analysis. The five scenarios which have been chosen consist of the two most extreme scenarios, an intermediate scenario and two scenarios in which just the level of EV and PV penetration is altered from the intermediate scenario. The calculated expected cost for these five scenarios and the five cases of data availability is shown in Fig. 7.

The expected cost for the different scenarios confirms the idea that the relative value of the measurement data is larger as the chance of having to reinforce is smaller. This can be seen as the relative difference between the 100% AMI case and the substation maximum case decreases as the amount of load growth in the scenarios increases. As the amount of load growth increases, the benefits of the usage of only half the AMI data decreases, as it is expected that the difference in loading between households becomes larger. For the scenario with the least load growth the expected costs are close to the case with 100% AMI data available, while for the highest load growth, the 50% AMI case generates almost no profit compared to the substation maximum case. The value of having measurement data available on a feeder level instead of just at the substation level is significantly higher for the extreme scenarios in comparison to the intermediate scenarios. For the extreme scenarios, the difference between the penetration rates at different time steps for different feeders can be larger, giving rise to an increased value of knowing the loading of the feeder.

5. Conclusion

An approach for the valuation of the use of measurement data for the LV-network planning process is presented. The approach is based on using a single LV-network planning approach that allows for the use of different levels of measurement data, through altering the applied load values. This approach allows for the determination of the value of measurements in the LV-network, from a network planning perspective. With this approach, the cost-benefit analysis for installing an AMI can be improved. Based on the proposed method the value of the possible deferral of the LV-network reinforcements can also be included. The approach has been applied to the LV-network of a Dutch DNO. In general, an increase in available measurement data leads to lower reinforcement cost as the network-wide results show. The cost reduction that can be achieved in practice does, however, differ significantly based on the scenario and the type of feeder which is used in the analysis. Lower load growth scenarios and more robust feeders, which still require some level of finely-tuned reinforcements are the main beneficiaries of the uncertainty reduction through additional measurement data. The analysis also shows that on average a level of 65% of AMI data availability can already achieve a cost reduction of 50%.

This work allows the DNO to decide whether or not to invest in additional measurement devices in the LV-network, by not only including the operational benefits for an AMI but also the benefits from the deferral of LV-network reinforcements. As the use of additional measurement data is required for the implementation of many smart grid solutions such as demand side management (DSM), the approach can be extended to include DSM alternatives in the planning process. As future work, a more accurate assessment of the benefits of applying these smart grid technologies can be performed by their inclusion in the LV-network planning methodology.

Appendix A. Load scenarios

For the evaluation of the network expansion planning, the residential load for the coming decades needs to be determined, as LV-cables tend to have a lifetime easily eclipsing the 30 years. The main technologies which are expected to increase in penetration rate and have a large influence on the LV-network are PV, EV and heat pumps. The penetration rate is defined as the number of households that use a certain technology. Next to the increasing penetration of these technologies, the changes in electricity usage should also be incorporated in the scenarios. Scenarios on the economic growth and the change in appliance efficiency are also used. The scenarios are modelled through the use of the following equation:

$$pt(yr) = \frac{pt_{\max}}{1 + e^{-a(yr)}} \qquad (9)$$

with pt the penetration level at year yr, pt_{max} the maximum penetration level and a the penetration rate parameter. This approach is applied for the all the scenario drivers, expect for the economic

Table A.5
Overview of the different parts of a scenario and how they are implemented.

	High		Medium		Low	
	pt_{max}	a	pt_{max}	a	pt_{max}	a
PV	82	0.16	32	0.18	7.2	0.19
EV	63	0.09	28	0.11	4.5	0.11
Heat pump	46	0.06	5.2	0.06	0.1	0.05
Efficiency	0.52	0.13	0.59	0.12	0.96	0.17
Growth	1.25%		0.5%		0%	

growth where an exponential growth function is used. In Table A.5 an overview of the employed scenario parameters is given.

For the EV and PV the following assumptions have been made with respect to the modelling of these technologies. The rated power of the PV systems are assumed to uniformly distributed between 2 and 5.5 kW per household. For the EV charging, slow charging at a rate between 3 and 8 kW is assumed. The distribution of departure and arrival times of the EVs as well as the distance driven are modelled based on data from mobility studies according to the method described in [32]. The EVs start charging as soon as they arrive back at the household.

References

[1] M. Nijhuis, M. Gibescu, J. Cobben, Assessment of the impacts of the renewable energy and ICT driven energy transition on distribution networks, Renew. Sustain. Energy Rev. 52 (August) (2015) 1003–1014, Available: http://linkinghub.elsevier.com/retrieve/pii/S1364032115007716.

[2] D.A.G. Vieira, M.M.A. Cabral, T.V. Menezes, B.E. Silva, A.C. Lisboa, Measuring the spatial error in load forecasting for electrical distribution planning as a problem of transporting the surplus to the in-deficit locations, Proceedings – 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012, vol. 2 (December 2012) 475–480.

[3] M. Neaimeh, R. Wardle, A.M. Jenkins, J. Yi, G. Hill, P.F. Lyons, Y. Hübner, P.T. Blythe, P.C. Taylor, A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts, Appl. Energy 157 (January) (2015) 688–698.

[4] E. Liu, M.L. Chan, C.W. Huang, N.C. Wang, C.N. Lu, Electricity grid operation and planning related benefits of advanced metering infra-structure, 2010 5th International Conference on Critical Infrastructure, CRIS 2010 – Proceedings (September 2010) 1–5.

[5] J. Foosns, E. Tnne, T. Pynten, Power system planning in distribution networks today and in the future with smart grids, in: 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013), No. 1426, June 2013, http://dx.doi.org/10.1049/cp.2013.1233.

[6] D. Bernardon, G. Iop, M. Gundel, J. Fernandes, A real-time operation analysis system for distribution networks, Electr. Power Syst. Res. 78 (March (3)) (2008) 346–352, Available: http://linkinghub.elsevier.com/retrieve/pii/S037877960700048X.

[7] A. Alam, J. Zhu, V. Frey, R. Zhao, T. Hoang, S. Larson, S. Mundade, S. Ruelas, A. Herasimava, Improving operational planning study models with historical generation and load data, in: 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), vol. 2016, July, IEEE, May 2016, pp. 1–5, Available: http://ieeexplore.ieee.org/document/7520038/.

[8] A.A. Mohamed, O.A. Mohammed, Real-time load emulator for implementation of smart meter data for operational planning, in: 2012 IEEE Power and Energy Society General Meeting, IEEE, July 2012, pp. 1–6, Available: http://ieeexplore.ieee.org/document/6345619/.

[9] X. Zhang, S. Grijalva, M.J. Reno, A time-variant load model based on smart meter data mining, in: 2014 IEEE PES General Meeting – Conference & Exposition, vol. 2014, October, IEEE, July 2014, pp. 1–5, Available: http://ieeexplore.ieee.org/document/6939365/.

[10] R. Brakken, T. Ait-Laoussine, S. Steffel, D. Sikes, Integrating customer load and GIS data for improved distribution planning & operations (updated), in: 2005/2006 PES TD, IEEE, 2006, pp. 1056–1063, Available: http://ieeexplore.ieee.org/document/1668648/.

[11] W. Luan, D. Sharp, S. Lancashire, Smart grid communication network capacity planning for power utilities, Transmission and Distribution . . . (April 2010) 1–4, Available: 10.1109/TDC.2010.5484223%5Chttp://ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=5484223.

[12] W. Luan, D. Sharp, S. LaRoy, Data traffic analysis of utility smart metering network, in: 2013 IEEE Power & Energy Society General Meeting, IEEE, 2013, pp. 1–4, Available: http://ieeexplore.ieee.org/document/6672750/.

[13] P. Siano, Demand response and smart grids – a survey, Renew. Sustain. Energy Rev. 30 (November) (2014) 461–478.

[14] Y.L. Lo, S.C. Huang, C.N. Lu, Transformational benefits of AMI data in transformer load modeling and management, IEEE Trans. Power Deliv. 29 (April (2)) (2013) 742–750.

[15] B.P. Hayes, M. Prodanovic, State forecasting and operational planning for distribution network energy management systems, IEEE Trans. Smart Grid 7 (March (2)) (2016) 1002–1011.

[16] R. Singh, B.C. Pal, R.B. Vinter, Measurement placement in distribution system state estimation, IEEE Trans. Power Syst. 24 (May (2)) (2009) 668–675.

[17] G. Roupioz, X. Robe, F. Gorgette, First use of smart grid data in distribution network planning, 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013) (June 2013) 1–4, Available: http://www.ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=6683438.

[18] J. Pan, S. Rahman, Multiattribute utility analysis with imprecise information: an enhanced decision support technique for the evaluation of electric generation expansion strategies, Electr. Power Syst. Res. 46 (2) (1998) 101–109, Available: http://www.sciencedirect.com/science/article/pii/S0378779698000224.

[19] A.H. van der Weijde, B.F. Hobbs, The economics of planning electricity transmission to accommodate renewables: using two-stage optimisation to evaluate exibility and the cost of disregarding uncertainty, Energy Econ. 34 (January (6)) (2012) 2089–2101, http://dx.doi.org/10.1016/j.eneco.2012.02.015.

[20] H. Willis, Power Distribution Planning Reference Book, vol. 2, CRC Press, March 2004.

[21] M. Nijhuis, M. Gibescu, J. Cobben, Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data, Energy Build. 112 (January) (2016) 121–129, Available: http://linkinghub.elsevier.com/retrieve/pii/S0378778815304436.

[22] F.L. Quilumba, W.J. Lee, H. Huang, D.Y. Wang, R.L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities, IEEE Trans. Smart Grid 6 (March (2)) (2015) 911–918.

[23] M. Nijhuis, R. Bernards, M. Gibescu, J. Cobben, Stochastic household load modelling from a smart grid planning perspective, in: 2016 IEEE International Energy Conference (ENERGYCON), IEEE, April 2016, pp. 1–6, Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7513954.

[24] W.H. Kersting, Distribution System Modeling and Analysis, 2002.

[25] V. Miranda, J.V. Ranito, L.M. Proena, Genetic algorithms in optimal multistage distribution network planning, IEEE Trans. Power Syst. 9 (January (4)) (1994) 1927–1933, Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=331452.

[26] I. Ziari, G. Ledwich, A. Ghosh, Optimal integrated planning of MVLV distribution systems using DPSO, Electr. Power Syst. Res. 81 (October (10)) (2011) 1905–1914, Available: http://linkinghub.elsevier.com/retrieve/pii/S0378779611001271.

[27] C. Gamarra, J.M. Guerrero, Computational optimization techniques applied to microgrids planning: a review, Renew. Sustain. Energy Rev. 48 (2015) 413–424.

[28] R. Shi, C. Cui, K. Su, Z. Zain, Comparison study of two meta-heuristic algorithms with their applications to distributed generation planning, Energy Proc. 12 (2011) 245–252.

[29] M. Nijhuis, M. Gibescu, J. Cobben, Clustering of low voltage feeders from a network planning perspective, Cired 2015, Lyon, France (June 2015).

[30] R. Li, F. Li, N.D. Smith, Multi-resolution load profile clustering for smart metering data, IEEE Trans. Power Syst. 31 (November (6)) (2016) 4473–4482, Available: http://ieeexplore.ieee.org/document/7428953/.

[31] R. Al-Otaibi, N. Jin, T. Wilcox, P. Flach, Feature construction and calibration for clustering daily load curves from smart-meter data, IEEE Trans. Ind. Inf. 12 (April (2)) (2016) 645–654, Available: http://ieeexplore.ieee.org/document/7404272/.

[32] R.A. Verzijlbergh, The Power of Electric Vehicles – Exploring the value of flexible electricity demand in a multi-actor context (Ph.D. dissertation), October 2013, Delft.

Modelling of Load Uncertainty for Multi-Stage LV-network Expansion Planning

Michiel Nijhuis, *Member, IEEE,* Pedro M.S. Carvalho, Madeleine Gibescu, *Member, IEEE,* and Sjef Cobben

Abstract—In LV-network expansion planning a high load uncertainty is present, due to the volatility of the household load. Especially for LV feeders with a small number of consumers, this uncertainty needs to be taken into account in the planning process. In this paper, an approach to model this uncertainty and implement it in the LV-network expansion planning is described. The approach consists of decomposing the load uncertainty into two components: the uncertainty caused by the volatility of the household load and the uncertainty associated with the long-term load evolution. The former can be modelled by a probability density function. To incorporate this in the network expansion planning methodology, a probabilistic load flow instead of a deterministic one needs to be applied. The long-term load evolution can be captured by utilising different scenarios. To be able to apply these scenarios in the optimisation, a reduced set of scenarios needs to be obtained. This is done with a reduction methodology based on the Wasserstein distance. The resulting scenarios can then be evaluated with a minimum maximum-regret approach. This new approach with respect to the modelling of household load uncertainty has been tested on a real LV-network and is shown to be feasible.

Index Terms—Residential load uncertainty, Power system planning, Decision analysis, Genetic algorithms, Probabilistic load flow, Scenario techniques, Stochastic optimisation

I. Introduction

IN the drive towards sustainable energy systems, the residential electricity usage is changing, due to the electrification of heating and transportation sectors, and the penetration of rooftop photovoltaic (PV)-systems. This energy transition has a large effect on the adequacy of the low voltage (LV)-network (<1kV). In most developed countries the LV-network has been developed in the previous century. The ageing assets need to be replaced as they are obsolete from a reliability point of view. Moreover the transition towards a sustainable energy system was not yet taken into account when this network was created. The LV-network needs to be reinforced in order to deal with the changes in the residential electricity consumption and generation [1]. In the coming years, the reinforcement and replacement of the LV-network will, therefore, become more important.

In Europe nearly 60% of the total conductors within the power system are in the LV-network. These lines are most often (56%) underground cables [2]. Reinforcing these systems

M. Nijhuis, M. Gibescu and J.F.G. Cobben are with the Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: m.nijhuis@tue.nl

P.M.S. Carvalho is with the Department of Electrical and Computer Engineering, IST and INESC-Id, Lisbon, Portugal

J.F.G. Cobben is also with Assetmanagement, Liander N.V., Arnhem, The Netherlands

is generally capital cost intensive, as the cost for the trenching is much large than the cost of the actual LV-cable. When a DNO installs a new cable it wants to be sure that the cable would remain operational for the next coming decades, as the technical lifetime of new LV-cables is so long that failure rate statistics collection is deemed uneconomical [3]. When reinforcements are required in the LV-network, the network needs to be analysed to ensure the chosen reinforcement option is robust enough. In order to do this, the DNO applies an expansion planning optimisation to the network. The main challenge with the evaluation of the LV-network is the large load uncertainty. This uncertainty consists of both the long-term load evolution as well as the volatile nature of household load. Therefore the DNO should incorporate the load uncertainty into the LV-network expansion planning. How these uncertainties can be modelled and implemented within the confines of the network expansion planning problem remains an open question.

For the distribution network, many expansion planning strategies under load uncertainty have been described. Most work has been done with respect to the inclusion of the uncertain output of distributed generation in the expansion optimisation process [4]–[7]. Load uncertainty in these approaches is however not taken into account. The use of a fuzzy description of the load [8]–[10] or the type of load [11] is one of the ways to deal with the load uncertainty in the network planning. Load curves for different load magnitudes [12]–[14], and daily load curves [15], [16] are also employed to bound the uncertainty of the load. All these approaches will generate a limited amount of load uncertainty, which is suitable for modelling the household load for the MV network, as the loads are already aggregated to a certain extent, but not for analysis on the LV-network. Another way to deal with uncertainty is by using scenario analysis. This can be done in multiple ways, by incorporating scenario-based load growth in the conventional planning procedure [17] [18] or by adjusting the network optimisation under scenario-based load uncertainty [19]. The final future load can also be seen as a random variable, rather than a deterministic value [20] [21]. Though these approaches are good for evaluating the future load from a system-wide perspective, the uncertainty on the level of the individual household loads is not taken into account.

The implementation of the long-term load evolution uncertainty in combination with the uncertainty caused by the volatile nature of the household load needs to be investigated, in order to use these approaches to accurately estimate the required LV-network strength in the LV-network expansion

planning problem. A method for including both these types of uncertainties in the network expansion planning is presented in this paper. First, a description of these two types of uncertainties and how they can be modelled and implemented is given in Section II. This is followed by the presentation of the network expansion planning problem and how it should be altered to allow for the implementation of both types of uncertainty. To provide insight into how the proposed modelling of uncertainty will affect the outcome of the network expansion planning, the proposed methodology is applied to an LV-network in Section IV. The paper finishes with the conclusion on the modelling and implementation of these uncertainties in the network expansion planning.

II. LOAD UNCERTAINTY IN THE LV-NETWORK

When it comes to the planning of LV-networks, on of the main uncertainties that exists is the load uncertainty. Households consumers make up the largest group of connections to the LV-network. Not only is the household load volatile because it is based on the behaviour of individuals, also how the household load will develop in the future is not known in advance. To be able to accurately assess the best reinforcement options for an LV-network, both these sources of uncertainty need to be taken into account. When taking these uncertainties into account, the modelling of the uncertainty is the first important step. The typical ways of modelling uncertainty are by either continuous or discrete probability density functions, fuzzy sets, lower and upper bounds or scenarios [22]. The load uncertainty in the LV-network takes different forms and therefore two kinds of load uncertainty are assessed. First, the uncertainty which comes from the volatility of the household load and secondly the uncertainty which comes from the long-term load evolution. These uncertainties are modelled through a continuous probability density function and through scenarios, respectively. In the following two subsections, their modelling is explained in more detail. This is followed by a section on the network expansion planning approach and a section on how these modelled uncertainties are taken into account in a specific case study.

A. Household load volatility

In the LV-network, most connected loads are household consumers. Even though these consumers behave in fairly the same way when looking at a large group of consumers, their individual behaviour is highly volatile. To illustrate this fact Fig. 1 has been created.

In Fig. 1, the distribution of the peak load values for different sizes of groups of households is shown. The figure is created by randomly selecting a certain number of household load curves from a database and subsequently calculating the yearly peak value of the group of households. By using 10 000 Monte Carlo samples for the selection of the load curves the distribution of the peak load values for different household group sizes is created. From the figure, it can be seen that the peak load of the aggregated households decreases as the number of households increases as the coincidence factor comes more and more into play. When looking at
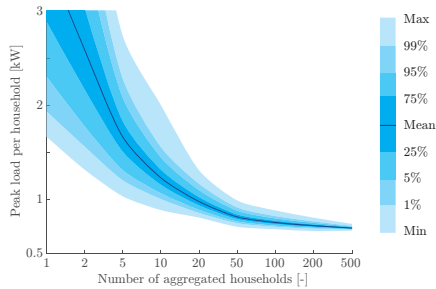
Fig. 1. The distribution of the average peak load of a group of households scaled back to a single household

the uncertainty bound around the household peak load, it also becomes clear that this decreases as the number of households increases. From the perspective of medium and high voltage levels, the residential load can almost be seen as a deterministic parameter, as the number of households which are grouped at a single bus can easily be more than 200. However, for an LV-feeder this is not the case and the uncertainty should be taken into account.

With the introduction of smart meters, this uncertainty should be reduced for the existing network, as for each household a load curve would be available. The household composition in a neighbourhood is however constantly changing (people move away, family situations change, etc.). These changes would have a significant effect on the energy usage in the neighbourhood as the household energy consumption is primarily dependent on the behaviour of the individuals who make up the household [23]. The use of measurement data to reduce this uncertainty is an adequate solution for the short time, though when looking further ahead the measurement data loses its applicability.

When it comes to the planning of the LV-network the general approach to deal with this uncertainty in the load is to take the 95th or 99th percentile value of the curve shown in Fig. 1. The use of this method has one major drawback. All the household loads use the exact same peak value. Through the distribution of the load over a feeder and the non-linearities in the relation between power and voltage, the estimated voltage may not be at the same percentile value. This is because the effect of households at the end of a feeder on the maximum voltage deviation is much larger than for households at the beginning of the feeder. Also if the highly loaded households are all connected to the same branch, the voltage deviation may be high, while the peak load at the feeder level can still be relatively low. In these ways, either more or less risk is taken with the network adequacy than was initially expected based on the chosen percentile value. This is illustrated in Fig. 2.

In this figure, the probability density function of the maximum voltage deviation at the end of the IEEE LV test feeder [24] is given. The maximum voltage deviation is computed

Fig. 2. Probability density function of the maximum voltage deviation at the end of the feeder, also indicated the value of the voltage deviation with all loads assumed equal to the 95th percentile of the peak load
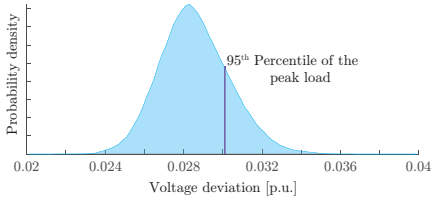
	High		Medium		Low	
	y_{max}	a	y_{max}	a	y_{max}	a
PV	82	0.16	32	0.18	7.2	0.19
EV	63	0.09	28	0.11	4.5	0.11
Heat pump	46	0.06	5.2	0.06	0.1	0.05
Efficiency	0.52	0.13	0.59	0.12	0.96	0.17
Growth	1.25%		0.5%		0%	

by performing load flow calculations with a selection of household load curves from the same database as used for Fig. 1. By assigning different load curves to the households by means of a Monte Carlo simulation, the probability distribution is generated. In the figure the value of the voltage deviation for the 95th percentile of the aggregated peak load obtained by distributing this load proportionally over all households is also shown. The value of the voltage deviation based on this 95th percentile value of the peak load corresponds in this case to the 81st of the maximum voltage deviation (0.03p.u.). By analysing the network based on the percentile of the peak load would thus, in this case, lead to a risk of a shortfall of 19% instead of the much lower 5% risk that would normally be expected when looking at the 95th percentile.

In order to eliminate this increase in risk level when non-linearly mapping the peak powers to nodal voltages, a stochastic approach to the evaluation of each individual household load should be employed. By employing a probabilistic load flow in the evaluation of the LV-network, a better risk assessment can be made. When a probabilistic power flow is employed rather than a deterministic one, a probability density function will be obtained for the current in each branch and the voltage in each node [25] [26]. The use of a probability density function allows for a more advanced approach with respect to modelling the volatility of the household load.

B. Long-term load evolution

In addition to the uncertainty due to the volatility of the household load, also the general trend with respect to the future load is uncertain. When reinforcing the LV-network, the implemented reinforcement option should last for at least the coming thirty years within the network. This is caused by the capital cost-intensive nature and long lifetime of underground cable networks, as explained in the introduction. The load of the households should thus be estimated for the coming decades. In this period the load can change dramatically, leading to a very high uncertainty of this future load. The usual approach to dealing with this kind of uncertainty is by making scenarios for the future load [27] and assessing the reinforcement strategy based on these scenarios in a multi-stage approach. Therefore in this paper a two-stage approach is assumed with T_1 after 10 years and T_2 after 30 years. For the construction of scenarios containing the penetration rate

of new appliances, an S-curve is used:

$$y = \frac{y_{max}}{1 + e^{-a(t-t_0)}} \qquad (1)$$

in which y_{max} is the maximum penetration level, a determines the rate of adoption of the appliance or technology and t_0 the initial penetration rate. In this paper, the values given in Table I are used for the generation of the scenarios. The parameter values describing these scenarios are based on a number of reports which focus on the forecasting of one or more of the different scenario elements [28]–[35]. For the scenario with respect to the economic growth, a simple growth rate is used, the applied growth rate is also given in Table I.

The scenarios are generated by making a combination of the different elements as shown in Table I. These scenarios should be addressed by the network expansion planning approach, while incorporating the volatility of the household load. A Bottom-up Markov Chain Monte Carlo scenario-based method for generating the household load is therefore applied, based on [36]. With this method, a large number of individual household load curves can be created for each given scenario and each moment in the future. Based on these curves, the distribution of possible household load values during conventional and PV peak times can be generated. These probability density functions would give a better estimation of the load than a simple peak value as they allow for variation in the generated scenario peak [37]. For the scenario options shown in Table I the peak load is shown in Fig. 3. In the figure also the initial time (T_0), the second investment time T_1 (after 10 years) and the final network time T_2 (after 30 years) are indicated.

The figure is generated by using each unique combination of the different scenario drivers and thus consists of 243 ($= 3^5$) scenarios. From the figure, the large spread among scenarios becomes apparent. The network cannot be analysed on all these possible scenario outcomes, as that would be computationally very expensive. Therefore, a subset of scenarios needs to be selected. The selection of the subset of scenarios is based on minimising the Wasserstein distance [38] between the original and reduced set. The Wasserstein distance can be seen as the minimum required transportation cost to move all the data from the original scenario set to the reduced subset. According to [39] this can be estimated by applying a k-medians heuristic [40] which leads to the following minimisation problem:

$$\min_{u \in \{1,...,N\}} \sum_{\substack{i=1 \\ i \neq u}}^{N} \mathcal{P}_i c(s_u, s_i) \qquad (2)$$
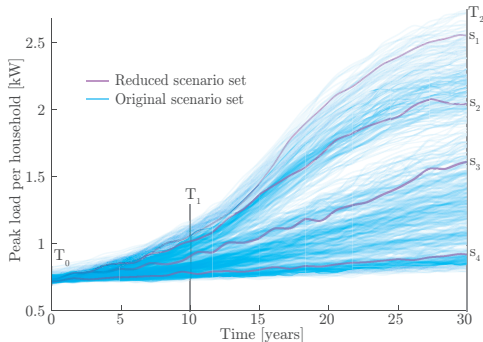
Fig. 3. Possible peak load scenarios for the household load and the four scenarios obtained through scenario reduction

where c is the cost to transport from an original scenario s_u to a scenario in the reduced scenario set s_i and $\mathcal{P}_i$ is the probability of belonging to scenario s_i. The final subset of scenarios can be estimated by starting from the complete set of scenarios and performing a backwards scenario selection process. This entails applying the minimisation (2), generating the set of N-1 scenarios and repeating the selection procedure until the desired number of scenarios is obtained. This will generate a reduced set of k scenarios in which each of the scenarios will have a probability that is based on the probabilities of the closest scenarios from the original set. The approach is applied with the following metrics for each scenario: the average load, the peak load and the off-peak load, all three measured at both T_1 and T_2. In this paper, a subset of $k = 4$ scenarios is used for the generation of the results. The resulting scenarios are also indicated in Fig. 3 as s_1, s_2, s_3 and s_4.

For the analysis of these different scenarios, it must be noted that scenarios are generated at a national level. This implies that if a certain scenario from the set of generated scenarios pans out, all of the reinforced networks will experience this scenario. Taking a risk-neutral expected value approach (as technically done with the local load uncertainty) is therefore not advisable as basically all the investments within the same scenario area have to be pooled and seen as a single invest-ment. A more risk averse approach is therefore adopted. The different network reinforcement options are analysed based on the minimum regret over the selected scenarios. For the calculation of the minimum regret, the lowest cost per scenario is estimated by applying a deterministic perfect information approach to the network expansion problem.

III. Network Expansion Planning Approach

The general formulation of the LV-network expansion plan-ning problem is presented in this section. The scope of this planning problem is up to and including the MV/LV transformer. The MV network is only taken into account in a shallow manner. As the planning of the MV-network is

different, for instance, the reconfiguration plays an important role, while in the LV network this is deemed uneconomical. The main objective of the LV-network planning is to achieve the lowest possible investment cost and operational cost related to energy losses and reliability while meeting the actual and future energy demand. The reliability cost estimation function is given in [41] and the loss estimate is taken from [42]. Both have been adjusted to be implemented in a planning algorithm. These two costs in combination with the investment cost can be captured in the following function for the total cost:

$$C_{tot}(T) = C_{inv}(T) + C_r(T) + C_{loss}(T) \tag{3}$$

with,

$$C_{inv}(T) = C_{inv,fix} + \sum_{d \in D_{cbl}} (dL_d C_{cbl}) + \\ \sum_{d \in D_{fdr}} (dL_d C_{cbl} + C_{fdr}) + \tag{4} \\ \sum_{d \in D_{sub}} (dL_{sub} C_{cbl,mv} + dC_{sub})$$

$$C_r(T) = p_{av} \sum_{f=0}^{FDR} (L_{fdr} \lambda_{cbl} HH_f Tr_{cbl}) \tag{5}$$

$$C_{loss}(T) = p_e \frac{Z_{tot} \sum_{n=1}^{N_B} I_n(t)^2}{F_d} \tag{6}$$

where C_{inv} is the investment cost, C_r the reliability cost, C_{loss} is the cost of the LV-network losses, T the time instance, $C_{inv,fix}$ a fixed investment charge, d a binary variable indi-cating whether an reinforcement will be made, D_{cbl} the list of branches to be possibly reinforced, L_d the length of option d, C_{cbl} the cost of laying a meter LV-cable, D_{fdr} the list of possible new feeders, C_{fdr} the cost of adding a feeder to a substation, D_{sub} the list of possible new substation, $C_{cbl,mv}$ the cost of laying a meter of MV-cable, C_{sub} the cost of a new substation, p_{av} the price of an outage minute for a household, FDR the set of all feeders, L_{fdr} the total length of feeder f, λ_{cbl} the failure rate of an LV-cable, HH_f the number of households connected to feeder f, Tr_{cbl} the mean time to repair a damaged LV-cable, p_e the electricity price, Z_{tot} the total impedance within the network, N_B the number of buses, $I_n(t)$ the current injected at bus n and F_d a distribution factor given by:

$$F_d = \frac{\sum_{n=1}^{N_B r} \left(l_{br,n} \sum_{k=n}^{N_{B,n-ds}} S_{k,max}^2 \right)}{\sum_{n=1}^{N_B} S_{n,max}^2} \tag{7}$$

where $N_B r$ is the number of branches in the network, $l_{br,n}$ is the length of branch n, $S_{k,max}$ the maximum power injected at bus k and $N_{B,n-ds}$ the downstream buses of branch n.

The problem is described as a two-stage investment pro-blem, where investments can be made at T_0 and T_1 and the final network is assessed at T_2. This leads to the following optimisation problem:

$$\min \left[C_{tot}(T_0) + \frac{C_{tot}(T_1)}{(1+r)^{10}} + \frac{C_{tot}(T_2)}{(1+r)^{30}} \right] \tag{8}$$

where r is the depreciation rate. The optimisation is typically constrained by the allowable voltage deviations and the current ratings of the components:

$$\Delta V_n < 0.05 \tag{9}$$

$$|I_c| < I_{c,rated} \tag{10}$$

where ΔV_n is the per unit change in voltage from the nominal value of 1 [p.u.] at node n, usually taken at 0.05 to keep the voltage within the 0.95-1.05 band, allowing for some voltage fluctuations within the MV-network and I_c the current flowing trough and $I_{c,rated}$ the rated current carrying capacity of component c (branch or transformer). In the optimisation, loads will have to be assigned to each of the buses with a connected household. As discussed in Section II-A, each load will be modelled as a probability density function. This modelling is incompatible with the conventional constraints (9) and (10), as the voltage and current are no longer deterministic variables. The chance of violating either the current loading or the voltage deviation constraints can be estimated from the probability density functions obtained from the probabilistic load flow. When either of these two constraints are violated, the DNO should reinforce the network to ensure it is operating again within the limits. The constraints with respect to the voltage and current can thus be compounded and replaced by a constraint with respect to the chance of overloads or voltage violations:

$$\mathcal{P}(\Delta V_n > 0.05) + \mathcal{P}(I_c > 1 | \Delta V_n < 0.05) \tag{11}$$

where $\mathcal{P}(\Delta V_n > 0.05)$ indicates the chance that the voltage is more than 0.05 from the nominal value and $\mathcal{P}(I_c > 1 | \Delta V_n < 0.05)$ is the chance that the current is more than the nominal value while the voltage is within the 0.05 limits. This approach is needed to ensure that the instances at which the voltage and the current are outside the limits are not counted twice, as the risk of a voltage violation and a current overloading have a certain level of dependency. The best reinforcement option available to increase the network if the chance of network inadequacy becomes too high can be determined within the optimisation procedure. Based on this reinforcement option, constraint (11) can be relaxed into a penalty function, generating an additional cost component in the objective function. This cost component is given by:

$$C_{pen} = C_{inv,s}(r-1)r^T \times \\ (\mathcal{P}(\Delta V > 0.05) + \mathcal{P}(I_c > 1 | \Delta V < 0.05)) \tag{12}$$

where $C_{inv,s}$ is the investment cost of the subsequent reinforcement option, e.g. the best investment option for T_1 at T_0. As the investment cost associated with the subsequent best reinforcement options will be accounted for when the network is altered and reassessed based on this reinforcement, only the cost of moving the investment forwards should be accounted for when calculating the penalty cost. In this way, the uncertainty which comes from the volatile household load is taken into account. In the two-stage approach there is no investment at T_2, however, an estimate of this investment is necessary to be able to calculate (12). The amount of investment at this stage, therefore, needs to be estimated when applying this approach.

IV. CASE STUDY

To illustrate how the change in the modelling and implementation of the load uncertainty affects the network expansion planning, a case study has been performed. In the case study, the modelling of uncertainty as described in the paper is applied to an existing LV-network. The network expansion planning optimisation has been performed with and without the modelling and implementation of uncertainty as proposed in Section II. To be able to perform the network expansion planning optimisation, a generic approach for solving the optimisation has to be established first, hereafter the results of the case study can be discussed in more detail.

A. Approach

The optimisation problem as formulated in Section III should be solved while taking into account the load uncertainties as described in Section II. With these uncertainties, the linearisation of the problem as formulated seems infeasible. Therefore, a heuristic solution method is applied. Many different heuristics have been applied for solving power system planning problems, e.g. genetic algorithms, particle swarm optimisation and Tabu search. Due to their ability to successfully tackle large combinatorial problems, genetic algorithms have often been applied to solve multi-stage network expansion planning problems [6], [43]–[46]. Therefore, a genetic algorithm is applied in this paper as well. The applied genetic algorithm has the following steps:

Step 1) Generate reduced scenario set to capture the load evolution
Step 2) Initialise population of network reinforcements
Step 3) Determine optimal timing for the reinforcements of each individual under each scenario
Step 4) Evaluate the population and determine maximum regret
Step 5) Selection, cross-over and mutation of the population of reinforcement solutions
Step 6) Repeat steps 3-5 until number of iterations reached

With the evaluation of the different options, the optimum investment level for each scenario is determined separately at T_1. In the first stage, T_0, the investments which are required are all the same for each scenario, as these are investments which should be done at the present moment in time. No information on which scenario will materialise is present. At T_1 there might be more information present, so the applied reinforcement options can differ between scenarios. This is incorporated into the optimisation by not necessarily implementing all the reinforcements at the second stage, T_1. The required level of reinforcements at the second stage is estimated for each scenario during the determination of the optimal timing.

To evaluate the optimisation with respect to the handling of uncertainty, two simulations are performed. With these simulations, the correctness of the applied scenario reduction approach, as well as the advantages of using a probability density function rather than a deterministic value for the household load, are illustrated. The different parameters which are used in the optimisation are given in Table II. The household

TABLE II
DATA USED FOR THE NETWORK EXPANSION PLANNING OPTIMISATION

Symbol	Component	value
C_{cbl}	LV-Cable [€\mtr]	65.5
$C_{cbl,mv}$	MV-Cable [€\mtr]	77
C_{fdr}	LV-busbar expansion [€]	4200
p_e	Electricity price [€\MWh]	0.73
C_{sub}	Additional MV/LV substation [€]	40000
$C_{inv,fix}$	Fixed investment cost [€]	8300
λ_{cbl}	Cable failure rate [\(yr km)]	0.038
Tr_{cbl}	Repair time [min]	200

TABLE III
CHARACTERISTICS OF THE LV-NETWORK USED IN THE CASE STUDY

Feeder	Length [m]	Connections	Mean impedance [$\Omega\backslash km$]
Orange	340	41	0.335
Green	457	62	0.467
Purple	564	57	0.393
Light blue	371	45	0.569
Red	436	55	0.280
Blue	241	41	0.310

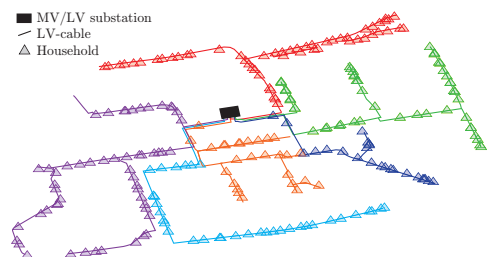■ MV/LV substation
⌐ LV-cable
△ Household

Fig. 4. The network used for the case study

load which is used in the evaluation is based on a bottom-up load model [36] with all possible scenario combinations implemented as indicated in Table I. The probability density functions of the household load are estimated by a Gaussian mixture distribution and evaluated with a probabilistic load flow designed for Gaussian mixture distributions [47].

For the simulation, a LV-network from a Dutch DNO is used. The network is fed by a 400kVA transformer, consisting of 6 feeders connecting 301 customers. The network was originally built in the sixties and is now operating close to its limits. A graphical overview of the network is given in Fig. 4, while the main characteristics of the LV-network are shown in Table III.

As this network is operating close to its limits, it will be reinforced soon. The network expansion planning optimisation is applied to this network in a number of different ways.

1) The optimisation is applied on the reduced scenario set with the modelling of the load uncertainty as described in Section II.
2) The network is planned by using a deterministic approach to the household load volatility rather than a probabilistic one. The load of all households on a single feeder is assumed to be equal. The level of the load is determined by taking the 95th percentile of the
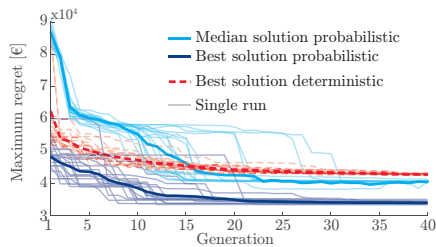

Fig. 5. Evaluation of the performance of the genetic algorithm for probabilistic versus deterministic approach with a reduced scenario set

aggregated peak load of the feeder. The other aspects of the optimisation remain unchanged.
3) The proposed optimisation approach is applied to the full scenario space containing 243 scenarios, rather than a reduced set of 4 scenarios.

By solving the same optimisation problem in three different ways, a better understanding of the effect of modelling and implementing load uncertainty is provided.

B. Results

The results of the three network expansion planning optimisation problems are given in the following section.

1) Probabilistic vs. deterministic approach: First, a closer look is taken at the performance of the genetic algorithm when applying a probabilistic approach, where the loads are modelled through probability density functions. This is compared to a deterministic approach, where all the loads have a constant value. Fig. 5 has been created, to show the differences. In this figure, the results are shown for each generation in the genetic optimisation. The figure shows the minimum maximum regret (dark blue) and the median value of the population (light blue) for the probabilistic approach as well as the minimum maximum regret when using a deterministic evaluation (red). As a genetic algorithm has inherent variability, the optimisation has been performed 20 times. The thick lines shows the average over the 20 trials, while the thin lines show a single trials.

In the figure, a number of interesting things can be seen. First of all, it shows that both the deterministic and the probabilistic approaches optimise and converge to a certain minimum value. The spread around this minimum is slightly higher for the probabilistic approach than for the deterministic one, but the spread remains under the 5%. The median value also saturates after 23 generations. The population appears to have converged from that point onwards with the higher median coming from the mutations in the population. As the initialisation of the algorithm is done purely random, it can be seen that the algorithm finds the optimal even for the probabilistic-based evaluation. Another observation that can be made from the figure is that the deterministic solution converges similarly to the probabilistic one. All things considered, the probabilistic evaluation can be deemed a feasible approach

TABLE IV
RESULTING COST OF THE OPTIMISATION WITH DETERMINISTICALLY AND
PROBABILISTICALLY MODELLED HOUSEHOLD LOADS

	Probabilistic	Deterministic
Initial	13 110	16 202
Scenario I	43 837	54 627
Scenario II	32 490	37 860
Scenario III	29 245	33 128
Scenario IV	31 722	36 826

TABLE V
RESULTS OF THE OPTIMISATION WITH REDUCED SCENARIO SPACE AND
FULL SCENARIO SPACE

	Full	Reduced
Initial investment at T_0 [€]	13 110	13 110
Maximum regret [€]	49 771	43 837
Computation time [min.]	957	21.1

from an algorithmic point of view. Lastly, the deterministic solution has a significantly higher maximum regret compared to the probabilistic one, as expected. In order to determine where this difference comes from, Table IV is created. In the table, the results of the probabilistic and deterministic solutions are compared, based on their initial investment and the resulting expected value of each scenario.

Table IV shows that for every scenario the expected costs are higher for the deterministic approach than for the probabilistic one. This does not mean that the solution provided by the deterministic case is necessarily worse compared to the probabilistic one. As both approaches are evaluated in a different way the estimation cost would therefore differ. The higher value for the deterministic case comes from the need to invest up to a point where the constraints are met, while in the probabilistic case the use of a penalty function based on the chance of violating the limits can defer the investments. As deferring the investments until they are actually needed is the common approach to LV-network expansion, the probabilistic method generates a more realistic cost estimate than the deterministic one. In the table, also the initial investment, at T_0, has been given. The level of the initial investment corroborates this explanation, as the initial investment of the deterministic case is also higher than that of the probabilistic one. These results show that the proposed approach with respect to modelling and implementing the volatility in the household load offer clear advantages over the commonly used approach.

2) Full vs. reduced scenario set: The next comparison which has been performed is the comparison between the optimisation with the reduced (4 scenarios) and the full set of scenarios (243 scenarios). The maximum regret, initial investment and computation time of the two network expansion planning approaches are given in Table V. From the table, the first thing that becomes apparent is the difference in computational time. For the reduced scenario set, the computational time is a factor 45 lower compared to 60 times fewer scenarios which need to be evaluated. A computational time reduction of a factor 60 is not obtained as the time it takes to run the scenario reduction algorithm is included and also because in the full scenario set relatively many low growth scenarios have been

included which require less time to optimise. The maximum regret when applying the reduced scenario set is slightly lower compared to the full scenario set case. The extreme scenarios, the upper and lower bounds of the scenario space, are not included in the reduced scenario set (indicated as $s_1 - s_4$ in Fig. 3). These scenarios generate the highest regret when applying the full scenario space, the absence of these scenarios in the reduced set thus leads to a lower maximum regret. The initial investment for both scenario sets is the same, indicating the feasibility of the proposed approach to the scenario reduction.

V. CONCLUSION

In this paper, a new approach to the modelling and implementation of uncertainty in the LV-network expansion planning is applied. The main idea behind the approach is to decompose the load uncertainty into two parts: the uncertainty coming from the volatility of the household load and the uncertainty coming from the long-term household load evolution. The volatility of the household load has been incorporated by modelling the household load with a probability density function and evaluating this load based on a probabilistic load flow. The long-term household load evolution is modelled through scenarios. To be able to apply these scenarios in an optimisation, a reduced set of scenarios has been generated and a minimum maximum regret approach was applied to evaluate this reduced scenario set.

The proposed approach has been validated with simulations ran over a typical Dutch LV network. Simulation results have proven the feasibility of the approach proposed as well as its advantages over the current practice of using peak loads Advantages include a more accurate estimation of the total network investment cost as well as a reduced level of initial investment. The approach to the reduction of the scenario set has shown that even though the maximum regret differs between solutions found, the initial investment remains the same, demonstrating that the reduction of the scenario space has little effect on the outcome of the network expansion planning optimisation.

REFERENCES

[1] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Assessment of the impacts of the renewable energy and ICT driven energy transition on distribution networks," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1003–1014, aug 2015.

[2] Eurelectric, "Power distribution in Europe - Facts and figures," Tech. Rep., 2013. [Online]. Available: http://www.eurelectric.org/media/113155/dso_report-web_final-2013-030-0764-01-e.pdf

[3] W. A. Thue, "Cable Performance," in *Electrical Power Cable Engineering*. CRC Press, dec 1975, pp. 395–400.

[4] R. Hemmati, R.-A. Hooshmand, and N. Taheri, "Distribution network expansion planning and DG placement in the presence of uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 665–673, 2015.

[5] K. Zou, A. P. Agalgaonkar, K. M. Muttaqi, and S. Perera, "Distribution system planning with incorporating DG reactive capability and system uncertainties," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 112–123, 2012.

[6] C. L. T. Borges and V. F. Martins, "Multistage expansion planning for active distribution networks under demand and Distributed Generation uncertainties," *International Journal of Electrical Power and Energy Systems*, vol. 36, no. 1, pp. 107–116, jan 2012.

[7] M. E. Samper and A. Vargas, "Investment decisions in distribution networks under uncertainty with distributed generation - Part i: Model formulation," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2331–2340, jan 2013.

[8] N. Kagan and R. Adams, "Electrical power distribution systems planning using fuzzy mathematical programming," *International Journal of Electrical Power & Energy Systems*, vol. 16, no. 3, pp. 191–196, 1994.

[9] I. Ramirez-Rosado and J. Dominguez-Navarro, "New Multiobjective Tabu Search Algorithm for Fuzzy Optimal Planning of Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 224–233, 2006.

[10] M. Skok, S. Krajcar, and D. Skrlec, "Dynamic Planning of Medium Voltage Open-Loop Distribution Networks under Uncertainty," in *Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems*, vol. 2005. IEEE, nov 2005, pp. 551–556.

[11] M. A. Matos and M. T. de Leão, "Electric distribution systems planning with fuzzy loads," *International Transactions in Operational Research*, vol. 2, no. 3, pp. 287–296, 1995.

[12] M. Jalali, K. Zare, and M. T. Hagh, "A multi-stage MINLP-based model for sub-transmission system expansion planning considering the placement of DG units," *International Journal of Electrical Power and Energy Systems*, vol. 63, pp. 8–16, 2014.

[13] S. A. Taher and S. A. Afsari, "Optimal location and sizing of DSTA-TCOM in distribution systems by immune algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 60, pp. 34–44, 2014.

[14] R. S. Rao, K. Ravindra, K. Satish, and S. V. L. Narasimham, "Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 317–325, 2013.

[15] M. Sadeghi and M. Kalantar, "Multi types DG expansion dynamic planning in distribution system under stochastic conditions using Covariance Matrix Adaptation Evolutionary Strategy and Monte-Carlo simulation," *Energy Conversion and Management*, vol. 87, pp. 455–471, 2014.

[16] B. Arandian, R.-A. Hooshmand, and E. Gholipour, "Decreasing activity cost of a distribution system company by reconfiguration and power generation control of DGs based on shuffled frog leaping algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 61, pp. 48–55, 2014.

[17] B. B. Souza, E. G. Carrano, O. M. Neto, and R. H. Takahashi, "Immune system memetic algorithm for power distribution network design with load evolution uncertainty," *Electric Power Systems Research*, vol. 81, no. 2, pp. 527–537, feb 2011.

[18] D. Wang, L. N. Ochoa, and G. Harrison, "Modified GA and data envelopment analysis for multistage distribution network expansion planning under uncertainty," *2011 IEEE Power and Energy Society General Meeting*, vol. 26, no. 2, pp. 1–1, jan 2011.

[19] E. G. Carrano, F. G. Guimarães, R. H. C. Takahashi, O. M. Neto, and F. Campelo, "Electric distribution network expansion under load-evolution uncertainty using an immun system inspired algorithm," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 851–861, jan 2007.

[20] E. G. Carrano, C. G. Tarôco, O. M. Neto, and R. H. Takahashi, "A multiobjective hybrid evolutionary algorithm for robust design of distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 645–656, dec 2014.

[21] N. C. Koutsoukis, P. S. Georgilakis, and N. D. Hatziargyriou, "A Tabu search method for distribution network planning considering distributed generation and uncertainties," in *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, jul 2014, pp. 1–6.

[22] P. M. Verderame, J. A. Elia, J. Li, and C. A. Floudas, "Planning and Scheduling under Uncertainty: A Review Across Multiple Sectors," *Industrial & Engineering Chemistry Research*, vol. 49, no. 9, pp. 3993–4017, may 2010.

[23] A. Druckman and T. Jackson, "Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model," *Energy Policy*, vol. 36, no. 8, pp. 3167–3182, jan 2008.

[24] D. Test and F. Working. (2015, jan) Distribution Test Feeders. [Online]. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html

[25] A. M. L. da Silva, M. B. D. Coutto Filho, S. M. P. Ribeiro, V. L. Arienti, and R. N. Allan, "Probabilistic load flow techniques applied to power system expansion planning," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1047–1053, jan 1990.

[26] H. Yu, C. Y. Chung, K. P. Wong, and J. H. Zhang, "A chance constrained transmission network expansion planning method with consideration of

load and wind farm uncertainties," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1568–1576, jan 2009.

[27] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456–462, jan 2014.

[28] J. J. C. Bruggink, "The next 50 years: Four European energy futures," Tech. Rep., may 2005.

[29] E. Veldman, "Power Play Impacts of flexibility in future residential electricity demand on distribution network utilisation," Ph.D. dissertation, dec 2013.

[30] Pwc, "Energie-Nederland Financial and economic impact of a changing energy market," Tech. Rep. March, mar 2013.

[31] C. Hellinga, "De energievoorziening van Nederland Vandaag (en morgen ?)," Tech. Rep., oct 2010.

[32] S. Teske, G. Onufrio, A. Gianni, E. Rainer, and H. Rahlwes, "Energy [r]evolution: A sustainable Netherlands energy outlook," Tech. Rep., may 2013.

[33] S. Moorman and M. Kansen, "Naar duurzaam wegverkeer in 2050 Een verkenning van mogelijke opties," Tech. Rep., nov 2011.

[34] DNV KEMA, "Rol van gas in nieuwe wijken," Tech. Rep. april, apr 2013.

[35] "Analyze Routekaart 2050," Tech. Rep., nov 2011.

[36] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data," *Energy and Buildings*, vol. 112, pp. 121–129, jan 2016.

[37] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, may 2010.

[38] G. Pflug, "Scenario tree generation for multiperiod financial optimization by optimal discretization," *Mathematical Programming*, vol. 89, no. 2, pp. 251–271, jan 2001.

[39] H. Heitsch and W. Römisch, "A note on scenario reduction for two-stage stochastic programs," *Operations Research Letters*, vol. 35, no. 6, pp. 731–738, nov 2007.

[40] M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys, "A Constant-Factor Approximation Algorithm for the k-Median Problem," *Journal of Computer and System Sciences*, vol. 65, no. 1, pp. 129–149, aug 2002.

[41] G. Celli, E. Ghiani, S. Mocci, and F. Pilo, "A multiobjective evolutionary algorithm for the sizing and siting of distributed generation," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 750–757, may 2005.

[42] P. Rao and R. Deekshit, "Energy Loss Estimation in Distribution Feeders," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1092–1100, jul 2006.

[43] E. Carrano, R. Cardoso, R. Takahashi, C. Fonseca, and O. Neto, "Power distribution network expansion scheduling using dynamic programming genetic algorithm," *IET Generation, Transmission & Distribution*, vol. 2, no. 3, p. 444, 2008.

[44] M. S. Nazar, M. R. Haghifam, and M. Nažar, "A scenario driven multiobjective PrimarySecondary Distribution System Expansion Planning algorithm in the presence of wholesaleretail market," *International Journal of Electrical Power & Energy Systems*, vol. 40, no. 1, pp. 29–45, sep 2012.

[45] H. Falaghi, C. Singh, M. R. Haghifam, and M. Ramezani, "DG integrated multistage distribution system expansion planning," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 8, pp. 1489–1497, 2011.

[46] E. Naderi, H. Seifi, and M. S. Sepasian, "A dynamic approach for distribution system planning considering distributed generation," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1313–1322, 2012.

[47] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Gaussian Mixture Based Probabilistic Load Flow for LV-Network Planning," *IEEE Transactions on Power Systems*, pp. 1–1, 2016.

Low Voltage Multi-Stage Network Expansion Planning Under Uncertainty: a Bilevel Evolutionary Optimisation Approach

Michiel Nijhuis, *Member, IEEE,* Pedro M.S. Carvalho, Madeleine Gibescu, *Member, IEEE,* and Sjef J.F.G. Cobben

Abstract—In this paper, a bilevel optimisation approach for the problem of low voltage (LV) network expansion planning under uncertainty is formulated. Uncertainty is modelled by a set of scenarios. Specific genetic algorithms (GA) are designed and used (i) to find the initial optimal network investments (upper-level optimisation) and (ii) to find the optimal sequence of such investments over the planning horizon (lower-level optimisation). By formulating the LV-network expansion problem as a bilevel optimisation problem, the initial investment can be evaluated with more detail, while subsequent investments can still be taken into account. Details of the scenario modelling approach are given together with the GA framework used to solve the problem of optimisation under uncertainty. Results of a case study with real LV-networks are presented and compared with the results from an integrated approach to show that the proposed method reduces computational burden while preserving the optimality of investment.

Index Terms—Decision theory, Genetic algorithms, Power system planning, Probabilistic load flow, Uncertainty modelling, Scenario analysis, Bi-level optimisation

I. Introduction

THE loading of the low voltage (LV) network is changing. With the introduction of electric vehicles and rooftop photovoltaics, the network peak load is expected to rise and reverse power flows are becoming possible. These changes have much influence on the adequacy and the operation of the LV-network [1]. In order to handle these changes, the distribution network operator (DNO) needs to reinforce the LV-network in many places over the coming years. When the LV-network is close to its operating limits or the limits have already been reached, the DNO will evaluate the possible reinforcement options. When evaluating the reinforcement options the DNO has to take into account the long economic lifetime of additional LV-cables (>30 years) and the negative asset recovery value (removing a cable that is no longer required comes at a cost for the DNO) for underground cables. The DNO, therefore, assesses the LV-network over a long horizon to ensure the network could handle almost all foreseen load changes. This LV-network expansion planning problem is, from an optimisation standpoint, a mixed integer non-linear problem [2]. As the solution should span a large time horizon,

multi-stage approaches are generally preferred, thus greatly increasing the complexity of the problem. The LV-network expansion planning is therefore often solved by using heuristic as opposed to classic optimisation techniques.

The main techniques that are applied for distribution network planning are simulated annealing [3], Tabu search [4], partial swarm optimisation [5], ant colony optimisation [6] and genetic algorithms [7]. These types of algorithms have been compared in a number of typical power system problems [8]–[11]. These comparisons show that the actual implementation of a certain algorithm is more important than the actual type of algorithm that is used. A careful consideration of the description of the multi-stage LV-network expansion planning problem is thus required. The calculation of all the stages in a multi-stage approach is often done in a single integrated approach [12]–[15], which, through the combinatorial nature of the problem, vastly increases the computational complexity of the problem. The solutions for a multi-stage integrated problem are not only the reinforcement options, that are applied, but also the timing of the reinforcement options. Both the scheduling of the reinforcements [16] and the choice of which reinforcements are to be applied are already complex problems. From a practical point of view, only the investment in the first stage and an indication of the expected total cost is of interest. Therefore, the problem can be split into two parts, the optimisation of the first investment and the optimisation of the subsequent investment decisions. By applying a bilevel optimisation, the differences in the requirements with respect to the first stage versus the sequential stages can be fully exploited. This kind of decomposition of the problem is commonly applied when distributed generation is included in the distribution network expansion planning through the use of an optimal power flow [17]–[19].

In this paper, we propose a genetic algorithm (GA) specifically suited for the LV-network expansion planning problem. The GA will accurately determine the initial investment the DNO has to make, and only provide a deterministic estimation of the subsequent investments. The paper is organised as follows. In Section II the description of the multi-stage LV-network expansion problem is given. In Section III the implementation of this multi-stage LV-network expansion problem in a GA is discussed. Section IV shows the performance of the proposed approach by applying the developed approach to an LV-network, followed by the conclusions in Section V.

M. Nijhuis, M. Gibescu and J.F.G. Cobben are with the Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: m.nijhuis@tue.nl

P.M.S. Carvalho is with the Department of Electrical and Computer Engineering, IST and INESC-Id, Lisbon, Portugal

J.F.G. Cobben is also with Asset Management, Liander N.V., Arnhem, The Netherlands

II. MULTI-STAGE NETWORK EXPANSION PLANNING UNDER UNCERTAINTY

In this section, the formulation of the LV-network expansion planning problem is given, followed by a discussion on how this problem can be tackled in a computationally feasible manner.

A. LV-network expansion planning

LV-network expansion planning is performed as a cost minimisation problem that takes into account the investment, operational and reliability costs. These costs are evaluated in a multi-stage approach where the load uncertainty is modelled by using multiple scenarios as given by:

$$\min C_{inv}(T_0) + C_r(T_0) + C_{loss}(T_0) + \qquad (1)$$
$$\frac{1}{||S||} \sum_{s \in S} \sum_{i=1}^{n} \frac{C_{inv}(T_i)(s) + C_r(T_i)(s) + C_{loss}(T_i)(s)}{(1+r)^{T_i}}$$

where C_{inv} is the investment cost, C_r is the reliability cost, C_{loss} is the cost of the energy losses, T_0 the initial time, T_i the ith time step, T_n the final evaluation time, S the set of scenarios s and r the depreciation rate. The costs are computed as combined costs over all scenarios. The different cost components within this network expansion planning problem can be calculated by applying the following equations:

$$C_{inv}(T_i) = \sum_{j=0}^{||X||} X_j(T_i) \qquad (2)$$

$$X_j = C_{inv,fix} + L_{b_n \to b_{n+1}} C_{cbl} \qquad \text{or}$$
$$X_j = C_{inv,fix} + L_{b_n \to b_0} C_{cbl} + C_{fdr} \qquad \text{or}$$
$$X_j = C_{inv,fix} + L_{mv} C_{cbl,mv} + C_{sub} + L_{b_n \to b_{sub}} C_{cbl}$$

$$C_r = p_{av} \sum_{f=0}^{FDR} (L_{fdr} \lambda_{cbl} HH_f Tr_{cbl}) \qquad (3)$$

$$C_{loss} = p_e \sum_{sn=0}^{SN} \left(\frac{\sum_{t=1}^{24} S_{sn}(t)^2}{\max S_{sn}} \alpha_{L,sn} \right) \qquad (4)$$

where X the set of reinforcement investments which includes reinforcing branches, adding feeders and adding substations, $C_{inv,fix}$ a fixed investment cost, $L_{b_n \to b_{n+1}}$ the length of the branch which would be replaced, C_{cbl} the cost of laying a meter LV-cable, $L_{b_n \to b_0}$ the length from bus n to bus 0 (the MV/LV substation), C_{fdr} the cost of adding a feeder to a substation, $C_{cbl,mv}$ the cost of laying a meter of MV-cable, C_{sub} the cost of a new substation, L_{mv} the length to the new substation from the nearest substation, C_r is the reliability cost, C_{loss} is the estimated cost of the transport losses, p_{av} the price of an outage minute for a household, FDR the set of all feeders, L_{fdr} the total length of feeder f, λ_{cbl} the failure rate of an LV-cable, HH_f the number of households connected to feeder f, Tr_{cbl} the mean time to repair a damaged LV-cable, p_e the average electricity price, sn the sub-network connected to a single transformer, $S_{sn}(t)$ the apparent power load at hour t in the average load curve and $\alpha_{L,sn}$ a loss factor of sub-network sn. This load factor is determined for each of the installed MV/LV transformers separately and is calculated by

assuming a linear relationship between the load and the loss which can be estimated from the load flow.

The network expansion planning is constrained by the voltage limits and the current rating of the branches and transformers:

$$\Delta U < 0.05 \qquad (5)$$
$$|I_c| < I_{c,rated} \qquad (6)$$

where ΔU is the per unit deviation in voltage from the nominal value of 1 [p.u.], usually taken at 0.05 to keep the voltage within the 0.95-1.05 band to allow for voltage fluctuations within the MV-network and I_{rated} the rated current carrying capacity of a component c. The evaluation of these constraints is carried out by load flow calculations. The simple optimisation problem formulation in (1)-(6) needs to be adjusted to allow for the inclusion of uncertainty in the multi-stage approach.

The main drawback of a formulation of the network expansion planning problem as a multi-stage problem, with the inclusion of load uncertainty based on scenarios, is the increased computational challenge that it brings. For the inclusion of scenarios to model the uncertainty, the evaluation complexity is at least multiplied by the number of points within a scenario-tree and even more if a more realistic scenario representation is applied. This increase in the number of possible solutions is such that the problem as formulated should be adjusted in order to be solvable. The multi-stage problem can be split into a bilevel optimisation problem where the reinforcements that are required are selected separately from the timing of the reinforcements. This will allow for using solution techniques specifically designed for these types of combinatorial problems.

B. Multi-stage network expansion planning under uncertainty

To determine how the multi-stage problem formulation discussed in the previous section can best be altered to make a multi-stage approach computationally feasible, a closer look at why a multi-stage approach is preferable should be taken. The two main reasons are: the difference in magnitude of required subsequent reinforcements and the reduced set of subsequent reinforcement options available.

Depending on the chosen reinforcement option, the resulting network strength can differ significantly. If a single-stage approach is applied the reinforcement should be sufficient to adequately handle the load at the evaluation time. Whether or not an additional incremental change in the load would lead to a violation of the constraints is not taken into account. A network capable of handling an additional load change would provide additional value; however, in the single stage approach, this is neglected.

When a certain reinforcement option is applied to a network, it changes the space of possibilities of future reinforcements. For instance, when depleting most incremental reinforcement options, the flexibility to deal with future load changes diminishes, while placing a new substation would lead to more incremental reinforcement options in the future. This should

also be accounted for in order to accurately evaluate a certain reinforcement option.

Next to this, the use of a multi-stage approach also resembles how the network reinforcements are carried out in reality. The normal procedure when reinforcing LV-networks is to wait until the network becomes overloaded, experiences voltage problems or the cables are no longer reliable enough, and then take action. Assuming one will reinforce the network at set intervals is, therefore, a crude representation of the actual reinforcement procedure. When reinforcements are applied, the network is made robust enough to be able to handle the load changes over the coming decades, as the reinforcements are generally capital intensive, with a negative asset recovery value. When a reinforcement investment is applied at a later stage (i.e. T_1 to T_n) the network will be analysed at that time to determine the optimal reinforcements (e.g. the value of X_1 will be re-evaluated at T_1, instead of meticulously following the plan determined at time T_0).

The evaluation of the later stages of the best reinforcement options is thus mainly done to gain a better estimate of the expected value of the initial options over a longer period of time. The importance of the sequential investments $X_1, ..., X_n$ is reduced due to the time value of money and the uncertainty that exist about future required network strength and reinforcement options. In an LV-network expansion planning approach, it is, therefore, justified to focus on determining the first investment option more accurately. The proposed approach will focus on the first investment X_0 and determine the sequential investments $X_1, ..., X_n$ with a more computationally efficient calculation. For the solution of this bilevel optimisation problem a genetic optimisation approach is applicable [20].

III. PLANNING APPROACH

The approach to solving the network expansion planning problem is discussed in this section. The approach consists of two optimisation loops: an upper-level optimisation to determine the optimal first stage, T_0, set of investments X_0; and a lower-level optimisation to determine the sequential investment options, $X_1, ..., X_n$ at $T_1, ..., T_n$. Based on these sequential investments, the expected cost C_s for each scenario s can be calculated.

A. Upper-level: First investment option

The upper-level optimisation is designed to optimise the first stage investment X_0 over the scenario space $s_1, ..., s_n$. A flow chart of the upper-level optimisation procedure is given in Fig. 1. An overview of the steps in this upper-level optimisation is depicted in this figure; each of the steps will be discussed in more detail in the following subsection.

A.1) Initialisation of GA based on the estimated expected cost per scenario $s_1, ..., s_n$

The network expansion planning approach starts with the estimation of the expected cost for each scenario. An estimation of the cost is necessary to gain insight into the required amount of investments for the entire scenario space as well as
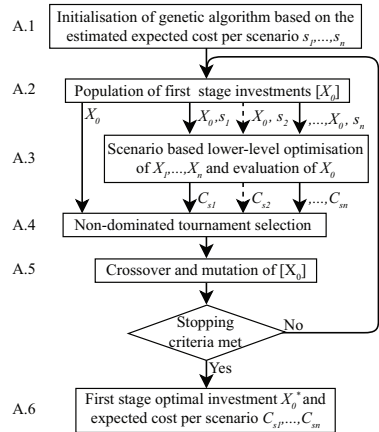

Fig. 1. Flowchart of the upper-level optimisation which determines the optimal first investment X_0 of the multi-stage network planning approach

for the individual scenarios. The use of the deterministically estimated expected cost eliminates the possible drift away from a scenario if the initial population only contains solutions in which the expected cost of a scenario is by far overestimated. The generation of the expected cost for each scenario is based on the network utilisation. The network utilisation k_u is defined as:

$$k_u = \max\left[\frac{|\Delta U_{max}|}{\Delta U_{lim}}, \frac{I_{max}}{I_{lim}}\right] \quad (7)$$

where ΔU_{max} is the maximum voltage deviation, I_{max} is the maximum current encountered for a given scenario and the subscript lim indicates the enforced limits. The network utilisation is calculated for the entire time horizon and the estimated expected cost $C_{s,e}$ for a certain scenario s is calculated as follows:

$$C_{s,e} = \sum_{t=0}^{T_n} [k_u(t) - 1]^+ \times k_u(0) \times C_{grid} \quad (8)$$

where t indicates the time, $^+$ indicates only the positive values, and C_{grid} the investment cost of the original network. The population of first stage investments $[X_0]$ can be generated based on these cost estimates. The amount of selected reinforcement options is estimated by the lowest expected cost over all scenarios $\min_{s_1, ..., s_n} C_{s,e}$. For the initial population, reinforcements are randomly assigned to the population individuals.

A.2) Population of first-stage investments $[X_0]$

The population of first stage investments $[X_0]$ is initialised in Step A.1. The coding of these possible first stage reinforcement investments is based on a multi-chromosome approach. First of all for each feeder within the LV-network a different chromosome will be used, as usually not all feeders within the same network require reinforcement. Next to this,

the reinforcements are encoded over different parts of the chromosome, with the replacement of cables in the beginning of the chromosome and the addition of feeders and substations at the end of the chromosome. The reinforcement options are encoded as a binary vector, where the place in the vector corresponds to a certain reinforcement option and a value of 1 to the implementation of a certain reinforcement [21]. For each of the first stage investments in the population, the coded network reinforcements are applied to the original network.

A.3) Scenario based lower-level optimisation of $X_1, ..., X_n$ and evaluation of X_0

Each of the first investments X_0 in the population need to be evaluated for each scenario in the scenario space $s_1, ..., s_n$. A scenario based lower-level optimisation is used to determine the sequential investments $X_1, ..., X_n$. Based on these and on the initial investment X_0, the expected cost C_s are determined for each scenario s and initial investment option X_0. A detailed explanation of this procedure is given in Section III-B. The result of this step will be the expected cost for each scenario, $C_{s1}, ..., C_{sn}$, for all the initial investment options X_0 in the population.

A.4) Non-dominated tournament selection

With the expected cost for each scenario $C_{s1}, ..., C_{sn}$ known, the best investments X_0 can be selected from the population. This selection is based on a binary tournament selection. First, all first stage investments are ranked according to how many other first stage investments they dominate based on their fitness value. Subsequently, two first stage investments are randomly selected from the population and compared on their dominance rank. The first stage investment which dominates the other is selected as one of the parents for the crossover. A non-elitism approach is taken, with respect to the selection of the next generation, which means the new generation contains only the offspring of the old generation. The best individual X_0 as well as its expected cost $C_{s1}, ..., C_{sn}$ are stored. If in a later iteration of the optimisation a better individual is found, the stored value is replaced.

A.5) Crossover and mutation of X_0

The crossover between the two individuals from the population of first stage investments is based on two types of operators: the exchange of a whole part of the chromosome and a path-based crossover. For the part of the chromosome containing the possible additional feeders and substations, the whole part of the chromosome is exchanged. This can be done, as these types of reinforcements are seen as non-complementary for a single feeder. For the reinforcement of the branches, the generally applied crossover strategies of one/two-point or uniform crossover [22], are not applicable due to the little amount of meaningful topological information which is transferred. To counter this problem, a spanning tree [23] or an edge-set encoding has been proposed in [24]. Both approaches are designed for greenfield network planning, where the crossover could result in unconnected networks.
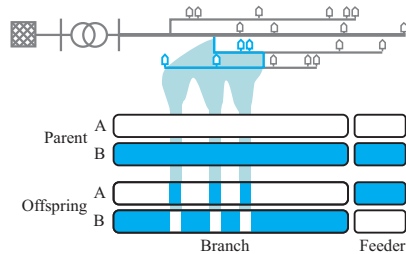

Fig. 2. The crossover process for a single feeder, with the left part of the chromosome the reinforcement of the branches and the right part the additional feeders/substations, the blue path indicates the selected path for the path-based crossover process

For network expansion planning (brownfield) this problem is not present, therefore these approaches are adjusted to a path-based crossover approach. At the beginning of the optimisation, a number of paths within the LV-network are selected. These paths are defined as the shortest path between the endpoint of a branch and the main branch. From these predetermined set of paths, a random path is drawn and all the genes corresponding to this path are interchanged between the parents. This crossover approach is depicted in Fig. 2.

Mutation is applied based on the bit string mutation, i.e. the value of each gene of the chromosome can be mutated from one to zero or vice-versa based on a mutation chance, dependent on the length of the chromosome. The generated new population of first-stage investments $[X_0]$, cycles in the loop starting from Step A.2 unless the stopping criterion is met.

A.6) First stage optimal investment X_0^* and expected cost per scenario $C_{s1}, ..., C_{sn}$

After the stopping criterion has been met, the stored value of the best first stage investment X_0^* is considered to be the optimal investment to undertake at the current time T_0.

B. Lower-level: Scenario based evaluation of X_0

The lower-level optimisation, which is Step A.3 in the upper-level optimisation loop, is designed to determine the lower bound of the expected cost for a scenario C_s given the initial investment X_0. An overview of the scenario-based lower-level optimisation and evaluation procedure is given in Fig. 3. The steps in this figure are explained in more detail in this subsection.

B.1) Generate population of sequential investments $[X_1, ..., X_n]$

With the initial investment X_0 already implemented in the network, a new population consisting of other possible investment options will be constructed. This population will contain the investments after the first stage $X_1, ..., X_n$. The number of sequential reinforcements is based on the estimated expected cost $C_{s,e}$ for the scenario, as well as that of the first
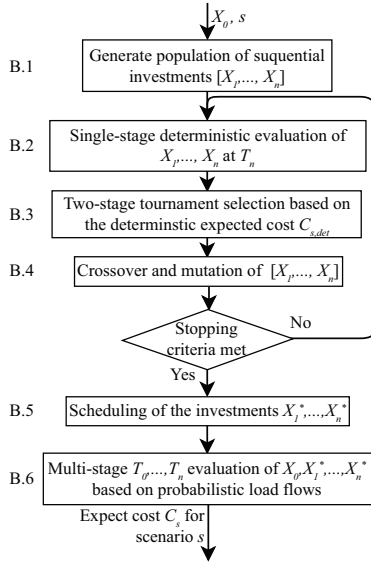
Fig. 3. Flowchart of the lower-level optimisation to calculate the sequential investment options $X_1, ..., X_n$ and the expected cost C_s for a scenario s

stage investment. The initialisation is done based on randomly selected reinforcements.

B.2) Single-stage deterministic evaluation of $X_1, ..., X_n$ at T_n

Each individual of the sets of investments $[X_1, ..., X_n]$ will be analysed in a single-stage approach. The single-stage assumes the network will be reinforced at T_0 and should be able to handle the load at T_n. In this single-stage evaluation of the network cost, only the investment cost is taken into account when calculating the expected cost. The other costs due to energy losses and outages are assumed to have a small difference between the different sequential reinforcement options and are therefore neglected. These single stage investment costs $C_{s,det}$ for each scenario s are calculated as follows $C_{s,det} = \sum_{i=1}^{n} X_i$. The evaluation of the network is based on two deterministic load flow calculations, one during the middle of the day to capture the PV peak and one during the conventional (evening) peak hours. All the households are assumed to have identical load curves equal to the average load of a large number (> 200) of households. Based on the two load flow calculations it is determined whether the network is still operating within its voltage and current limits at T_n. If a violation of the limits occurs, the solution is penalised with an additional investment cost equal to $C_{grid} \times k_u(0)$ to ensure this solution has higher investment cost than all feasible solutions.

B.3) Two-stage tournament selection based on expected cost C_s

When the single stage investment costs $C_{s,det}$ are determined for each of the sequential investments $X_1, ..., X_n$ in the population, the best alternatives are selected from the population. The selection of the parents of this generation is done based on a two-stage tournament. A two-stage tournament selection has a stronger selection pressure as the best parent is selected from four individuals. The stronger selection pressure is necessary to gain a quick estimate of the optimal solution. Two solutions are selected from the whole generation, the best of the two is used as a parent for the creation of the new generation. Next to this, the alternatives with the lowest $C_{s,det}$ are automatically passed on to the new generation to ensure they will not get lost due to crossover or mutation.

B.4) Crossover and mutation of $X_1, ..., X_n$

From each pair of parents determined in Step B.3 two offspring are generated based on two types of crossover. For the chromosomes with additional feeders and substations, the complete part of the chromosome can be exchanged while for the branches it is path-based, similarly to the mutation and crossover as discussed in Step A.5. Steps B.2-B.4 are repeated until the stopping criteria are met. The best sequential single stage investments $X_1^*, ..., X_n^*$ have been selected by this lower-level optimisation. To be able to evaluate the solution $X_1^*, ..., X_n^*$ at each of the time intervals $T_0, ..., T_n$ these investments will need to be scheduled.

B.5) Scheduling of the investments $X_1, ..., X_n$

To determine the best investment sequence, the optimal reinforcement options for the T_{n-1} stage are determined by calculating the expected cost $C_{s,det}$ with the removal of each of the available reinforcement options. This procedure is repeated for time T_{n-2} and so on until all investment options have been placed in time. The sequence of the investments is amended with X_0 to generate the final set of investments, that is used in the evaluation. Then $X_0, X_1^*, ..., X_n^*$ is stored for the evaluation of the complete solution.

B.6) Multi-stage $T_0, ..., T_n$ evaluation of $X_0, X_1^*, ..., X_n^*$ based on probabilistic load flows

To preserve computation efficiency, the evaluation of the solutions has so far only been done based on deterministic peak load/generation values which were assumed equal for all households and only based on the investment costs. In order to provide a more extensive evaluation of the investment options, a detailed evaluation of each solution is performed in this stage. The evaluation of the investments $X_0, X_1^*, ..., X_n^*$ is done by applying a probabilistic approach instead of a deterministic one. The net household loads are characterised as Gaussian mixture distributions of the peak load as well as the PV-peak load [25]. The application of a probabilistic load flow results in probability density functions for the node voltages and branch currents. This allows for the application of a penalty function with a value equal to the value of deferral

TABLE I
CHARACTERISTICS OF THE LV-NETWORK USED IN THE CASE STUDY

Feeder	Length [m]	Connections	Mean impedance [$\Omega\backslash km$]
Orange	340	41	0.335
Green	457	62	0.467
Purple	564	57	0.393
Light blue	371	45	0.569
Red	436	55	0.280
Blue	241	41	0.310

TABLE II
THE FOUR SCENARIOS USED IN THE CASE STUDY

Scenario	Load growth	EV	PV
Low	0%/year	$\dfrac{4.5}{1+e^{-0.11(T_i-T_0)}}$	$\dfrac{7.2}{1+e^{-0.19(T_i-T_0)}}$
High EV	1%/year	$\dfrac{63}{1+e^{-0.09(T_i-T_0)}}$	$\dfrac{7.2}{1+e^{-0.19(T_i-T_0)}}$
High PV	1%/year	$\dfrac{4.5}{1+e^{-0.11(T_i-T_0)}}$	$\dfrac{82}{1+e^{-0.16(T_i-T_0)}}$
High	1%/year	$\dfrac{63}{1+e^{-0.09(T_i-T_0)}}$	$\dfrac{82}{1+e^{-0.16(T_i-T_0)}}$

of the next investment X_{i+1} over length of the current time step $T_i - T_{i-1}$: $C_{s,pen}(T_i) = X_{i+1}/T_i - T_{i-1}$. This penalty, $C_{s,pen}$, will be multiplied with the chance of violating the voltage or current limits obtained from the probabilistic load flow calculations. When the chance of violating the limits becomes equal to the chance of violating the limits at T_0, without the realisation of X_0, the next investment option is implemented instead of applying the penalty function. The cost of outages and the cost of transportation losses within the network are also evaluated at each time step $T_0, ..., T_n$. This will yield an expected cost for a certain scenario based on the first stage investment X_0.

IV. RESULTS

To determine whether the proposed approach to solving the LV-network expansion planning problem is feasible, the approach is applied to a case study. For the case study a residential network from the Dutch DNO Liander has been selected. An overview of the network is given in Fig. 4 and the main characteristics of the network are given in Table I. The network consists of 6 feeders connecting 301 customers and fed by a 400 kVA transformer. The loads in the network are generated by using a bottom-up Markov Chain Monte Carlo approach as described in our previous work [26]. To characterise the long-term evolution of the load, four scenarios are used. The scenarios contain a load growth rate and increasing penetration rates of PV and EV, selected as main drivers for reinforcement in the LV networks. The four scenarios are defined in Table II.

In order to perform the network expansion planning optimisation, a number of parameters still need to be defined. In Table III the required constants for the network planning approach are given.

With these parameters defined, the network expansion can be carried out. To demonstrate the performance of the network planning approach, four different results are discussed. First of all the convergence of the GA is discussed, followed by
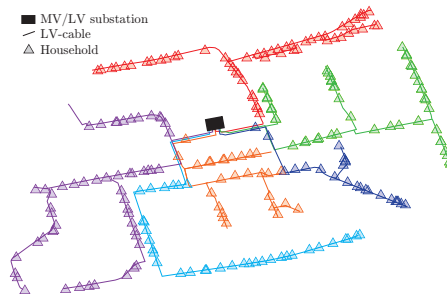

Fig. 4. Geographic overview of the LV-network used for the analysis

TABLE III
DATA USED FOR THE NETWORK EXPANSION PLANNING OPTIMISATION

Symbol	Component	value
C_{cbl}	LV-Cable [€\mtr]	65.5
$C_{cbl,mv}$	MV-Cable [€\mtr]	77
C_{fdr}	LV-busbar expansion [€]	4200
p_e	Electricity price [€\MWh]	0.73
p_{av}	Outage price [€\min]	0.05
C_{sub}	Additional MV/LV substation [€]	40000
$C_{inv,fix}$	Fixed investment cost [€]	8300
λ_{cbl}	Cable failure rate [\(yr km)]	0.038
Tr_{cbl}	Repair time [min]	200

a closer look at the convergence of the initial investments. The performance of the lower-level optimisation is discussed hereafter and the section is concluded with a look at the convergence of a bi-level versus an integrated optimisation approach.

A. Bi-level optimisation

The optimisation approach is applied to the network depicted in Fig. 4 with the scenarios as described in Table I. For each generation of the GA, the minimum expected value and average expected value of the population is logged. As the GA has a stochastic nature, the algorithm has been run for 20 times. The results are shown in Fig. 5, with the thin lines indicating a single run and the thick lines indicating the average of the 20 runs.

From the figure, it can be seen that the best solution converges steadily to a value of €280 000 after 14 generations. The expected cost of the average solution shows a decrease for the first 14 generations hereafter the expected value of the average solution is increasing. This is because the population moves closer to the optimal network and the unused capacity of the network decreases. For these networks a small change will more often result in a network which is no longer strong enough, leading to far higher cost for the network.

B. Initial investments

As the initial investment is the only investment which will be directly implemented, it is necessary to look at the convergence of this investment in more detail. In Fig. 6 the
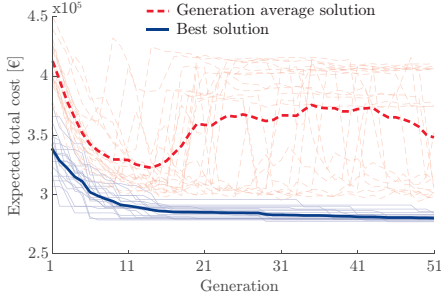
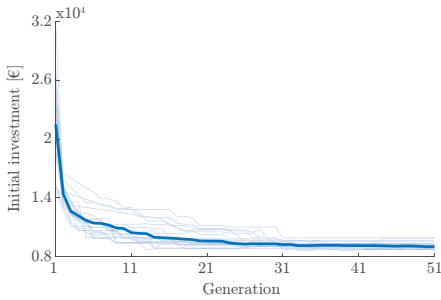Fig. 5. Convergence of the average and minimum expected cost solution

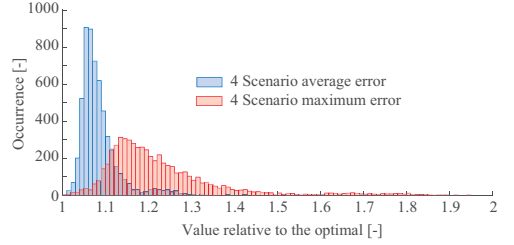Fig. 6. Convergence of the initial investment

Fig. 7. The performance of the optimisation of the investments after the initial investment

Fig. 8. The computational performance of the bilevel optimisation versus an integrated approach

optimal initial investment for consecutive generations has been given.

From the figure, it can be seen that most of the initial investments converge to a certain value within 15 iterations. For four runs it takes longer for the initial investment to reach more or less ($\pm 6\%$) the same value. When comparing these results to the results of the expected value as given in Fig. 5, it can be seen that even though the expected value already seems to have converged, the initial investments can still change. There are a number of solutions within a very narrow range of the expected value, but with larger differences when it comes to the initial investments, which causes the change in initial investments between generations 16 and 31, as can be seen in Fig. 6.

C. Performance lower-level optimisation

The lower-level optimisation of the future investments after the initial investment has been decided is performed for a limited set of evaluations. To determine whether this will lead to acceptable values for the subsequent investment, the whole optimisation approach (Steps A.1-A.6) is applied without the initial investment, i.e., set to zero, $X_0 = \varnothing$. By keeping the initial investment constant, the performance of the bilevel optimisation can be tested with the application of the probabilistic evaluation of the upper-level GA as described in Step B.6.

By removing the initial investments the number of required investment options is highest, so the expected error should also be highest. The performance of the lower-level GA is given in Fig. 7, where both the average expected value when analysing four scenarios and the maximum error of these four scenarios are plotted relative to the optimal value for a very large number of runs of the lower-level optimisation.

From the figure, it can be seen that in 95% of the cases the error introduced by the simple (single stage, backwards induction and deterministic) evaluation and the limited number of evaluations is limited to 15% when taking the average expected value over the four scenarios and 35% for the maximum error of the four scenarios. Especially, if multiple scenarios are used, the difference between multiple computations of the optimal subsequent investments, $X_1^*, ..., X_n^*$, is small, indicating that the lower-level optimisation gives a good approximation of the true optimal subsequent investments.

D. Bilevel versus an integrated optimisation

To gain insight into whether the application of a bilevel optimisation approach is beneficial from a computational point of view without much loss of accuracy, the bilevel approach is compared to an integrated approach for the LV-network expansion planning. The convergence of the bilevel optimisation versus that of an integrated approach is plotted in Fig. 8.

Fig. 8 shows that both algorithms converge to the same level of the expected value. The bilevel optimisation converges

faster compared to the computation time. The number of required generations to yield convergence is much lower for the bilevel optimisation case, and this can already be seen from the initial starting point of both graphs. The starting point indicates how long the evaluation of a single generation takes, with the bilevel optimisation approach taking more than three times longer to evaluate a single generation.

V. CONCLUSION

A bilevel genetic optimisation approach with a focus on the initial investment has been formulated. The initial investment is optimised in an upper-level optimisation procedure, where a probabilistic load flow is used to evaluate the volatile load of individual households. The subsequent investments are determined in the lower-level optimisation procedure. The uncertainty of the load is implemented by using a scenario based stochastic optimisation approach in the lower-level optimisation.

The application of this approach to a case study taken from a Dutch grid operator has shown to be a feasible alternative to a single level optimisation approach. Though a more coarse optimisation approach is applied, the results of the lower-level optimisation are still acceptably accurate. The paper also shows that the bilevel optimisation converges much faster than an integrated approach, while the expected monetary value only shows a very slight difference. Concluding, the proposed bilevel approach can obtain similar accuracy to an integrated approach, at a lower computational cost.

REFERENCES

[1] M. Nijhuis, M. Gibescu, and J. Cobben, "Assessment of the impacts of the renewable energy and ICT driven energy transition on distribution networks," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1003–1014, aug 2015.

[2] A. M. Cossi, R. Romero, and J. R. S. Mantovani, "Planning and projects of secondary electric power distribution systems," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1599–1608, 2009.

[3] J. M. Nahman and D. M. Peric, "Optimal planning of radial distribution networks by simulated annealing technique," *Ieee Transactions on Power Systems*, vol. 23, no. 2, pp. 790–795, 2008.

[4] A. M. Cossi, L. da Silva, R. Lazaro, and J. Mantovani, "Primary power distribution systems planning taking into account reliability, operation and expansion costs," *IET Generation, Transmission & Distribution*, vol. 6, no. 3, p. 274, 2012.

[5] S. Ganguly, N. C. Sahoo, and D. Das, "A novel multi-objective PSO for electrical distribution system planning incorporating distributed generation," *Energy Systems*, vol. 1, no. 3, pp. 291–337, 2010.

[6] A. M. El-Zonkoly, "Multistage expansion planning for distribution networks including unit commitment," *IET Generation, Transmission & Distribution*, vol. 7, no. 7, pp. 766–778, jan 2013.

[7] V. Miranda, J. V. Ranito, and L. M. Proena, "Genetic algorithms in optimal multistage distribution network planning," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1927–1933, jan 1994.

[8] C. Gamarra and J. M. Guerrero, "Computational optimization techniques applied to microgrids planning: A review," *Renewable and Sustainable Energy Reviews*, vol. 48, pp. 413–424, 2015.

[9] R. Shi, C. Cui, K. Su, and Z. Zain, "Comparison study of two meta-heuristic algorithms with their applications to distributed generation planning," in *Energy Procedia*, vol. 12, 2011, pp. 245–252.

[10] P. Reche-López, N. Ruiz-Reyes, S. García Galán, and F. Jurado, "Comparison of metaheuristic techniques to determine optimal placement of biomass power plants," *Energy Conversion and Management*, vol. 50, no. 8, pp. 2020–2028, 2009.

[11] S. Kannan, S. M. R. Slochanal, and N. P. Padhy, "Application and comparison of metaheuristic techniques to generation expansion planning problem," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 466–475, 2005.

[12] S. Heidari, M. Fotuhi-Firuzabad, and S. Kazemi, "Power Distribution Network Expansion Planning Considering Distribution Automation," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1261–1269, may 2015.

[13] J. Salehi and M.-R. Haghifam, "Long term distribution network planning considering urbanity uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 321–333, nov 2012.

[14] M. Rahmani, R. Romero, and M. J. Rider, "Strategies to Reduce the Number of Variables and the Combinatorial Search Space of the Multistage Transmission Expansion Planning Problem," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2164–2173, aug 2013.

[15] C. L. T. Borges and V. F. Martins, "Multistage expansion planning for active distribution networks under demand and Distributed Generation uncertainties," *International Journal of Electrical Power and Energy Systems*, vol. 36, no. 1, pp. 107–116, jan 2012.

[16] E. Carrano, R. Cardoso, R. Takahashi, C. Fonseca, and O. Neto, "Power distribution network expansion scheduling using dynamic programming genetic algorithm," *IET Generation, Transmission & Distribution*, vol. 2, no. 3, p. 444, 2008.

[17] S. S. Al Kaabi, H. H. Zeineldin, and V. Khadkikar, "Planning active distribution networks considering multi-DG configurations," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 785–793, mar 2014.

[18] H. Falaghi, C. Singh, M.-R. R. Haghifam, and M. Ramezani, "DG integrated multistage distribution system expansion planning," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 8, pp. 1489–1497, oct 2011.

[19] G. P. Harrison, A. Piccolo, P. Siano, and A. R. Wallace, "Hybrid GA and OPF evaluation of network capacity for distributed generation connections," *Electric Power Systems Research*, vol. 78, no. 3, pp. 392–398, mar 2008.

[20] Y. Yin, "Genetic-Algorithms-Based Approach for Bilevel Programming Models," *Journal of Transportation Engineering*, vol. 126, no. 2, pp. 115–120, mar 2000.

[21] S. Najafi, S. Hosseinian, M. Abedi, A. Vahidnia, and S. Abachezadeh, "A Framework for Optimal Planning in Large Distribution Networks," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1019–1028, may 2009.

[22] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, jun 1994.

[23] P. Carvalho, L. Ferreira, and L. Barruncho, "On spanning-tree recombination in evolutionary large-scale network problems - application to electrical distribution planning," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 6, pp. 623–630, 2001.

[24] F. Rivas-Davalos and M. Irving, "The Edge-set Encoding in Evolutionary Algorithms for Power Distribution Network Planning Problem Part I: Single-objective Optimization Planning," in *Electronics, Robotics and Automotive Mechanics Conference (CERMA'06)*, vol. 1. IEEE, sep 2006, pp. 203–208.

[25] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Gaussian Mixture Based Probabilistic Load Flow for LV-Network Planning," *IEEE Transactions on Power Systems*, pp. 1–1, 2016.

[26] M. Nijhuis, M. Gibescu, and J. Cobben, "Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data," *Energy and Buildings*, vol. 112, pp. 121–129, jan 2016.

Demand Response: Social Welfare Maximization in an Unbundled Energy Market Case Study for the Low-Voltage Networks of a Distribution Network Operator in The Netherlands

Michiel Nijhuis, *Member, IEEE*, Muhammad Babar, *Member, IEEE*,
Madeleine Gibescu, *Member, IEEE*, and Sjef Cobben

Abstract—With the introduction of smart meters, dynamic pricing, and home energy management systems, residential customers are able to react to changes in electricity prices. In an unbundled market, the energy supplier and the network operator may have conflicting interests with respect to demand response (DR) programs. As customer participation is essential to a well-functioning DR program, it is needed to assess which DR programs offer most customer benefits. Two DR program options are analyzed for low-voltage feeders: a program from the energy supplier based on the electricity price, and a DR program from the network operator based on the loading of the network. Depending on the network topology the benefits can change significantly between the two DR programs. DR from an energy supplier point of view might induce undervoltages which lead to additional network reinforcements, while load shifting from a network point of view can generate higher electricity cost.

Index Terms—Demand response (DR), demand side management, distribution networks, energy management, load management, peak shaving, power system operation, smart grids.

I. INTRODUCTION

THE energy transition will significantly alter the residential electricity usage. Local generation becomes possible with the introduction of rooftop photovoltaic (PV) systems. In addition, the household (HH) electricity consumption is expected to rise significantly through the electrification of heating and transportation. With these developments, the low-voltage (LV) network requires strengthening. Technology advancements in information and communications technology (ICT) have resulted in smart metering infrastructures and new possibilities to control HH appliances through the use of home energy management systems [2]. This will enable retail customers to alter their energy use based on real-time price signals and can result in monetary rewards when demand response (DR) programs are introduced [3] or when local renewable energy generation is fed back into the network [4].

For HH consumers in the Netherlands, the electricity prices are regulated with a constant or two-tier time of use (low price from 11 P.M. until 7 A.M. and normal price from 7 A.M. until 11 P.M.) electricity price and a fixed network tariff. To assess the benefits of DR, and the shift to more time-varying pricing structures, multiple DR pilot projects have been started. The *Jouw Energie Moment* [5] pilot is operated by a distribution network operator (DNO) to minimize network loading, while the *PowerMatching City* [6] pilot is evaluating a local market-based DR program. These two pilot projects can, however, highlight the conflicting interests between actors in the unbundled electricity market. A price-based DR scheme can introduce higher peak loads as users increase their loading during periods of low prices, while peak shifting based on network constraints could shift demand to times of higher energy prices, however reducing the investments necessary in the distribution network.

Therefore, it is necessary to assess which approach to the creation of the DR program could secure more social welfare and under which conditions. In order to assess the social welfare, first a small overview is given of the electricity market and distribution network in the Netherlands. The modeling of the network based on generic LV feeders, the Markov chain Monte Carlo modeling of the HH load, the DR modeling through price elasticity, and the creation of the DR incentive functions are discussed subsequently. Thereafter, the simulations and results are discussed.

II. DUTCH ELECTRICITY SYSTEM

An initial evaluation of DR programs is required to benchmark the social welfare of the electricity supply and demand in the Dutch electricity system. For this evaluation, it is important to establish the market and network structure present in the Netherlands. Thus, this section first introduces the current environment of the Dutch electricity market, and second explains the main characteristics of the Dutch distribution network.
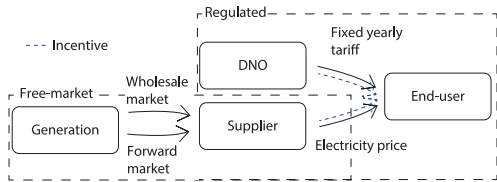
0093-9994 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
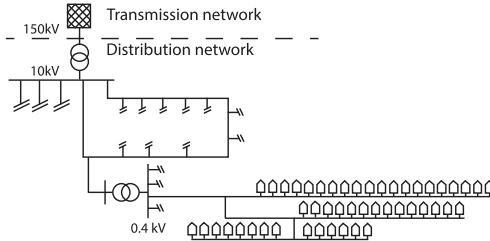
Fig. 1. Overview of the market structure.

Fig. 2. Overview of a typical distribution network layout in the Netherlands.

A. Electricity Market

The Dutch electricity supply chain is unbundled, meaning that a market participant either can be involved in the transmission and distribution of electricity, or in the generation and retail market (selling electricity to end-user consumers) and cannot be active in both. However, due to the liberalization of the electricity market, the vertical integration of generation and supply to end users is no longer a necessity. The transmission and DNOs operate in a regulated monopoly and charge a yearly fixed network tariff for HH consumers, independent of the amount of electricity used. For this research, an energy supplier (ES) is defined as: A market participant which buys electricity from generators through the wholesale market and/or forwards and/or has its own generation portfolio and sells electricity to the end users. In the current situation, the ES can offer HH consumers only a fixed or a two-tier electricity price. A small overview of the market structure is shown in Fig. 1.

For the inclusion of DR in this market structure, the current market framework can be left unchanged. The DNO and the ES gain the possibility of supplying consumers with an incentive (positive and negative) on top of their current electricity tariff/price. In this way, the electricity price and network tariff become variable prices on which the consumer can react. In principal, the overall cost for the consumer will be lower if he alters his load based on the variable price. The DNO is able to supply this incentive to the consumer as the use of the DR allows for investment deferral of their assets. The ES is capable of providing the incentive because it can buy the electricity at lower prices from the wholesale market, or/and reduce the amount of electricity bought during periods of high prices.

B. Distribution Network

An overview of the typical structure of the distribution network in the Netherlands is shown in Fig. 2; this shows the different voltage levels too. The medium-voltage (MV) level of 10 kV generally has a ring structure with multiple MV/LV transformers connected. The LV network has radial structure. The feeders consist of underground cables and can have up to 100 connected HHs per feeder.

The electrification of transportation or the increased penetration of PV will generate overloading and voltage limit violations in the LV network. In almost all of the analyzed LV feeders, the voltage limits are reached before the overloading limits of the feeder; thus, only the voltage limits are taken into account for this research. The DNO has to provide the required network

capacity to the HHs. This is usually achieved by reinforcing the cables of the LV network. With the use of the DR program, the DNO can promote peak shifting at the HH level which reduces the network loading and ensures the voltage limits are not violated.

III. MODELING

The information about the market and power system conditions in which the DR program will be analyzed has been established. A more detailed explanation on how these conditions are modeled is discussed in this section. The electricity prices that the customers experience consist mainly of two components which can be influenced by DR: the network (transportation) cost and the energy (commodity) cost. To assess the influence of the DR program on these components, both have to be estimated. The calculation of the energy price is done by multiplying the HH load with the real-time electricity price. The network costs are based on the current connection tariffs plus an estimate of the replacement cost of overloaded network assets.

To calculate these components, three inputs are needed: a real-time electricity price, the topology of the LV feeders to which the customers are connected, and the load curves of the customers. The real-time electricity price is constructed by utilizing the electricity prices of the German intraday market (15 min periods with the close 1 h ahead) [7] in combination with weather variables [8]. A Bayesian density-based forecast method for electricity prices [9] is utilized for the construction of the price signal.

A. LV Network

For the assessment of the DR programs, LV feeders are used. The subset LV feeders which are used come from a set of generic networks constructed from the actual networks of the Dutch DNO Liander by the means of a clustering algorithm [10]. From the output of the clustering, a number of networks are chosen for the assessment of the effects of the DR programs. The chosen networks should not be too weak, as reinforcements are definitely needed despite the use of the DR program, or too strong, as the introduction of the DR program has no significant effect on the loading of the feeder. In these cases, the ES-driven DR program would always be better, as the DNO cannot offer any incentive if there is no possibility of investment deferral.

TABLE I
OVERVIEW OF THE CHARACTERISTICS OF THE LV FEEDERS USED IN THE
SIMULATIONS

	Length	HH/non-HH	$\overline{R}$ [mΩ/km]	Branches	$\overline{R_{POC}}$ [mΩ]
A	310 m	92/7	300	1	43
B	275 m	28/1	620	2	94
C	520 m	41/1	430	9	105
D	500 m	61/1	370	4	75

Four feeders are chosen, which illustrate the possible effects of DR on the LV network. These feeders represent a set of relatively weaker networks which correspond to feeders to which 27% of all the LV customers in the network of Liander are connected. The rest of the customers are connected to stronger or weaker networks and will not be considered in this analysis. An overview of the main characteristics of the four feeders which are used in the analysis is presented in Table I.

For the feeders, the total feeder length, the number of HH and non-HH connections, the average cable resistance per kilometer ($\overline{R}$), the number of branches within a feeder, and the average resistance between the beginning of the feeder and each point of connection ($\overline{R_{POC}}$) are shown. These characteristics give an indication of the loading, the distribution of the loads, and the thickness of the cables used within the feeder. Feeder A represents a recently built feeder with relatively short distances and a high number of connected customers, i.e., a feeder in a city center. Feeder B represents an old feeder with a modest amount of customers, typical of a feeder in the suburbs. Feeder C represents an old feeder which has been expanded in the different directions in the last decades and feeder D represents a more or less average feeder typical of a Dutch DNO.

B. Household Load Profiles

The HH load profiles are modeled by using a Markov chain Monte Carlo approach based on time-of-use surveys, statistics on appliance characteristics, and weather variables [11]. Each appliance is modeled separately on a 15-min interval basis to be able to assess the amount of flexible load available in this interval. The accuracy of the model is verified using smart meter data from Dutch HHs. As the energy transition will change the loading compared to the current situation, the model of the HH load is augmented with an increased penetration of PV (60% of the HHs) and electric vehicles (EV) (80% of the HHs). An overview of the average loading of the HH is shown in Fig. 3.

The figure shows the breakdown of the energy use based on the load: EV, PV, base load, washing loads (dishwasher, washing machine, and dryer), and the other loads. The EV's (when present at the HH) and the washing/drying loads are assumed to be flexible [12]. From the figure, it becomes apparent that most flexible loads are available during the evening. The washing load can be shifted for up to 8 h, while the EV load should be fully charged when the EV leaves. The HHs connected to the feeder, all have the same profile to eliminate the effects of the distribution of HH loads over the feeder.
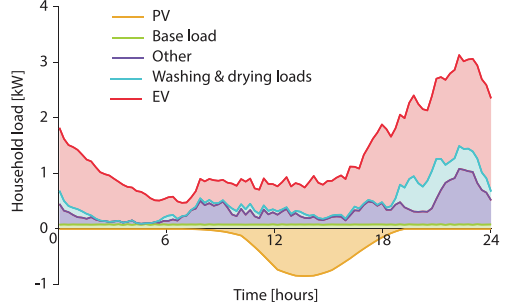

Fig. 3. Cumulative load of different household load classes for a single winter day.

C. Demand Response Model

To model the DR, the relationship between the price and the amount of electricity consumed is utilized. This is done via the (per unit) slope of the demand curve, which is usually referred as price elasticity of demand or demand elasticity [13], which can be expressed as follows:

$$\varepsilon = \frac{\frac{\partial D}{d_0}}{\frac{\partial P}{p_0}} \tag{1}$$

where ε is the elasticity coefficient which indicates the level of DR for a commodity which results from a change in price of that commodity, D is the demand, d_0 is the initial demand, P is the price, and p_0 is the initial price. This general economic concept can also be applied to the electricity price and consumption. In this way, the concept of demand elasticity can be used to model the change in electricity demand, when a change in electricity price occurs due to the incentive-based DR programs [14]. This is modeled through the following equation:

$$d(i) = d_0(i)\left\{1 + \sum_{j=1}^{k} \varepsilon_{ij}\left(\frac{p(i) - p_0(i) + \alpha(i)}{\frac{p(i) + p_0(i)}{2}}\right)\right\}$$

$$\forall(i,j) : i \triangleq \{1, \ldots, k\} \wedge j \triangleq \{1, \ldots, k\} \tag{2}$$

where $d(i)$ is the new customer's demand at time instance i, $p(i)$ is the spot price, $d_0(i)$ is the initial demand value, $p_0(i)$ is the initial price, $\alpha(i)$ is an incentive payment during the ith interval, and j is the time instance running from $i - 8$ until $i + 8$. ε_{ij} is the demand elasticity matrix, which can be divided into two regions: when $i = j$, ε_{ij} is called self-elasticity (percentage of load the consumer can shift) and when $i \neq j$, ε_{ij} is called the cross elasticity (effect of changes in the flexibility due to the DR in previous and future intervals). The self-elasticity of consumer can be represented using a Gaussian distribution [15]. Based on the electricity consumption of a Dutch HH, it is assumed that the aggregate self-elasticity of HHs has a Gaussian distribution with $\mu = -0.3$ and $\sigma = 0.11$. This means that on average 30% of the flexible loads which are present at the HH in this moment will react to the DR-induced price changes. The

cross elasticity values can be split into two categories: the postponing cross elasticity, when $j < i$, and the advancing cross elasticity, when $j > i$. The postponing cross elasticity represents the effects from actions in the previous time intervals, i.e., if the HH reduced its demand in the previous interval, it is less likely it will still be able to reduce the demand in the current interval as well. From data on Dutch HHs, it is observed that the postponing cross elasticity can be expressed as a Weibull distribution, i.e., $W(8, 1, 1)$, if the past eight intervals are considered. In this paper, it is assumed that the advancing cross elasticity is given by the same Weibull distribution as the postponing cross elasticity. The cross elasticity is calculated for 16 intervals across the ith interval, i.e., $[(i-8) \ldots i \ldots (i+8)]$, where $[(i-8) \ldots (i-1)]$ are the stored self-elasticity values and $[(i+1) \ldots (i+8)]$ are calculated by using a Weibull distribution, i.e., $W(k; \lambda, \kappa) \rightarrow W([(i+1) \ldots (i+8)]; 1, 1)$.

D. Incentive Functions

The incentive function of the DR program is either based on the energy price or the network loading. For the ES-based DR program, the price plus incentive is simply replaced by a real-time energy price. To avoid continuous load changes on even the smallest of price changes, a dead band is introduced in the price signal. If there are price deviations of less than 0.02€ from the average price, no price signal is received by the HHs.

For the DNO-based DR program, a more complex incentive function needs to be defined. The incentive function should translate the energy consumption of the HH to an incentive based on the impact on the network of this energy consumption. The DR approach is based on the loading of the HHs as this information can be made available through the use of smart meters. These loading conditions are converted to a signal which indicates the relative amount of DR required to keep the voltages within the limits; based on the reinforcement cost of the feeder, this signal is converted to an incentive for the customer. As the voltage limits should not be violated, the incentive requires a rise when the voltage deviations get closer to the limits. If the customer reacts on the incentive from the DNO, the network will be restored to a position well within its limits. This reduces the incentive from the DNO and will then increase the loading of the HHs, causing the incentive to increase again. To avoid this oscillation, the incentive from the DNO is only reduced after a time lag of 15 min and the rate in which the incentive can change is limited. The resulting incentive functions for the ES and the DNO-based DR programs are given in the following equations:

$$\alpha(i)_p = \begin{cases} \overline{p_E} - p_E(i) - 0.02, & \text{if } |\overline{p_E} - p_E(i)| \geq 0.02 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$\alpha(i)_n = \begin{cases} \frac{(|1-V(i)|-0.04)^3}{T_{DR}} \frac{(C_{rf} - C_{rf} \cdot d)}{C_{lf}}, & \text{if } |V(i) - 1| \geq 0.04 \\ 0, & \text{otherwise} \end{cases}$$

$$(4)$$

where $\alpha(i)_p$ is the incentive the HHs get from the ES, $\alpha(i)_n$ is the incentive from the DNO, p_E is the real time electricity price, $\overline{p_E}$ is the yearly average electricity price, T_{DR} is a factor to ensure
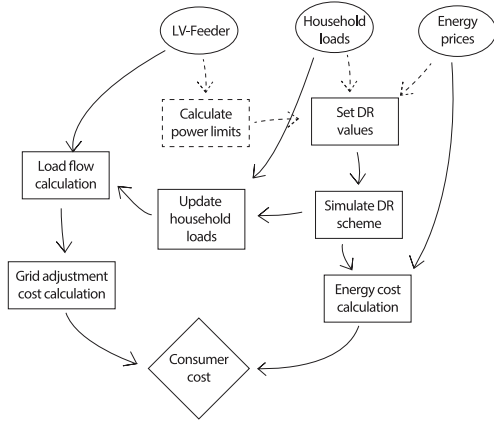
Fig. 4.   Flow chart of the assessment process of the customer electricity cost for both demand response programs.

the incentives supplied by the DNO do not exceed the profits from the investment deferral ($T_{DR} = \sum_{i=0}^{i_Y} |1 - V(i)| - 0.04$), i_Y is the number of 15 min intervals in a year, C_{rf} are the reinforcement cost, and C_{lf} is a factor (between 1.02 and 1.05 depending on the network topology) to decrease the investment deferral with the expected reduction in losses when reinforcing the network. Examples of the incentive signals which the customer receives can be seen in Fig. 5(b).

Fig. 4 shows a schematic overview of the procedure used to calculate the yearly energy cost for customers for the scenarios analyzed.

The electricity price for the customer is calculated based on the real-time price signal for one winter and one summer week multiplied with the HH loading for these weeks and extrapolated to the yearly energy cost. The network costs are based on the current network cost plus the cost of network reinforcements if the network is operating outside of the limits for any interval in the summer or winter week. The reinforcement costs are based on the replacement cost (available from the data of the Dutch DNO Liander) of the cables and transformers in the LV network.

IV. SIMULATION

The simulations of the ES- and DNO-driven DR programs are done based on the four selected feeders, for one summer and one winter week, to include effects of seasonal variation in load as well as in electricity price. An example of the incentive functions, the resulting load, and the effect on the voltage deviations is shown in Fig. 5 for feeder D.

Fig. 5(a) shows the load of an average HH and how this is affected by the DR programs. It can be seen that the loads are mainly shifted from the evening peaks to the night. The energy price incentive shows a much smoother profile compared to the incentive from the DNO, as can be seen in Fig. 5(b). This is caused by the direct feedback loop in the incentive from the

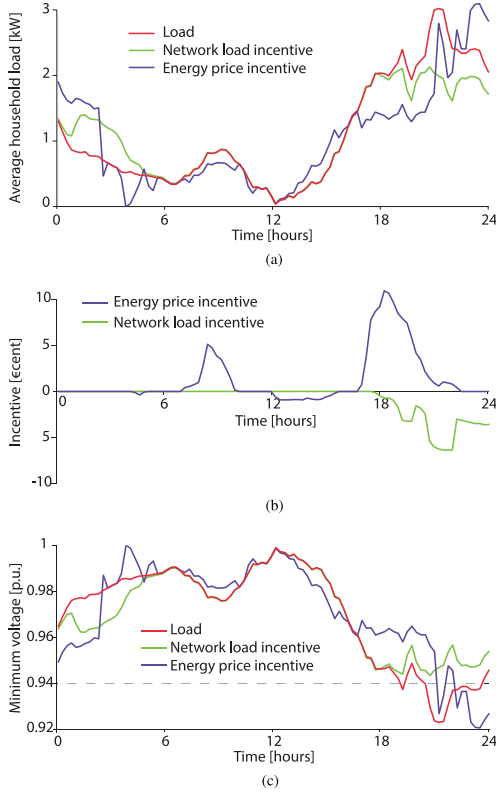Fig. 5. Loading under the two DR programs and the incentive functions used to simulate the DR during a winter weekday for feeder D. (a) Network load, (b) incentive function, and (c) minimum voltage.

DNO, while for the incentive from the ES, it is assumed that the changes in this LV network are too small to have an effect on the price. The voltage profile in Fig. 5(c) shows that both the normal HH load and the HH load with DR based on the ES violate the lower voltage limit of 0.94 p.u. (the actual voltage regulations allow larger voltage deviations, but as the voltage deviations in the MV network are not modeled, the 0.94 limit is used to account for this). The DR from the DNO point of view reduces the voltage deviations. It can also be seen that the DNO will only start to give an incentive if the voltage goes below the 0.96 p.u. mark.

V. RESULTS

The results from the simulations are assessed on the yearly energy cost. To examine whether the results of the simulations of the DR programs are still valid when the price levels change, a sensitivity analysis is performed.
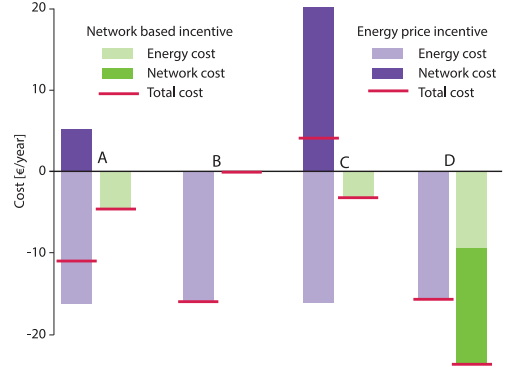
Fig. 6. Overview of the yearly energy cost per household for two DR programs with respect to the business as usual situation for the feeders A–D.

A. Cost

The evaluation of the two DR programs is done on their effect on the yearly energy cost. The yearly energy cost is calculated using the following formula:

$$C_{\text{EY}} = C_{\text{fixed}} + \frac{(C_{rf} - C_{rf} \cdot d)}{C_{lf}} + 26 \sum_{i=0}^{i_W} P_W(i) \cdot p_{\text{EW}}(i)$$

$$+ 26 \sum_{i=0}^{i_W} P_S(i) \cdot p_{\text{ES}}(i) \qquad (5)$$

where C_{EY} is the yearly energy cost, C_{fixed} the fixed part of the energy cost (taxes and various standing charges), i_W the number of intervals in a week, $P_W(i)$ and $P_S(i)$ the average power used by the HH during the simulated winter or summer week, and $p_{\text{EW}}(i)$, $p_{\text{ES}}(i)$ the real-time price during the winter/summer week. It is assumed that half of the year consists of summer weeks and other half of winter weeks to extrapolate to yearly energy cost. For the evaluation of the two DR programs, the difference in yearly energy cost with respect to the base case (no DR) is used.

For the base case, the yearly energy cost is calculated for the four feeders without the use of any DR. This is simulated for every 15 min value of both a winter and a summer week with all loads assumed to be nonflexible. This results in voltage deviations outside of the limits for feeder D (weakest feeder), the other feeders remain within the limits. The same winter and summer weeks are then simulated with the inclusion of the DR program of the DNO and the ES. The resulting yearly energy costs with respect to the base case are depicted in Fig. 6.

In the figure, the difference in yearly energy cost with respect to the base case is depicted, so a negative cost means that the costs of the DR program are lower than those without DR. The light bars indicate the contribution of the changes in energy cost due to the DR, the dark bars indicate the difference in network cost with respect to the base case, and the red bars indicate the combined effect of both the network and the energy cost on

Fig. 7. Sensitivity of the cost of the two DR programs with respect to the energy cost and the network reinforcement cost.

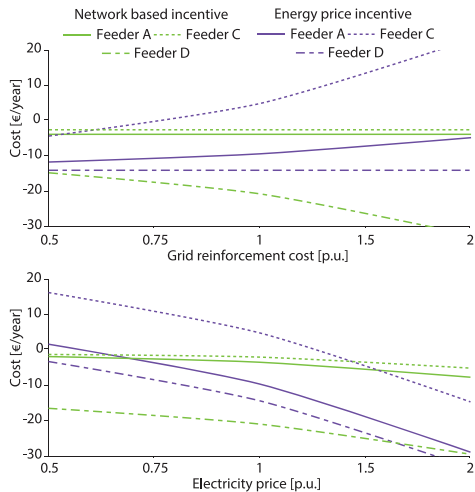the yearly cost. The introduction of the DR program seems to have generally a positive effect for the customer; however, the decrease in energy cost is limited. The exceptions are feeder C with the DR program from the ES point of view, which is more expensive than the base case, and feeder B with a DNO-controlled DR program, which results in negligible cost savings (<1€ cent per year). For feeders A and B, the ES-driven DR program is more profitable, while for feeders C and D, the DNO-driven DR scores best. For the feeders A and C, the cost of the network reinforcements increases as the ES-driven DR program increases the peak loading. For feeder A, this does not result in a higher total cost as the investment costs per connection are much lower due to the higher connection density. The DNO-driven DR program is capable of reducing the peak loading of feeder D enough, so that network reinforcement is no longer necessary. This leads to the largest total saving, as the energy cost is also reduced due to the predominant positive correlation between network loading and energy prices.

B. Sensitivity

The incentives, the DNO and the ES can offer, depend largely on the level of the electricity price and the cost of reinforcing the network. To gain insight into the susceptibility of changes in the results, if these prices change, the simulations are repeated for different electricity prices (with a constant ratio between average price and variation in price) and reinforcement cost levels. Feeder B is not assessed in this analysis as the DNO-driven program will never become more profitable as the level of electricity prices, at which overloading occurs, generates much higher profits for the ES-driven DR program than can be obtained from the DNO-driven program. In Fig. 7, the results of the simulations are plotted.

The sensitivity with respect to the reinforcement cost shows that the optimal DR program from a customer point of view will change for feeder C only if the reinforcement costs are 40% lower than the current costs. For feeders A and D, the cost should be more than doubled or halved to generate a change in the optimal DR program.

The sensitivity of the DR programs with respect to the electricity price is slightly higher. For the feeders, there is an electricity price level at which the optimal approach to the DR changes from DNO driven to ES driven. When there are no major changes, i.e., less than ±30% in the level of the electricity price, the optimal DR program is constant.

VI. CONCLUSION

Two DR programs, one driven by the ES with the goal of lowering the electricity cost and another driven by the DNO to avoid network reinforcements, are assessed. The DR is modeled based on the concept of demand elasticity, and price alterations are generated through incentives based on either real-time electricity prices (ES driven) or on the loading of the network (DNO driven). Simulations are executed for one winter and one summer week for four generic LV feeders from the network of a Dutch DNO. From the simulations, the yearly energy cost and the required cost of network reinforcements are computed. The simulations results show that it is highly dependent on the feeder characteristics which of the DR programs would generate a lower customer energy cost. The ES-driven DR program generates higher peak loads and in some cases these require investments in the network; depending on the topology, these reinforcement costs can dwarf the reduction in electricity cost. The sensitivity of the results with respect to the level of electricity prices and the reinforcement costs is found to be of lesser importance. Which DR program is most profitable from a societal point of view is thus highly dependent on the local network topology.

An initial assessment has shown that for the majority of the HHs, 73%, an ES-driven DR program is the option which generates the most social welfare. For these HHs, the network is strong enough to handle the additional peak which is generated by an electricity-based DR scheme, or the LV network is so weak, reinforcements are necessary despite the use of a network loading-driven DR program. For the other 27% of the LV feeders, a more detailed analysis has been performed to determine whether the ES- or DNO-driven DR program generates the most social welfare and the results differ from one feeder to another. As most of these feeder can benefit from both the ES- and the DNO-driven DR programs, a combined approach for these feeders might yield the best results; however, more research is necessary to determine whether the increased complexity is worth the decrease in network and energy cost. The ES-driven DR approach would yield in most cases a higher social welfare than the DNO-driven approach, so in the absence of further study, the ES-driven approach should be preferred.

The requirements for the DNO- and ES-driven DR programs are however significantly different, as for the DNO, the response of the DR is imperative, as a failure to respond may lead to

network failure. The simulations are performed with an equal load for all the HHs, as especially for the DNO-driven DR program, the HHs at the end of a feeder have much more impact. This allows for the limiting of the DR to just the HHs at the end of the feeder. In the ES-driven DR approach, the effect of many participating customers connected at the end of the cable will lead to an increased chance of violating the voltage limits.

REFERENCES

[1] M. Nijhuis, M. Babar, M. Gibescu, and J. F. G. Cobben, "Demand response: Social welfare maximisation in an unbundled energy market—Case study for the low-voltage networks of a distribution network operator in the Netherlands," in *Proc. 2015 IEEE 15th Int. Conf. Environ. Electr. Eng.*, Jun. 2015, pp. 944–949.

[2] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Jan. 2011.

[3] V. S. K. Murthy Balijepalli, V. Pradhan, S. A. Khaparde, and R. M. Shereef, "Review of demand response under smart grid paradigm," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol.-India*, Jan. 2011, pp. 236–243.

[4] M. A. A. Pedrasa, T. D. Spooner, and I. F. MacGill, "Coordinated scheduling of residential distributed energy resources to optimize smart home energy services," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 134–143, Jan. 2010.

[5] Jouw energie moment (Feb. 2015). [Online]. Available: http://jouwenergiemoment.nl/

[6] Powermatching city (Feb. 2015). [Online]. Available: http://www.powermatchingcity.nl/site/pagina.php?

[7] Epex spot se: Intraday (Nov. 2014). [Online]. Available: http://www.epexspot.com/en/market-data/intraday

[8] Knmi climate data (Nov. 2014). [Online]. Available: http://www.knmi.nl/climatology/daily_data/download.html

[9] A. Panagiotelis and M. Smith, "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions," *Int. J. Forecasting*, vol. 24, no. 4, pp. 710–727, Jan. 2008.

[10] M. Nijhuis, M. Gibescu, and S. Cobben, "Clustering of low voltage feeders form a network planning perspective," in *Proc. 23rd Int. Conf. Electr. Distrib.*, Jun. 2015, pp. 0680-1–0680-5.

[11] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Bottom-up Markov chain Monte Carlo approach for scenario based residential load modelling with publicly available data," *Energy Buildings*, vol. 112, pp. 121–129, Jan. 2016.

[12] R. Stammi *et al.*, "Synergy potential of smart appliances," Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, Tech. Rep., EIE/06/185//SI2.447477 D2.3, Nov. 2008.

[13] D. S. Kirschen, G. Strbac, P. Cumperayot, and D. de Paiva Mendes, "Factoring the elasticity of demand in electricity prices," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 612–617, Jan. 2000.

[14] S. D. Maqbool, M. Babar, and E. A. Al-Ammar, "Effects of demand elasticity and price variation on load profile," in *Proc. 2011 IEEE Power Electr. Soc. Conf. Innov. Smart Grid Technol.-Middle East*, Jan. 2011, pp. 1–5.

[15] N. Venkatesan, J. Solanki, and S. K. Solanki, "Residential demand response model and impact on voltage profile and losses of an electric distribution network," *Appl. Energy*, vol. 96, pp. 84–91, Jan. 2012.

Incorporation of On-Load Tap Changer Transformers in Low-Voltage Network Planning

M. Nijhuis & M. Gibescu
Electrical Energy Systems
Eindhoven University of Technology
Eindhoven, The Netherlands

J.F.G. Cobben
Asset Management & Electrical Energy Systems
Alliander & Eindhoven University of Technology
Arnhem & Eindhoven, The Netherlands

Abstract—**The introduction of distributed generation in the low-voltage (LV) network can generate bi-directional power flows and thus voltage increases instead of decreases from consumers along the feeder towards the substation. The new generation installed at the consumer premises may induce voltage problems while the loading of the cables is still under nominal values. Conventionally possible resulting voltage violations are solved by reinforcing the network, however smart grid alternatives like voltage control in the LV network can also alleviate the network problems. LV-networks are traditionally designed with medium to low voltage transformers equipped with off-line tap changers. The addition of an on-load tap changer (OLTC) for voltage control can decrease the voltage violations in the network, however this needs to be considered within the optimisation method applied for the planning of the LV-network. In this paper the smart grid alternative of installing an OLTC in this optimisation has been performed. By assessing different OLTC control strategies under conditions with increasing distributed generation over many types of networks, the effectiveness of the OLTC becomes apparent. The OLTC is included in the optimization problem formulation by the introduction of additional voltage constraints and relaxing the constraints in the form of a penalty function. When the introduction of an OLTC is more efficient rather than the conventional strengthening of the network is demonstrated with a case study on the impacts of distributed generation.**

Index Terms—**Distribution network planning, smart grids, power system optimisation, on-load tap changer**

I. Introduction

Conventional MV/LV transformers are equipped with an off-line tap changer. The introduction of generation in the LV network increases the absolute magnitude voltage deviations. To maintain these voltage deviations within the required limits, an on-load tap changer (OLTC) can be installed. The use of an OLTC is therefore also often considered in smart grid concepts. The OLTC is however currently not applied and most of the time not even evaluated as one of the possible alternatives of reinforcements in the LV network. The implementation of an OLTC at the MV/LV transformer is now mostly performed on a case-by-case basis and for smart grid pilot projects. To gain more insight into the value that an OLTC can have in the smart grid, a framework to evaluate how the OLTC can be applied within the planning process is needed.

The use of an OLTC in distribution networks has been discussed before. However, the research has been mostly focussed on more advanced operational approaches of combining the control of the OLTC with the control of distributed

generation [1][2] and/or EV [3], solving unbalance problems through individual phase voltage control [4], or using multi-objective optimisation approaches [5]. The control of the OLTC alone has also been a subject of research. While in [6] the control is based on state estimation from measurements within the distribution network, the work in [7] shows that a considerable amount of these advantages can already be achieved by just measuring at the first bus. Different strategies for the application of an OLTC to mitigate over-voltages have been presented in [8]. Though these strategies can all increase the hosting capacity of the network, they are only assessed from an operational perspective. The implementation of an OLTC in the planning process has been touched upon by [9] in a specific case study. The generalisation of this approach and the implementation in the network planning optimisation problem are however lacking. Therefore a framework for the implementation of an OLTC in the LV-network planning process is proposed in this paper.

To develop a framework for the implementation of an OLTC in the LV-network planning process, first of all a closer look is taken at existing LV-network planning methods, and a hypothesis is made regarding what an OLTC would add to planning methods. Secondly, the possible control strategies for the use of an OLTC in the network planning process are discussed in section III. Next in section IV the framework for incorporating an OLTC in the network planning process is defined. Finally, the framework is applied to the case study of three typical LV-networks in section V. Based on this case study, the conclusions on the definition and the application of the new planning framework will be discussed.

II. OLTC and Network Planning

In studies of future loading scenarios for the Dutch LV-network it is shown that the majority of feeders experience over-voltages prior to experiencing overloads [10]. The improvement – i.e. bringing the voltage within an acceptable range of the nominal value of 1 [p.u.] – within a LV-feeder can thus be sufficient to at least defer investments in the cables of the LV-network. The use of an OLTC as an alternative to traditional network reinforcement (the laying of additional or thicker cables), can thus be preferable for the DSO. Whether or not an OLTC can lead to improvements and if the OLTC can defer the network investments long enough to become

an economically viable alternative needs to be determined in the network planning phase. If an OLTC can only delay the investments in cables for a relatively short time, the cost of installation of the OLTC might surpass the deferral of the investments in the cables. The control strategy of an OLTC should also be considered as some control strategies require the installation of additional measurement devices and communication infrastructure. This generates additional investments for the implementation of the OLTC in the distribution grid. These investments should also be considered during the planning phase. To determine how the OLTC can be included in the network planning, first the optimisation method for the reinforcement of the LV-network needs to be defined.

The planning of the LV-network is traditionally performed by the following optimisation procedure:

$$\min_{d} \sum_{t=0}^{T} f_i(t, d)$$

s.t.

$$\mathbf{U}_{t,d} \leq 1.05 \qquad \forall t, d$$
$$\mathbf{U}_{t,d} \geq 0.95 \qquad \forall t, d$$
$$\mathbf{S}_{t,d} \leq 1 \qquad \forall t, d$$
$$\mathbf{I}^T \mathbf{A}_d \mathbf{I} = N - 2 \qquad \forall t, d$$
$$\mathbf{d} \in \{0, 1\} \qquad \forall d \in \mathbb{D}$$

in which the vector $\mathbf{d}$ is the binary variable containing all the possible traditional reinforcement options, T the time horizon for which the network is analysed (40 years), $f_i(t, d)$ are the cost of reinforcing the grid according to the decision variable d incurred at time t, $\mathbf{U}_{t,d}$ the bus voltages and $\mathbf{S}_{t,d}$ the branch apparent power loadings at time t under the set of decisions d, $\mathbf{A}_d$ the adjacency matrix and N the number of buses in the network. In which the voltage limits of ± 0.05 are chosen. This allows for voltage variations of up to ± 0.05 in the MV network, while ensuring the voltage in the LV stays within the $\pm 10\%$ limits.

In the conventional formulation, this optimisation does not include the use of an OLTC. Since the OLTC increases the ability of the network to handle load and local generation by either increasing or decreasing the voltage at the secondary side of the MV/LV transformer, it actually induces changes in the constraints of the network rather than changing the network topology. This will require a change in the optimisation problem formulation rather than just adding another option to the set of possible reinforcement options. To evaluate how the optimisation problem should be changed, the effects of an OLTC on the LV-network will be investigated in the next section.

III. CONTROL STRATEGY

To gain the most out of the application of an OLTC in an LV network, a dedicated control strategy needs to be conceived. As currently there are almost no measurements in the LV network, more advanced control strategies would require investments in measurement and communication infrastructure,

in addition to the investments in the OLTC itself. However these control strategies would result in even larger allowable voltage variations within the LV-network. To determine the value of an OLTC within the network planning process the different applicable control strategies need to be defined. In this paper 4 main strategies are introduced. The strategies vary in complexity and in the amount of communication they require. Each of the strategies is discussed shortly below.

A. $U_{MV/LV} = 1[p.u.]$: The most simple control strategy is to regulate the voltage at the busbar of the MV/LV substation $U_{MV/LV}$ to 1 [p.u.]. This control strategy eliminates the voltage deviations from the MV network and thus increases allowable voltage deviations in the underlying LV-network. In this approach, only a local voltage measurement at the secondary side of the transformer is needed.

B. $U_{MV/LV}(I_{MV/LV})$: Most HV/MV transformers in the distribution network are equipped with an OLTC. The control strategy of these OLTC's is based on the amount of current flowing through the transformer. For the MV/LV transformers a similar strategy can also be applied. In this approach a current measurement at the secondary side of the transformer needs to be installed.

C. $U_{MV/LV}(I_1, ..., I_n)$: The current for different LV-feeders connected to a single substation can differ significantly, especially if PV is introduced at only some of them. From the total current flowing through the transformer it cannot be distinguished whether it is a period of low load, or of high load and high generation at different feeders. Therefore, with the current at the beginning of each feeder $(I_1, ..., I_n)$ a control strategy which allows for even larger voltage deviations can be accomplished.

D. $U_{MV/LV}(U_{e,1}, ...U_{e,n})$: A more complete picture of the voltage distribution can generally be obtained by using the voltage at the end of each LV-feeder $(U_{e,1}, ...U_{e,n})$. With this information the uncertainty of the voltage within the LV-feeder will be smallest, allowing for the tightest control of the OLTC, however extra investments in ICT are needed.

IV. OLTC IMPLEMENTATION FRAMEWORK

With the introduction of an OLTC in the network planning process, the main aspect which should be determined is how many years the OLTC can defer the investments in larger or additional cables. If this is estimated for a given LV-network, a net present value calculation can be employed to determine if the expected years of deferral are worth the additional cost of the OLTC and related equipment. With the analysis of the OLTC it is important to note that the OLTC has a positive asset recovery value. The OLTC could be swapped with a normal transformer when the network needs upgrading, allowing the OLTC to be used in other LV-networks.

To what extent an OLTC will be feasible is very dependent on the load of the feeder and the network characteristics. For short networks with cables with a relatively small diameter, overloading problems will occur sooner, so the amount of investment deferral will be limited. The same goes for the case of large expected load growth. If the load changes within the

network are high, the moment for which additional network reinforcements are needed comes sooner, thus reducing the amount of investment deferral. The question as to whether the implementation of an OLTC is an effective method to keep the voltage within its limits should thus be determined per network and for different loading situations. In order to gain insight into the possible allowable voltage deviations, probabilistic load flow calculations are used to determine the chance of voltage limit violations for the different OLTC strategies. An overview of this approach is given in Fig. 1.
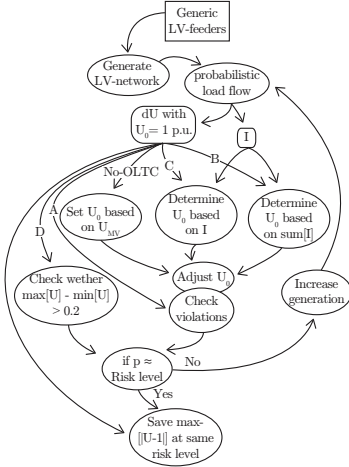

Fig. 1. Overview of the approach to assess the OLTC performance for the control strategies A-D and no-OLTC case

The amount of voltage deviations allowable within the LV-network should be determined for the different control strategies and compared to the reference case without the application of an OLTC. The first step in this analysis is to perform probabilistic load flow calculations to get a baseline for the voltages within the network. In this way, the probability distribution of the voltage at each node and current in each branch can be determined. In this calculation, the voltage at the LV busbar of the substation is assumed to be constant and equal to 1 p.u. This probabilistic load flow can be calculated with the implementation of different levels of generation present within the LV-network. In this way the baseline should extend to the situations in which the voltages deviations in the network are more than +20%. For the load, a Gaussian mixture estimate, at 12:00 (peak of the PV-production) and at 20:00 (peak of the load) based on a large number of household load curves is used. The load curves are created by using a bottom-up load model [11]. This Gaussian mixture distribution is estimated using the expectation minimisation algorithm on the generated load data of the households with and without PV. By varying the component proportions the Gaussian mixture,

the combined distribution can consist of more or less houses with PV. The distribution of the load at 12:00 is given in Fig. 2.
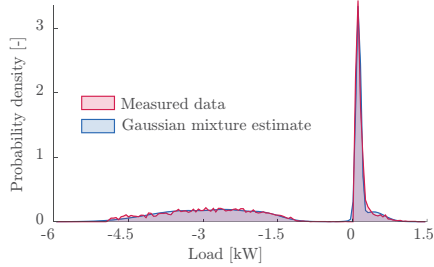
Fig. 2. Gaussian mixture estimate of the load with and without PV

In the case when there is no OLTC present within the network, the peak loading should also be assessed. As the voltage set point cannot change in between and during the peak load conditions, the LV busbar voltage should still be sufficiently high enough to keep the voltage within the limits. For the peak loading a distribution is based on the same generated household loads, considering only the evening residential peak time of 20:00 hrs.

For each of the different control strategies of the OLTC, the LV busbar voltage will change and thus also the voltages at each of the buses. The changes in bus voltages can be accurately estimated by utilising the calculated probability distribution of the voltages and the change at the LV busbar. How these final voltage deviations are calculated depends on the chosen control strategy of the OLTC. For each of the control strategies and the case without an OLTC the required calculation steps are described in more detail below.

No-OLTC: For the case without any OLTC the voltage deviations allowable in the LV-network only extend to $\pm0.05\%$, since the MV network is designed to have voltage deviations of up to 5%. Based on the risk level the DSO applies, somewhat higher or lower voltages may be allowed during the planning phase if the chance of occurrence is low. The off-line tap changer of the transformer is set to achieve an average voltage closest to 1 p.u. during high loading times. For the off-line tap changer, a discrete set of tap positions (-0.5kV,-0.25kV,...,0.5kV as seen from the primary side) is assumed. For the voltage at the MV-side of the MV/LV transformer the correlation between the loading of the transformer and the voltage at the MV level is used. Based on measurements, this correlation is found to be 0.94. The MV busbar voltage is estimated using Cholesky factorisation in combination with the calculated correlation and the computed transformer load. Based on the voltage set point and the estimated MV voltage, the new voltage at the LV busbar $U_{0,new}$ is calculated. The voltage distribution in the network is sequentially estimated in the following manner:

$$\mathbf{U} = \frac{U_{0,new}^2 - U_{0,old}^2}{U_{0,new}} + \frac{U_{0,old}}{U_{0,new}}\mathbf{U}_i \qquad (1)$$

where $U_{0,old}$ is the initial voltage at busbar and $\mathbf{U}_i$ the initially calculated voltage at bus i.

A. $U_{MV/LV} = 1[p.u.]$: If the voltage at the LV busbar is kept constant at 1 [p.u.] the voltage fluctuations allowable in the LV-network increase as the MV voltage fluctuations can be assumed zero in this case. This change does not require any additional analysis compared to the initial calculation, as with the initial calculation the voltage at the LV busbar is already assumed to be constant.

B. $U_{MV/LV}(I_{MV/LV})$: With the control of the OLTC dependent on the current flowing through the transformer, the magnitude of the voltage deviations should be determined with respect to the link between this current and the voltages at each bus. The control based on the maximum current tries to make the average bus voltage as close to 1 p.u. as possible. To generate this relationship, the bivariate distribution of the current through the transformer and the average voltage at the buses is established, by using the correlation between the total current and the average voltage. The correlation is determined by the load flow calculations and found to be between -0.78 and -0.84. In this way, for each value of the current the expected value of the average voltage can be calculated and the resulting LV-busbar voltage at which the average voltage at the buses are closest to 1 p.u. can be determined.

C. $U_{MV/LV}(I_1, ..., I_n)$: If the measurements of the currents are extended to each feeder, the distribution of the voltage deviations can be more accurately estimated in comparison with the case above, where only aggregated measurements are available. The approach to determining the voltage set point is the same as with the $U_{MV/LV}(I_{MV/LV})$ case, only the combination of the expected average voltages at each feeder is used separately to determine the best voltage set point. In this way the average system voltage still determines the setting of the OLTC, only the voltages in the feeders can be more accurately estimated.

D. $U_{MV/LV}(U_{e,1}, ...U_{e,n})$: For the assessment of switching the tap position based on the voltages at the end of the feeders, the allowable voltage range within the LV-network increases to 20% (as this will allow all voltages to have values between -10% and +10%). The assessment of the chance of voltage violations versus the voltage limits applied during the planning process is determined by the chance of violating this 20% range by calculating the difference between the lowest and highest voltages within the network.

To assess whether the network is still sufficiently strong with the OLTC in place, the risk of violating the voltage limits is calculated based on the distribution of the voltages within the network. For each sample in the distribution, the node with the highest average voltage deviation for the five cases (no-OLTC and the four different OLTC control options) is determined. At two percentile values (0.95 and 0.99) the distribution of the voltage at this (most critical)node is evaluated. If the percentile value is found to violate the voltage limits, the corresponding

voltage in the case of the constant LV busbar voltage is calculated. In this way the possible voltage limits which can be applied while assessing the network with a constant LV busbar voltage can be determined. As this is the way in which the voltage limits are usually assessed during the LV-network planning process.

V. RESULTS

The performance of the OLTC with the four different control strategies can now be assessed. For the assessment, the calculation of the voltage distribution at each bus will be done for a large variety of networks. To have a good mix of networks, networks will be generated randomly for a set of 94 generic feeders [12]. These networks are subsequently combined into a single LV-network and a large number of these networks will be used in the analysis. This is done to create a representative set of networks to determine if conclusions can be generalized. The resulting maximal allowable voltage deviations during the planning phase are given in Fig 3 and Fig. 4 for the 95th and 99th percentiles respectively.
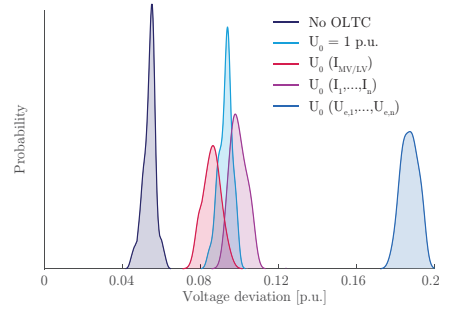

Fig. 3. The 95th percentile of the maximum allowable voltage deviation during the planning phase for the different control strategies
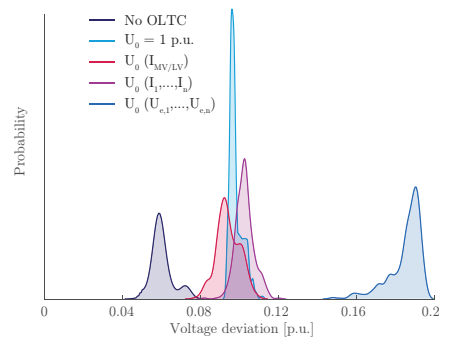

Fig. 4. The 99th percentile of the maximum allowable voltage deviation during the planning phase for the different control strategies

From both figures a number of things can be concluded. First of all, as most LV-feeders have been developed to have a maximum allowable voltage drop of 0.05 p.u., not including voltage variations within the medium voltage network. For the case without an OLTC the calculated allowable voltage deviation confirms this. Due to the stochastic nature of the assessment and the discrete increments of the generation within the LV-network, the value of the allowable voltage will vary around the 0.05 p.u. value. The 95th percentile values also show a much tighter distribution than the 99th percentile values. For the three control strategies $U_{MV/LV}(I_1, ..., I_n)$, $U_{MV/LV}(I_{MV/LV})$ and $U_{MV/LV} = 1[p.u.]$ the difference between the admissible voltages is not large, with the $U_{MV/LV}(I_{MV/LV})$ strategy slightly under-performing with respect to the others, The voltage used in the planning process can generally be increased to $\pm$ 10%. The use of strategies based on the current can therefore be discarded as viable options for a generic network, though there may be situations in which these strategies can be considered superior (i.e. uniform loading conditions). As expected, the use of the voltage measurements at the end of the feeders can increase the admissible voltages during the planning process the most, to on average allow for voltages up to +19%.

To include these resulting distributions in the network planning optimisation the voltage constraint which is now set at ± 0.05 p.u. can be replaced by a penalty function. This alters the constrained optimisation problem as follows:

$$f_p(\mathbf{X}) = f(\mathbf{X}) + \sum_{i=1}^{n} C_i \delta_i \quad (2)$$

in which $f_p(\mathbf{X})$ is the penalised objective function, $f(\mathbf{X})$ the original objective function, C_i the penalty which needs to be applied to constraint i and δ_i a binary variable indicating whether or not a constraint is violated. The original voltage constraint will be translated into a penalty with a weight equal to the cost of the OLTC, and an additional two constraints. The first one, is the violation of the ± 0.1 p.u. voltage limit and the second one is the violation of the ± 0.19 p.u. voltage limit. These voltage limits correspond to the maximum admissible voltage with an OLTC and with an OLTC and voltage measurements at the end of the feeder respectively. The weight of the penalty functions should be the cost of installing the measurements and a weight large enough to make voltages deviation higher than 0.19 p.u. prohibitive. With this optimisation defined, the OLTC has been incorporated into the network planning optimisation. The results of this are shown in a case study below.

VI. CASE STUDY

For the case study three typical LV-networks are selected: a rural, a sub-urban and an urban network. The main characteristics of the networks can be found in the Table I. All the networks are three-phase and have an operating voltage of 0.38kV line-to-line.

TABLE I
CHARACTERISTICS OF THE THREE LV-NETWORKS USED IN THE CASE STUDY

	Rural	Sub-Urban	Urban
Total Length[m]	5190	2740	1523
Connections	69	124	190
# Feeders	4	4	5

TABLE II
COST OF THE COMPONENTS USED IN THE CASE STUDY

Component	Cost [€]
LV-Cable [\km]	45 000
MV-Cable [\km]	70 000
Additional MV/LV substation	25 000
LV-busbar expansion for an additional feeder	7 000
OLTC-transformer	32 000
OLTC recovery value	25 000
Voltage measurements per feeder	1 000
Fixed investment cost	10 000

The costs as shown in Table II are applied for the components in this case study:

The asset recovery value of the OLTC should also be taken into account for the optimisation, as the DSO can remove the OLTC if cable reinforcements have been applied and use it somewhere else in the network against limited cost. A depreciation rate of 3% is applied to allow for the monetisation of the investment deferral. The losses in the network are neglected in this case study. To assess the LV-networks, the PV penetration rate is increased according to three scenarios. An overview of these scenarios is given in Fig. 5
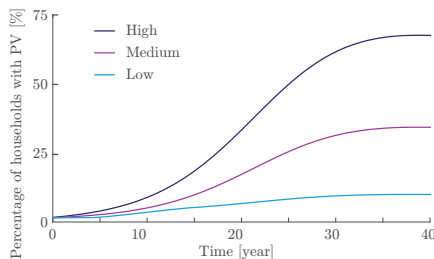
Fig. 5. PV penetration ratio for three scenarios of the OLTC case study

The figure shows a high, middle and low scenario which have been generated for the Netherlands according to [13]. First of all the middle scenario is employed and applied to all three LV-networks, rural, suburban and urban. This results in the reinforcement costs depicted in Fig 6. Note that all NPV calculations include the cost of the OLTC and the associated measurement infrastructure if the voltages at the end of the feeder are used as part of the control strategy.

From the figure it can be seen that in two out of three cases the OLTC generates a positive result with respect to reinforcements by replacing or laying additional cables. For the urban case the OLTC does generate a higher cost, however
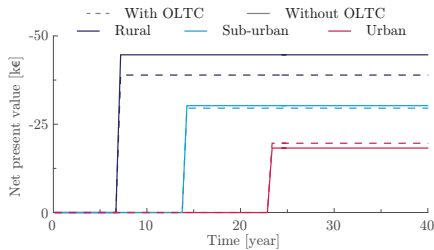
Fig. 6. Cost of the reinforcement of the three LV-networks with and without an OLTC, for the medium growth PV scenario

if the asset recovery value of the OLTC is included after 40 years, it would generate a more favourable result. Due to the relatively high fixed investment cost in both the rural and sub-urban case it is most effective to install the OLTC with measurements at the end of the feeder directly.

To gain insight into the effects with different loading scenarios. The analysis is also applied for the three different PV growth scenarios applied to the sub-urban network. The results are depicted in Fig. 7.
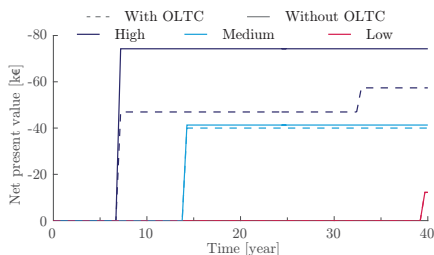
Fig. 7. Cost of the reinforcement of the sub-urban LV-network with and without an OLTC for the three PV penetration scenarios

In the figure it can be seen that the low scenario only requires investments after 38 years, and also that the amount of investments required are low enough to not warrant the use of an OLTC. Only the high scenario shows the effect of the OLTC in investment deferral. The OLTC can delay the investment in the cables of the network until the year 32. In this year, the OLTC can still be recovered. The initial investments which are done in year 7 for the high scenario in both the OLTC and in the no-OLTC case are needed in the network to ensure it is still sufficiently strong. In year 32 the voltage deviations in the network still not exceed the limits, however some cables within the network become overloaded in the OLTC case, therefore requiring some reinforcements.

VII. CONCLUSIONS

The incorporation of an OLTC in the network planning optimisation has been researched. Based on probabilistic load flow calculations for cases with and without an OLTC, the maximum admissible voltage in the underlying feeders can be determined for the different operating approaches. The admissible voltage is found to be independent of the network structure, allowing for an elegant implementation of the OLTC in the network planning optimisation. By relaxing the voltage constraints into penalty functions the use of an OLTC can be evaluated within the same framework as the conventional network planning. A case study of three typical LV-networks in the Netherlands, with different PV penetration scenarios, show that the OLTC can generate value for the DSO in a number of cases. The case study shows that the OLTC can be used to defer network investments or as a complete smart grid alternative to network reinforcements, especially for situations with long, rural or sub-urban feeders where a scenario of high growth of PV is anticipated.

REFERENCES

[1] K. M. Muttaqi, A. D. T. Le, M. Negnevitsky, and G. Ledwich, "A coordinated voltage control approach for coordination of oltc, voltage regulator, and dg to regulate voltage in a distribution feeder," *IEEE Transactions on Industry Applications*, vol. 51, no. 2, pp. 1239–1248, Mar 2015.

[2] D. Ranamuka, A. P. Agalgaonkar, and K. M. Muttaqi, "Online voltage control in distribution systems with multiple voltage regulating devices," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 617–628, Mar 2014.

[3] M. A. Azzouz, M. F. Shaaban, and E. F. El-Saadany, "Real-time optimal voltage regulation for distribution networks incorporating high penetration of pevs," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3234–3245, Nov 2015.

[4] J. Hu, M. Marinelli, M. Coppo, A. Zecchino, and H. W. Bindner, "Coordinated voltage control of a decoupled three-phase on-load tap changer transformer and photovoltaic inverters for managing unbalanced networks," *Electric Power Systems Research*, vol. 131, pp. 264–274, Oct 2015.

[5] J. R. Castro, M. Saad, S. Lefebvre, D. Asber, and L. Lenoir, "Optimal voltage control in distribution network in the presence of dgs," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 239–247, Dec 2015.

[6] S. N. Salih and P. Chen, "On coordinated control of oltc and reactive power compensation for voltage regulation in distribution systems with wind power," *IEEE Transactions on Power Systems*, p. 1, Nov 2015.

[7] C. Reese, C. Buchhagen, and L. Hofmann, "Enhanced method for voltage range controlled oltc-equipped distribution transformers," in *2012 IEEE Power and Energy Society General Meeting*, Jul 2012.

[8] C. Long and L. F. Ochoa, "Voltage control of pv-rich lv networks: Oltc-fitted transformer and capacitor banks," *IEEE Transactions on Power Systems*, p. 1, Apr 2015.

[9] A. Navarro-Espinosa and L. F. Ochoa, "Increasing the pv hosting capacity of lv networks: Oltc-fitted transformers vs. reinforcements," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2015, pp. 1–5.

[10] M. Nijhuis, M. Gibescu, and S. Cobben, "Scenario analysis of generic feeders to assess the adequacy of residential lv-grids in the coming decades," in *PowerTech 2015*, Jun 2015.

[11] M. Nijhuis, M. Gibescu, and J. F. G. Cobben, "Bottom-up markov chain monte carlo approach for scenario based residential load modelling with publicly available data," *Energy & Buildings*, vol. 112, pp. 121–129, Jan 2016.

[12] M. Nijhuis, M. Gibescu, and S. Cobben, "Clustering of low voltage feeders from a network planning perspective," in *CIRED 2015*, Lyon, France, Jun 2015.

[13] J. J. C. Bruggink, "The next 50 years: Four european energy futures," Tech. Rep., May 2005.

B

Generic LV-feeders

Table B.1: *Overview of the generic feeders obtained from the network clustering*

Num	Branch	Length [m]	Z [Ω/km]	Z at PoC [Ω] mean	Z at PoC [Ω] max	Connections Total	Connections HH [%]	Rate [%]
1	2	493	418	78	109	7	86	0.0
2	2	424	321	61	124	71	79	0.0
3	4	837	778	203	395	62	94	0.0
4	3	480	379	56	116	23	74	0.0
5	3	958	533	218	486	15	87	0.0
6	2	683	350	100	168	9	89	0.0
7	4	799	695	157	288	62	94	0.0
8	4	831	750	265	545	64	94	0.0
9	3	578	347	53	110	12	92	0.0
10	4	963	860	250	631	63	92	0.0
11	2	433	354	59	146	8	88	0.1
12	2	643	512	154	255	7	100	0.1
13	2	498	569	135	197	7	100	0.1
14	3	509	409	81	156	34	88	0.1
15	4	839	753	287	553	67	94	0.1
16	9	519	432	105	363	42	98	0.1
17	2	496	345	70	110	11	91	0.1
18	3	535	382	74	180	48	83	0.1
19	4	617	605	127	255	42	100	0.1
20	3	1023	316	112	193	14	93	0.1
21	2	639	533	130	236	8	100	0.1
22	4	721	649	146	317	49	94	0.1
23	2	530	318	74	97	7	100	0.2
24	1	258	340	47	88	9	100	0.2
25	2	90	208	77	138	28	100	0.2
26	3	1047	322	157	295	10	100	0.2
27	5	743	352	89	191	28	57	0.2
28	2	846	438	140	207	14	93	0.2
29	3	746	449	143	244	16	94	0.2
30	2	533	323	56	92	8	63	0.2
31	2	276	324	42	82	13	46	0.3
32	2	259	320	22	76	9	56	0.3
33	3	878	343	132	198	16	75	0.3
34	6	824	444	99	196	76	91	0.3
35	2	757	310	127	159	8	100	0.3

Num	Branch	Length [m]	Z [Ω/km]	Z at PoC [Ω] mean	Z at PoC [Ω] max	Connections Total	Connections HH [%]	Rate [%]
36	2	454	506	95	212	8	100	0.3
37	4	984	338	131	262	17	94	0.4
38	3	592	545	152	297	47	94	0.4
39	3	354	339	41	73	23	83	0.4
40	4	642	394	69	131	42	100	0.4
41	9	501	530	149	330	87	99	0.4
42	3	501	408	81	178	23	100	0.5
43	4	939	346	131	198	14	100	0.5
44	1	505	379	118	191	8	75	0.5
45	2	581	314	49	109	12	100	0.5
46	2	279	318	282	705	19	42	0.5
47	2	277	622	94	185	29	97	0.5
48	2	283	304	34	52	9	56	0.6
49	1	299	386	44	116	6	83	0.6
50	2	513	539	102	250	15	100	0.6
51	1	332	380	83	126	8	100	0.6
52	1	177	374	21	66	9	100	0.7
53	1	312	305	43	63	99	93	0.7
54	4	410	397	60	107	15	67	0.7
55	3	899	348	89	163	19	100	0.7
56	2	335	957	93	169	27	100	0.8
57	3	551	357	74	118	15	87	0.8
58	1	260	295	51	77	30	70	0.9
59	5	921	482	174	395	110	99	0.9
60	2	625	335	90	187	15	93	0.9
61	2	421	409	56	140	73	100	1.0
62	2	355	367	52	111	17	76	1.0
63	4	658	419	87	161	48	94	1.0
64	1	292	298	54	87	102	99	1.0
65	1	299	435	69	130	6	83	1.1
66	5	763	432	96	208	39	100	1.1
67	2	320	407	40	99	16	100	1.1
68	1	161	307	33	49	8	75	1.1
69	2	453	433	95	179	25	100	1.2
70	1	246	294	49	72	30	83	1.3

Num	Branch	Length [m]	Z [Ω/km]	Z at PoC [Ω]		Connections		Rate [%]
				mean	max	Total	HH [%]	
71	3	1011	468	156	394	58	100	1.5
72	3	392	390	53	96	26	96	1.5
73	1	146	343	17	50	12	100	1.6
74	5	498	435	68	185	46	100	1.6
75	2	595	417	120	233	33	91	1.7
76	1	190	543	62	103	13	100	1.8
77	1	187	322	40	60	11	82	1.8
78	1	233	298	27	69	67	100	2.1
79	2	354	386	62	111	26	92	2.1
80	4	567	419	101	218	51	100	2.2
81	1	227	291	27	66	38	95	2.5
82	4	499	371	75	126	62	98	2.6
83	1	277	404	77	112	13	92	2.8
84	3	371	535	68	165	33	100	2.9
85	2	439	396	65	145	35	100	2.9
86	1	320	379	86	121	27	100	3.2
87	4	617	383	108	211	74	100	3.3
88	2	312	361	36	77	51	100	3.4
89	2	274	496	41	83	29	100	3.4
90	3	347	349	42	97	35	100	4.1
91	1	208	485	57	101	20	100	4.4
92	1	255	307	45	78	38	100	4.5
93	2	224	327	24	57	24	100	4.5
94	1	165	276	24	46	18	100	6.4

C

Data used for the network expansion planning

Table C.1: *Data used for the network expansion planning optimisation*

Symbol	Component	value
C_{cbl}	LV-Cable [€\mtr]	65.5
$C_{cbl,mv}$	MV-Cable [€\mtr]	77
C_{fdr}	LV-busbar expansion [€]	4200
p_e	Electricity price [€\MWh]	0.73
p_{av}	Outage price [€\min]	0.05
C_{sub}	Additional MV/LV substation [€]	40000
$C_{inv,fix}$	Fixed investment cost [€]	8300
λ_{cbl}	Cable failure rate [\(yr km)]	0.038
Tr_{cbl}	Repair time [min]	200

List of Publications

Journal papers

2017

Nijhuis, M., Carvalho, P.M., Gibescu, M., and Cobben, J. F. G. (2017). Low Voltage Multi-Stage Network Expansion Planning Under Uncertainty; a bilevel evolutionary optimisation approach. *Submitted*

Nijhuis, M., Carvalho, P.M., Gibescu, M., and Cobben, J. F. G. (2017). Modelling of Load Uncertainty within a Multi-Stage LV-network Expansion Planning Optimisation. *Submitted*

Haque, A.N.M.M., Nijhuis, M., Ye, G., Nguyen, P.H., Bliek, F.W. and Slootweg, J. G. (2017). Integrating Direct and Indirect Load Control for Congestion Management in LV Networks. *Transaction on Smart Grids* (in press)

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2017). Valuation of Measurement Data for Low Voltage Network Expansion Planning. *Electric Power Systems Research*, 151:59–67

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2017). Analysis of Reflectivity & Predictability of Network Tariff Structures for Household Consumers. *Energy Policy*, 109:631–641

Ni, F., Nijhuis, M., Nguyen, P.H., and Cobben, J. F. G. (2017). Variance-Based Global Sensitivity Analysis For Power Systems. *IEEE Transactions on Power Systems* (in press)

Ye, G., Nijhuis, M., Cuk, V., and Cobben, J. F. G. (2017). Scenario-based Approach to Assess the Grid Impact of Harmonic Sources. *Energies* 10:372

Ye, G., Xiang, Y., Nijhuis, M., Cuk, V., and Cobben, J. F. G. (2017). Bayesian Inference-based Voltage Dip State Estimation. *IEEE Transactions on Instrumentation & Measurement* (in press)

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2017). Gaussian Mixture Based Probabilistic Load Flow for LV-Network Planning. *IEEE Transactions on Power Systems* 32(4):2878–2886

2016

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2016). Bottom-up Markov Chain Monte Carlo Approach for Scenario Based Residential Load Modelling with Publicly Available Data. *Energy and Buildings*, 112:121–129.

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2016). Risk-based Framework for the Planning of Low-Voltage Networks Incorporating Severe Uncertainty. *IET Generation, Transmission & Distribution*, 11(2):419–426.

Nijhuis, M., Babar, M., Gibescu, M., and Cobben, J. F. G. (2016). Demand Response: Social Welfare Maximisation in an Unbundled Energy Market – Case Study for the Low-Voltage Networks of a Distribution Network Operator in the Netherlands. *IEEE Transactions on Industry Applications* (in press)

2015

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2015). Assessment of the Impacts of the Renewable Energy and ICT Driven Energy Transition on Distribution Networks. *Renewable and Sustainable Energy Reviews*, 52:1003–1014.

2014

Nijhuis, M., Rawn, B., and Gibescu, M. (2014). Prediction of Power Fluctuation Classes for Photovoltaic Installations and Potential Benefits of Dynamic Reserve Allocation. *IET Renewable Power Generation*, 8(3):314–323.

Conference papers

2017

Ye, G., Nijhuis, M. (2017). Scenario-based Approach to Assess the Grid Impact of Harmonic Sources. In *IEEE Powertech 2017*, Manchester, United Kingdom

2016

Nijhuis, M., Bernards, R., Gibescu, M., and Cobben, J. F. G. (2016). Stochastic Household Load Modelling from a Smart Grid Planning Perspective. In *2016 IEEE International Energy Conference (ENERGYCON)*, Leuven, Belgium

Nijhuis, M. and Cobben, J. F. G. (2016). Effects of Power Quality Limits on LV-Network Design. In *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, pages 83–88, Belo Horizonte, Brazil

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2016). Application of Resilience Enhancing Smart Grid Technologies to Obtain Differentiated Reliability. In *IEEE EEEIC 2016*, Florence, Italy.

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2016). Incorporation of On-Load Tap Changer Transformers in Low-Voltage Network Planning. In *ISGT Europe 2016*, Ljubljana, Slovenia.

2015

Nijhuis, M., Babar, M., Gibescu, M., and Cobben, J. F. G. (2015). Demand Response: Social Welfare Maximisation in an Unbundled Energy Market – Case Study for the Low-Voltage Networks of a Distribution Network Operator in the Netherlands. In *IEEE EEEIC 2015*, Rome, Italy.

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2015). Scenario Analysis of Generic Feeders to Assess the Adequacy of Residential LV-Grids in the Coming Decades. In *2015 IEEE Eindhoven PowerTech, PowerTech 2015*, Eindhoven, The Netherlands

Nijhuis, M., Gibescu, M., and Cobben, J. F. G. (2015). Clustering of Low Voltage Feeders from a Network Planning Perspective. In *Cired 2015*, Lyon, France.

Nijhuis, M., Vonk, B., Gibescu, M., Cobben, J. F. G., and Slootweg, J. G. (2015). Assessment of Probabilistic Methods for Simulating Household Load Patterns in Distribution Grids. In *Cired 2015*, Lyon, France

2011

Nijhuis, M., Rawn, B.G., and Gibescu, M. (2015). Classification Technique to Quantify the Significance of Partly Cloudy Conditions for Reserve Requirements due to Photovoltaic Plants. In *PowerTech, 2011 IEEE*, Trondheim, Norway.

Acknowledgements

It would not have been possible to enjoy the last 4 years of doing PhD research, without the help and support of many people. I would like to take a moment to thank a number of them.

First of all, I would like to thank Alliander and Sjef Cobben for giving me the opportunity to do my PhD. From the moment I started with the research I had the feeling I was being entrusted with a very large amount of freedom to perform the research in the way I saw fit. This made the research a far more rewarding experience to me. Next to this, I would also like to thank Sjef for his guidance during my PhD, encouraging me to enjoy the research and asking critical questions when necessary.

Madeleine, even though the subject of low voltage networks was not your own at the beginning of my PhD you have been an invaluable part of my PhD research. Forcing me to clearly explain every step of what I have done and what I wanted to do, assisted me a lot in getting a better grip on my research and made me see where I needed to improve. Our discussions were often not directly connected to my PhD research, they strengthened me in the decisions I have made in my life. Thank you.

My PhD thesis would not have been the same without my visit to IST Lisbon. Pedro your guidance during this period gave me more insight into how to accurately express the research I have done and the application of heuristic optimisation. My thesis would not have been of the same quality without your help during my stay in Lisbon and I am grateful for the guidance you were willing to give me during this period. I would also like to thank Stichting 3E for granting me the opportunity to stay at IST Lisbon.

To the members of my PhD committee, Prof. Lars Nordström, Prof. Peter Palensky, Prof. Han Slootweg and Prof. Guus Pemen, I would like to extend my gratitude for carefully assessing my thesis and the discussions and comments which followed. These comments and discussions they improved the quality of this thesis and gave me additional insight into the work I have done and the gaps which still remain.

I would also like to thank Frans Provoost, for taking the time to teach me how the distribution network operates and for the feedback he has given me during my PhD. Without his guidance, in particular when I was still at Alliander, I would never have developed an interest in the distribution network let alone be able to do a PhD in the subject.

I would like to thank my family for the support they have given me during my PhD. Especially I would like to thank my father for his support and convincing me I made the right decision to do a PhD. I could not think of anyone who would have been more proud of what I have achieved during my 4 years of PhD research.

Lastly, I would like to extend my thanks to my colleagues at the EES group. You helped me through these 4 years by collaborating on research, learning about other cultures, having interesting discussions, drinking a beer, or just wasting time. Without you, I would never have enjoyed my PhD as much!

Curriculum Vitae

Michiel Nijhuis was born on the 24th of March 1986, in Maastricht, the Netherlands. He received his BSc degree in Aerospace Engineering in June 2009, from the Delft University of Technology. In September 2008 he switched to the applied physics department to start a master program in Sustainable Energy Technology also at Delft University of Technology. His thesis work for the master program was performed at the Faculty of Electrical Engineering, Mathematics and Computer Science, where he focussed on the effects of PV-induced power fluctuations on the reserves in the power system. In February 2011 he received his MSc degree. Hereafter he joined the Dutch distribution network operator Alliander as a trainee. He worked at Alliander for two and a half years at various departments. In November 2013 he left Alliander to pursue a PhD degree with the Electrical Energy Systems Group at Eindhoven University of Technology. His research focusses on the long-term adequacy of the low voltage network.