

Burmeister, Sascha Christian; Schryen, Guido

**Article — Published Version**

## Distribution network optimization: predicting computation times to design scenario analysis for network operators

Energy Systems

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Burmeister, Sascha Christian; Schryen, Guido (2023) : Distribution network optimization: predicting computation times to design scenario analysis for network operators, Energy Systems, ISSN 1868-3975, Springer, Berlin, Heidelberg, pp. 1-28, <https://doi.org/10.1007/s12667-023-00572-5>

This Version is available at:

<https://hdl.handle.net/10419/307053>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Distribution network optimization: predicting computation times to design scenario analysis for network operators

Sascha Christian Burmeister<sup>1</sup> · Guido Schryen<sup>1</sup>

Received: 29 December 2021 / Accepted: 13 February 2023  
© The Author(s) 2023

## Abstract

The increased use of renewable energies promotes decarbonization and raises the load on power distribution networks, forcing responsible distribution network operators to re-evaluate and re-design their networks. Infrastructure planners employ a rolling-horizon planning procedure with frequent recalculations to face informational uncertainty, which require solving multiple scenarios. Keeping complexity manageable is particularly challenging as distribution network areas may span multiple cities and counties. In this study, we focus on infrastructural decomposition, where the distribution network is decomposed into multiple parts and planning problems, which are then optimized separately. However, infrastructure planners lack the knowledge of how they should design a scenario analysis for a subnetwork to account for informational uncertainties subject to limited planning time and computing resources. Based on empirical requirements from literature and discussions with experts, we present a novel mixed integer linear optimization model that allows to use exact solution approaches for realistic large-scale distribution networks. Our approach considers the primary and secondary distribution network in an integrated way and designs a flexible topology for high reliability power distribution. We perform extensive computational experiments and a sensitivity analysis to determine correlations between the values of model parameters and computation times required to solve the resulting model instances to optimality. The results of the sensitivity analysis indicate that the combination of the number of buses, lines and the considered action scope have a considerable influence on the solving time. In contrast, a higher number of available transformers led to a better solvability of the model. From these computational insights, we derive implications for infrastructure planners who wish to perform scenario analysis for planning their power distribution networks.

**Keywords** OR in energy · Networks · Distribution network optimization · Linear programming

---

✉ Sascha Christian Burmeister  
[sascha.burmeister@upb.de](mailto:sascha.burmeister@upb.de)

Extended author information available on the last page of the article

# 1 Introduction

Renewable energy plays a decisive role in climate change mitigation. The transformation to renewable energy (e.g., solar power, wind power, or hydropower) can reduce greenhouse gas emissions and contribute to the fossil fuel phase-out progress [41]. The increased market liberalization shows a positive effect on the production of decentralized renewable energy. This effect will be promoted by the further development of renewable energy sources [39]. The consumer side of the energy market reflects this trend towards renewable energy: the number of heat pumps grows while an exponential increase of electric vehicles is expected for the next years [14, 43, 45, 58]. In addition to producers and consumers, there are also prosumers who generate renewable energy themselves and share surplus energy [60]. Prosumers include household, commercial and industrial actors of the energy market and enable opportunities for innovative energy business models [8].

The capabilities of market actors to obtain or feed in energy are limited by the underlying infrastructure, i.e., by the distribution network. Distribution network operators are responsible for the operation and strategic planning of the distribution network and are, thus, intermediaries for the energy market. When conventional distribution networks were planned, the assets (e.g., substations, transformers, and power lines) were not designed for the load of the emerging energy market. Therefore, distribution network operators must extend their distribution networks to meet future needs. In the case of Germany, the *German Energy Academy* estimates expenses up to 42.5 billion euros until 2030 [25] whereas the *Federal Ministry for Economic Affairs and Energy* assumes additional costs of up to 48.9 billion euros until 2032 [10].

In the field of “Electric Power System Research”, the *Power Distribution Planning* (PDP) Problem is defined as “finding the most economical solution with the optimal location and size of future substations and/or lines to meet the future demand” [20]. In operations research, this problem can be considered as a specialized network flow optimization problem that is suited to include power distribution network characteristics, such as power losses, the integration of electronic storage systems or charging stations for electric vehicles [3, 51]. Therefore, we refer to the abovementioned problem as the *Distribution Network Optimization Problem* (DNO). It is a strategic problem where a long-term solution is sought. Models and analytical techniques for solving the DNO were reviewed by Khattor and Leung [28], Sempértegui et al. [50], Ehsan and Yang [16], Georgilakis and Hatziaargyriou [20] and Saboori et al. [46]. Although the DNO is a strategic planning problem, which usually needs to be solved only infrequently with low requirements on the computational efficiency of solution methods, the consideration of the efficiency of solving DNO instances deems important for several reasons: (1) Infrastructure planners (e.g., distribution network operators or engineering offices) employ a rolling-horizon planning procedure with frequent recalculations. (2) Facing informational uncertainty, infrastructure planners often employ scenario analyses, which require solving multiple scenarios which differ

in their values of exogenous model parameters, including future energy needs. (3) Planning time and computing resources are limited. (4) The solution space has many local optima, and complexity increases rapidly with increasing network size [21]. Keeping complexity manageable is particularly challenging as distribution network areas may span multiple cities and counties. For instance, the distribution network operator *Westfalen Weser Netz* serves 658,000 customers in an area of 6500 km<sup>2</sup> around Paderborn, Germany. This involves 29,251 km of lines, 108 distribution substations, and 7477 transformer stations.

In general, there are several options for addressing the computational complexity of DNO problem instances: (1) design and application of heuristic procedures, (2) model simplification, and (3) infrastructure decomposition. Although heuristic approaches are popular due to their computational efficiency, a key issue with most heuristic procedures (not only for the DNO problem) lies in the unknown quality of generated partial and complete solutions due to the lack of computationally or analytically identified bounds [1, 11, 61]. DNO model simplification usually involves the simplification of technical assumptions or the application of linearization techniques in order to apply exact solution methods. These approaches often neglect or abstract from the low-voltage part of the distribution network, including voltage quality and the network's behavior during load peaks. However, due to the increasing penetration of low-voltage networks, the consideration of these issues is important [6, 19, 59]. When infrastructure decomposition is applied, the distribution network is decomposed into multiple parts and planning problems, which are then optimized separately. Although merging optimal solutions of subproblems do not necessarily lead to an optimal solution of the overall problem when subproblems are connected to each other, using exact procedures to solve (sub)problems resulting from decomposition allows determining bounds and the quality of found solutions for each subproblem. Thus, the decomposition approach is often useful and addressed in this article.

An infrastructure planner needs to answer two key questions when performing infrastructure decomposition: (1) the type of decomposition: How can the network be decomposed? (2) The scenario analysis design: given a decomposition of a network into subnetworks, how should a scenario analysis for a subnetwork be designed to account for informational uncertainties subject to limited planning time and computing resources? This study focuses on the latter question, which has often been neglected in the literature. When addressing this question, one key issue of computational scenario analysis is the discretization of value ranges for (uncertain) exogenous parameters, such as the future load, available assets, or investment opportunities. Associated with this issue, a second issue is the question of how parameter values should be combined to form a set of scenarios to be computationally analyzed. For example, it may turn out that a full factorial design is computationally intractable due to limitations in time and computing resources.

To guide our scenario analysis design, we study correlations between parameter values of scenarios with their solution times and pursue the following research question: How do scenario characteristics affect the model's computational solution time? Methodologically, we perform extensive computational experiments and a statistical analysis of computation times.

The contributions of our work consist of (1) the formulation of a linear optimization model which considers distribution networks with their technical details, (2) the conduction of extensive computational experiments for determining correlations between parameter values and computation times, and (3) the provision of recommendations for infrastructure planners who wish to perform scenario analysis for planning their power distribution networks.

The remainder of this paper proceeds as follows. In Sect. 2, we survey relevant literature. We provide a problem description and requirements which need to be considered in Sect. 3. We present a novel MILP formulation in Sect. 4 and evaluate our approach in Sect. 5. Then, we discuss our results in Sect. 6 before we conclude in Sect. 7.

## 2 Literature

Distribution network optimization has received a lot of attention in the literature. In Sect. 2.1, we present related research and show the extent to which optimization models ensure supply reliability as well as the effectiveness in terms of the size of solved distribution networks. In Sect. 2.2, we conclude research gaps and situate this study within the existing literature.

### 2.1 Recent research

To present the literature, we use the taxonomy of Georgilakis and Hatziaargyriou [20], who classify studies according to their objective, the problem type, the planning period, and distribution levels. The distinction according to the objective indicates the optimization goals. The problem type can either design a new distribution network or expand an existing one and adapt it to future needs. The planning period can be a static single-stage period, where a solution is determined for a certain demand. A dynamic planning period designs a successive expansion plan for the distribution network, divided into multiple stages. The distribution level indicates whether a primary or secondary distribution network is considered, which maintains medium and low voltage, respectively.

To better address the need for further research, we add the categories of uncertainty management and the network size, that are part of the taxonomy of Saboori et al. [46]. In the area of uncertainty management, we distinguish between reliability and demand forecasting. The reliability indicates whether the distribution network design takes supply interruptions (e.g., due to damaged lines) into account. The demand forecast describes whether the model considers unknown future demand. The network size highlights the number of nodes, that are used for the evaluation in the respective studies. Table 1 summarizes the considered literature.

Distribution network optimization pursues various objectives like minimizing investment and operation cost, decreasing energy losses and emissions, and increasing reliability and social benefit. However, multi-criteria decision-making is not

**Table 1** Classification of related work

	Objective					Type	Period	Level	Uncertainty		Size	Method					
	Investment cost	Operation cost	Energy losses	Unserved energy	Emission	Social benefit	New	Expansion	Static	Dynamic	Primary	Secondary	Reliability	Demand	Nodes	Formulation	Approach <sup>1</sup>
[12]	●	●	●					●	●			●			47	NLP	GA
[42]	●	●	●					●		●	●				54	NLP	TS
[49]	●	●	●	●				●		●	●				52	NLP	PSO
[7]	●	●	●					●		●	●				69	MILP	Heuristic
[63]	●	●	●		●			●		●	●				119	NLP	GA
[32]	●	●	●					●	●		●				36	MILP	Exact
[57]	●	●	●		●			●			●				32	NLP	Exact
[31]	●	●	●					●	●		●				47	NLP	CHA
[26]	●		●					●	●		●				136	MILP	Exact
[40]	●	●	●					●	●		●	●			75	MILP	Exact
[36]	●	●	●				●	●	●		●	●			322	unknown	GA
[37]	●	●	●	●			●	●	●		●	●			812	unknown	GA
[44]	●	●	●	●				●	●		●				84	MILP	SA
[30]	●	●	●	●			●	●	●		●				100	NLP	Seeker
[17]	●	●	●	●			●	●		●	●				9	MILP	Exact
[62]	●	●	●	●				●		●	●				205	NLP	PSO
[18]	●	●	●	●				●		●	●				45	QP	ABC
[22]	●	●	●	●				●		●	●		○		18	MILP	Exact
[27]	●	●	●	●				●		●	●		○		54	MILP	Exact
[38]	●	●	●	●				●		●	●	●	○		5000	NLP	GA
[33]	●	●	●	●				●		●	●		○	○	27	MILP	Exact
[4]	●	●	●	●		●		●		●	●		○	●	26	MILP	Exact
[47]	●	●	●	●	●			●	●		●		○	●	41	MILP	Exact
[51]	●	●	●	●				●	●		●		○	●	18	MILP	Exact
[52]	●							●	●		●			●	18	MILP	Exact
[3]	●	●						●	●		●			●	54	MILP	Exact
[29]	●		●	●			●	●	●		●			●	23	NLP	TS

● is considered,

○ is partially considered

<sup>1</sup>ABC artificial bee colony, CHA constructive heuristic algorithm,

GA genetic algorithm, PSO particle swarm optimization

necessary since the different goals can be valued monetarily. All studies reviewed in Table 1 consider investment costs. Operational costs are sometimes neglected [17, 26, 44, 52] or included in the investment costs [3, 29]. Asensio et al. [4], El-Fouly et al. [17], Haffner et al. [22], Jabr [26], Jooshaki et al. [27], Santos et al. [47] and Shen et al. [51] pay special attention to energy production and loss costs. Energy production and loss costs refer to the conversion of energy, e.g., when substations change the voltage level, or generators and storages feed in energy. Asensio et al. [4], Haffner et al. [22], Jooshaki et al. [27], Lotero and Contreras [33], Santos et al. [47] and Shen et al. [51] also consider the cost of unserved energy. These costs are also called value of lost load and refer to expenses associated with a disruption of supply. Santos et al. [47], Wong et al. [57] and Zidan et al. [63] price CO<sub>2</sub>-emissions and consider them to penalize investments in polluting generators. Asensio et al. [4] pay special attention to the social benefit of a solution in relation to the network's consumers. They design their objective function to consider the consumer's present costs as an available budget for investment and operative costs. The unused budget is maximized to prevent additional costs for consumers.

While few papers in the past newly construct distribution networks [17, 29, 30, 36, 37], most studies extend existing infrastructure. Arias et al. [3] keep a given initial solution and extend the distribution network. In addition to the expansion, Asensio et al. [4] allow the model to upgrade existing substations. Shen et al. [51] consider substations and lines and allow the replacement of existing parts of the distribution. This makes it possible to also scale back the distribution network and to better adapt it to the requirements.

Jabr [26], Li et al. [32], Paiva et al. [40], Santos et al. [47], Shen et al. [51] and Shu et al. [52] perform a static optimization. They calculate a distribution network for one certain time in the future. Prioritization and order of the expansion are left to the decision maker. Approaches with multiple time stages are more complex but offer the advantage of creating a more detailed expansion plan [3, 4, 17, 22, 27, 33].

Mendoza et al. [36], Nazar et al. [38] and Paiva et al. [40] state that the design of the secondary distribution network has an effect on the optimality of the overall design. Synergy effects between the two networks have a strong impact on investment costs. Thus, they consider an integral planning of the primary and secondary network. Cossi et al. [12] and Mendoza et al. [37] focus their work on the secondary distribution network. All other studies in Table 1 evaluate their approaches with primary network instances only and neglect the secondary network.

Another issue of distribution network optimization is the design of a reliable distribution network to prevent supply interruptions. To consider the reliability of the distribution network, the cost of unserved energy can be penalized to prevent unreliable networks [4, 18, 29, 30, 37, 38, 42, 44, 47, 49, 62]. Jooshaki et al. [27] and Lotero and Contreras [33] calculate a system average interruption frequency index to measure reliability and assign an objective coefficient. Instead of ensuring low supply interruptions, reliability can also be ensured by a flexible topology. A flexible topology is characterized by a redundant supply routes: if an interruption of supply occurs in a part of the network, it can be bypassed by changing the circuit and supplying the affected area via other feeders. For example, if one transformer of the network fails, the circuit of the network can be changed and affected customers can be supplied via other transformers. Shen et al. [51] formulate constraints to allow such a flexible topology. In the event of a supply interruption in the network, the topology can thus be flexibly switched and sections affected by supply interruptions can be covered via other routes. However, aforementioned approaches encourage reliability but not enforce it. Failures in the network do not receive sufficient attention, if the predicted costs of interruptions are too low. For this reason, they only partially meet the reliability criterion (cf. Table 1).

To account for different future situations, Asensio et al. [4] formulate their model as a scenario-based stochastic programming model. Arias et al. [3] introduce chance constraints to model uncertain conventional loads and the demand for electric vehicles. Koutsoukis et al. [29], Santos et al. [47], Shen et al. [51] and Shu et al. [52] use scenarios to map different cases of the future. The solutions of the future scenarios are collected into a solution pool. Experts can analyze this pool to make a selection for a final expansion strategy. Lotero and Contreras [33] calculate a solution for each year of the planning period to also provide an orientation in time. However, they provide no distinctive support for the decision maker.

Many studies choose a nonlinear model (QP/NLP) as formulation to model current flows with physical precision [12, 18, 42, 49, 62, 63]. A nonlinear model formulation results in a high problem complexity [55]. Therefore, heuristic methods are applied predominately to networks with a large number of load nodes. For example, Mendoza et al. [37], Nazar et al. [38] and Ziari et al. [62] use heuristic approaches to evaluate distribution networks with 812, 5000, and 205 load nodes, respectively. In the work in Table 1, exact methods are applied only on smaller networks with 136 load nodes or less [26].

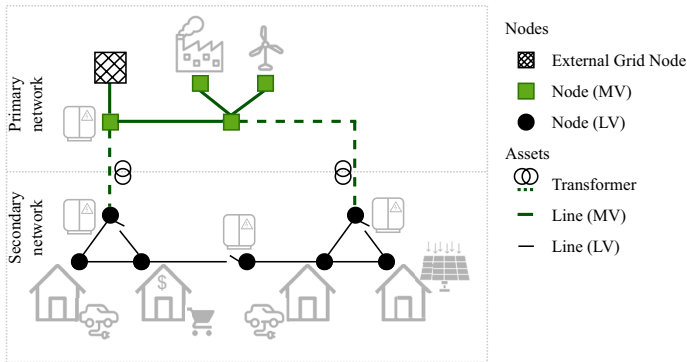
## 2.2 Research gaps

Recent research places various emphases in the area of distribution network optimization (cf. Sect. 2.1). We identify three research gaps that we address in our work: (1) solving large networks with exact optimization methods, (2) integrating the primary and secondary distribution network, and (3) designing flexible topologies for high reliability. (1) The presented studies focus on solving distribution networks with few load nodes exactly or apply heuristic methods when there are many load nodes. However, there are no model formulations that are solved with exact methods for networks with a high number of load nodes. (2) Considering networks with a high number of load nodes is important to also model the secondary network and enable integrated expansion planning. However, the secondary network is predominantly neglected, even though this is just as important for transporting energy to network customers [12, 40]. (3) Previous work has evaluated reliability using costs of undelivered energy. However, no study requires the design of a flexible topology.

To the best of our knowledge, there is no linear formulation that considers the primary and secondary networks in an integrated way, imposes a flexible topology, and can be solved exactly. We wish to extend the research of network optimization and close this research gap by presenting a novel MILP formulation. Our goal is to minimize investment and operation costs of the primary and secondary distribution network using an exact solution approach. In addition to physical requirements, the constraints must also ensure a flexible topology in order to ensure practically applicable results. We want to focus on a static planning horizon; dealing with an uncertain future is still possible by computing multiple future scenarios and consolidating an expansion plan by experts, as shown in other studies.

## 3 Problem description and requirements

The distribution network consists of different feed-in, demand, and switch nodes, and is divided into a *primary* and a *secondary* distribution network. As Fig. 1 shows, one node connects the primary network with the external grid and serves as the source of electricity. The primary network delivers the electricity under medium voltage (MV). It also distributes electricity to secondary networks, which operates a lower voltage (LV) level. Nodes represent producers (e.g., small power plants, photovoltaic panels), consumers (e.g., households, businesses), or junctions (e.g.,



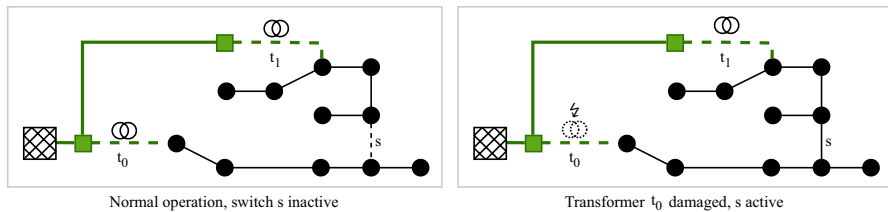
**Fig. 1** Topology of a distribution network

**Table 2** Requirements for the model

	#	Requirement	Description
Physical	I	Current flow balance	Supply and consumption must be equal
	II	Voltage level	The voltage level has to be within specified limits and voltage drop during transportation has to be considered
	III	Current limits	The current flow must not exceed the line's limits
	IV	Transformer loading	Transformers must not be overloaded
Topological	V	Radiality	The network may be meshed, but operates radially
	VI	Connectivity	Every node must be connected and its demand satisfied
	VII	Reliability	The network must have a flexible topology at desired points
	VIII	Logical constraints	Line compatibility and operational issues

distribution boards), that are associated with a demand. Lines and transformers are assets. Lines connect different nodes that have the same voltage level. Transformers connect nodes from different voltage levels. Hence, they represent connections from the primary to the secondary network and vice versa. For connecting two nodes, different transformer types and line types can be installed in the distribution network. Transformer types differ in their capacity of power they can supply to the distribution network, based on the number of windings and other characteristics. Similarly, line types differ in current flow limitations or voltage losses, based on their material and cross-section: For example, a line type made of aluminum with a cross-section of 150 mm<sup>2</sup> can carry a higher current flow and has a better voltage loss behavior than a line type of the same material with a smaller cross-section.

Literature provides requirements for realistic calculation and usability in practice [20, 55]. In Table 2, we present and explain the requirements for our model and distinguish between physical and topological requirements. We discussed the requirements and supplemented them in a workshop with practical users to ensure relevance and completeness.



**Fig. 2** Illustration of flexible topology

Physical requirements (I–IV) affect the current flow and voltage level: according to Kirchhoff's Current Law there has to be a power flow equality (I), such that the current entering a node is exactly equal to the current leaving it. The voltage level (II) has to be within specified limits and the voltage drop has to be considered. Also, the current flow through a line must not exceed its limit (III). Transformers must not be overloaded (IV) with current outside permissible limits, since this can lead to outages [3, 32, 55]

Topological requirements (V–VIII) influence the design of the network: A distribution network's design may be meshed or radial. A meshed network has multiple power sources and uses loops in its circuitry, resulting in better voltage maintenance and reliability. A radial network has only one energy source and distributes the energy using a tree structure. The majority of the world's power distribution networks are operated radially, because they are easier to plan, operate and maintain [13]. In our work, we consider the design of a German distribution network layout that uses a meshed topology, but is operated radially by using sectionalizing switches (V). The distribution network must satisfy the demand of every network customer (VI). For the network customer's supply, we require a high reliability (VII). A high reliability can be achieved by designing a redundant supply, such that the topology can be switched flexibly [32]. A flexible topology allows affected transformers and lines to be bypassed in the event of supply interruptions. Figure 2 illustrates an example of a reliable network whose circuit can be switched in the event of a transformer failure. Finally, logical constraints (VIII) must be met to ensure the operability of the network, e.g., only one asset type can be chosen, and adjacent lines must be compatible [55].

## 4 Optimization model

In this section, we introduce our mathematical model for the DNO. The distribution network is modelled as a directed graph  $G = (V, E)$ . Nodes  $V = \{1, \dots, n\}$  represent points of the network and the set  $E = \{e_1, \dots, e_m\} \subset V \times V$  defines edges  $e_{ij} := (i, j)$  between two nodes  $i$  and  $j$ .

To model the use of different line and transformer types, we introduce node and edge qualifications: node qualifications provide information about the nodes' voltage levels, so we can consider whether two nodes can be connected with a line or a transformer

type. The function  $c : V \rightarrow V_q = \{EG, MV, LV\}$  defines the node qualifications and provides information about whether the node belongs to the primary MV network or the secondary LV network. At least one node is part of the medium voltage network and connected to the external grid (EG) and thus serves as a source for energy. Edge qualifications provide information about the line and transformer types that can be used on an edge. The function  $d : E \rightarrow Q = Q_\lambda \cup Q_\tau$  provides the set of available qualifications for a particular edge. Therefore, the codomain  $Q$  can be subdivided into available line types  $Q_\lambda$  and transformer types  $Q_\tau$ . This means that the function  $d(e)$  reflects the set of potentially installable line and transformation types  $q \in Q$  for an edge  $e \in E$ : If an edge  $e = (i, j) \in E$  connects two nodes  $i, j \in \{i, j | (i, j) \in E \wedge c(i) = c(j)\}$  of the same voltage level,  $d(e)$  is limited to the line qualifications  $Q_\lambda$ . Otherwise, the edge requires a transformer and  $d(e)$  returns the set of transformer qualifications  $Q_\tau$ .

$$\min \sum_{i \in V} \sum_{j \in V} \sum_{q \in Q} c_{ijq} x_{ijq} \quad (1)$$

$$f_{ijs} = 0 \quad s = (i, j), s \in S \quad (2)$$

$$-\sum_q u_q^c x_{ijq} \leq f_{ijs} \leq \sum_q u_q^c x_{ijq} \quad (i, j) \in E, s \in S \quad (3)$$

$$\max_q \{u_q^c\} \varphi_{jis} \leq f_{ijs} \leq \max_q \{u_q^c\} \varphi_{ijs} \quad (i, j) \in E, s \in S \quad (4)$$

$$\sum_{i \in V} \varphi_{ijs} \leq 1 \quad j \in V, s \in S \quad (5)$$

$$\sum_j \eta_{ji} f_{jis} - \sum_j f_{ijs} = d_i \quad (6)$$

$$\{i | i \in V \wedge c(i) \neq EG\}, s \in S \\ l_i^v \leq v_{is} \leq u_i^v \quad i \in V, s \in S \quad (7)$$

$$v_{is} - \Delta_{ijq} f_{ijs} \leq v_{js} + u_i^v (2 - x_{ijq} - \varphi_{ijs} - \varphi_{jis}) \quad (8)$$

$$v_{is} - \Delta_{ijq} f_{ijs} \geq v_{js} - u_i^v (2 - x_{ijq} - \varphi_{ijs} - \varphi_{jis}) \quad (9)$$

$$\{(i, j) | (i, j) \in E \wedge d((i, j)) = Q_\lambda\}, q \in Q_\lambda, s \in S \\ l_q^\theta v_{is} \leq v_{js} + u_i^v (2 - x_{ijq} - \varphi_{ijs} - \varphi_{jis}) \quad (10)$$

$$u_q^\theta v_{is} \geq v_{js} - u_i^v (2 - x_{ijq} - \varphi_{ijs} - \varphi_{jis}) \quad (11)$$

$$\begin{aligned} & \{(i, j) | (i, j) \in E \wedge d((i, j)) = Q_\tau\}, q \in Q_\tau, s \in S \\ & \sum_q x_{ijq} \leq 1 \quad (i, j) \in E, q \in d((i, j)) \end{aligned} \quad (12)$$

$$\sum_q x_{ijq} = 0 \quad (i, j) \in E, q \notin d((i, j)) \quad (13)$$

$$x_{ijq} = x_{klq} \quad (14)$$

$$\begin{aligned} & \{(i, j), (k, l) | (i, j), (k, l) \in \kappa \wedge (i, j) \neq (k, l)\}, \kappa \in K, q \in Q \\ & x_{ijq}, \varphi_{ijs} \in \{0, 1\}, \quad f_{ijs}, v_{vs} \in \mathbb{R} \end{aligned} \quad (15)$$

Mixed-Integer Linear Problem Formulation of the Distribution Network Optimization Problem

The decision variable  $x_{ijq} \in \{0, 1\}$  indicates if edge  $(i, j)$  is qualified by qualification  $q \in d((i, j))$ . The Objective Function (1) minimizes the associated capital expenditures  $c_{ijq}$ . The maintenance of lines and transformers takes place at regular intervals and is usually independent of the amount of energy transported, which is why these costs can be estimated before an investment and included in the capital expenditures  $c_{ijq}$ .

Table 4 elucidates which requirements are considered by which constraint. To design a flexible topology and therefore maintain reliability (Req. VII), the set of failure scenario  $S$  lists edges  $s = (i, j) \in S \subseteq E$ , whose potential failure must be compensated. Variable  $f_{ijs}$  returns the flow of current from node  $i$  to  $j$  in the case that the edge  $s$  is unavailable. Hence, Constraint 2 sets the flow  $f_{ijs}$  equal to zero for  $s = (i, j)$ , which means that any flow from node  $i$  to  $j$  is prohibited when this particular edge is unavailable. If  $f_{ijs}$  is not restricted by  $s$ , the current flow is limited to the bound  $u_q^c$  of its qualification  $q \in d((i, j))$  as stated in Constraint 3 (Req. III and IV). A negative flow of  $f_{ijs}$  means that the current flows in the opposite direction, i.e. from node  $j$  to  $i$ .

To maintain radiality (Req. V), Constraint 4 forces a variable  $\varphi_{ijs}$  to indicate whether there is a positive current flow from node  $i$  to  $j$  during failure scenario  $s$ . In combination, Constraint 5 limits the sum of ingoing edges  $\phi_{ijs}$  for a certain node  $j$  to one. This causes that each node is supplied by only one predecessor, which results in a radial supply.

Concerning flow balance (Req. I), Constraint 6 matches the incoming flow  $f_{jis}$  and outgoing flow  $f_{ijs}$  with the node's demand  $d_i$  for each node  $i \in V$ . The flow balance does not apply for source nodes as they can serve an arbitrarily high demand. Parameter  $\eta_{ij}$  models current losses in the case that edge  $(i, j)$  is serviced by a transformer. This procedure is explained in more detail in Appendix A. Constraint 6 also ensures the connectivity (Req. VI) of all nodes with a demand. For rare real-world cases where a node does not have a demand  $d_i$ , a minimal demand can be assumed to maintain its connectivity.

The voltage value  $v_i$  of node  $i$  has to be within its bounds for each failure scenario  $s \in S$ , as stated in Constraint 7 (Req. II). The model considers the voltage drop that occurs when current flows through an edge. For lines, the model considers a voltage drop factor  $\Delta_{ijq}$  (see Appendix B) that occurs within an edge  $(i, j) \in E$  depending on the qualifying type of line  $q \in Q_k$ . The voltage level  $v_{is}$  of node  $i$  minus the voltage drop  $\Delta_{ijq} f_{ijs}$  has to be equal to the voltage level of the adjacent node  $j$  for each failure scenario  $s \in S$  (Constraints 8 and 9). This constraint only applies if qualifier  $q \in Q$  is active ( $x_{ijq} = 1$ ) and if there is a current flow  $f_{ijs}$  through edge  $(i, j) \in E$  in failure scenario  $s \in S$  ( $\phi_{ijs} = 1 \vee \phi_{jis} = 1$ ). Otherwise, another qualifier  $q \in Q$  is active or there is no current flow, so the constraint's right-hand side is unbounded.

For transformers, Constraints 10 and 11 sets the output voltage  $v_{js}$  of node  $j \in V$  within a certain voltage interval determined by the coefficients  $l_q^\theta$  and  $u_q^\theta$  (see Appendix A). Similar to Constraints 8 and 9, the constraints 10 and 11 are unbounded if qualifier  $q \in Q$  is not active or if there is no current flow  $f_{ijs}$  in failure scenario  $s \in S$ .

As a logical restriction, Constraint 12 defines the limitation that only one qualification  $q$  can be built between two nodes  $i$  and  $j$  at the same time. To meet requirement VIII, set  $K \subset \mathcal{P}(E)$  represents a set of distinct subsets of  $E$ . For each set in  $K$ , the same qualification has to be active (Constraint 14). To summarize, the model considers all imposed requirements. To guarantee the admissibility of computed solutions and their practicality, we present a simulative validation in Sect. 5.2.

The model contains binary as well as continuous variables, and all constraints are convex and linear. Thus, the optimization model is a mixed integer linear problem (MILP). To determine the size of the optimization model, the cardinality of vertices  $|V|$ , edges  $|E|$ , qualifications  $|Q|$ , failure scenarios  $|S|$  and line cluster's union  $|\bigcup K|$  are relevant. The total number of variables is  $3|E| \cdot |S| + |E| \cdot |Q| + |V| \cdot |S|$ . The number of constraints equals  $2|Q| \cdot |E| \cdot |S| + 3|V| \cdot |S| + 2|E| \cdot |S| + |E| \cdot |Q| + |\bigcup K| \cdot |Q|$ .

## 5 Computational experiments

In this section, we present our computational experiments. In Sect. 5.1, we introduce different instances based on empirical data. Solutions are checked for validity in Sect. 5.2. Section 5.3 presents the execution times and a sensitivity analysis of the execution times depending on different distribution network characteristics. The problem is solved with Gurobi 9.1.0 under standard parameterization on a Linux CentOS 7 operating system with an Intel Xeon Gold 6148 CPU with  $16 \times 2.4$  GHz and 190 GB main memory.

### 5.1 Instances and data generation

We create a set of instances with different characteristics to investigate their effects on solvability and execution time. A direct choice of the model's sets and parameters (cf. Table 3) is not suitable to create realistic distribution networks, since the underlying topology would be disregarded. Similarly, installation

**Table 3** Notation for the mathematical formulation

Notation	Description
<i>Sets</i>	
$G$	Directed Graph, representing the distribution network, $G = (V, E)$
$V$	Nodes, representing consumer, prosumer, producer and switches, $V = \{v_1, \dots, v_n\}$
$E$	Edges, representing transformers and lines between nodes, $E = \{e_1, \dots, e_m\}$
$S$	Failure scenarios, representing edges whose failure must not affect the stability of the network, $S \subseteq E$
$Q$	Set of edge qualifications, representing lines and transformer types that may be placed on an edge, $Q = \{q_1, \dots, q_p\}$
$Q_\lambda$	Edge qualifications for lines, representing available line types, $Q_\lambda \subset Q$
$Q_\tau$	Edge qualifications for transformers, representing available transformer types, $Q_\tau \subset Q$
$K$	Line cluster, representing contiguous sections of edges for which the same line type must be selected, $K \subset \mathcal{P}(E)$
<i>Variables</i>	
$x_{ijq}$	Binary indicator, 1 if $q \in Q$ is the qualification of edge $(i, j) \in E$ , 0 otherwise
$\phi_{ijs}$	Binary indicator, 1 if there is a positive current flow from node $i \in V$ to $j \in V$ during failure scenario $s \in S$ , 0 otherwise
$f_{ijs}$	Continuous variable for the current flow in edge $(i, j) \in E$ during failure scenario $s \in S$ , $f_{ijs} \in \mathbb{R}$
$v_{is}$	Continuous variable for the voltage of node $i \in V$ during failure scenario $s \in S$ , $v_{is} \in \mathbb{R}$
<i>Parameters</i>	
$c_{ijq}$	Associated cost of $x_{ijq}$ , representing the costs for installing qualification $q \in Q$ to edge $(i, j) \in E$
$d_i$	Current demand of node $i \in V$
$u_q^c$	Upper bound representing the maximum current flow through edge qualifier $q \in Q$
$l_i^v$	Lower bound representing the minimum voltage of node $i \in V$
$u_i^v$	Upper bound representing the maximum voltage of node $i \in V$
$l_q^\theta$	Lower bound of voltage after transformation through transformer $q \in Q_\tau$
$u_q^\theta$	Upper bound of voltage after transformation through transformer $q \in Q_\tau$
$\eta_{ij}$	Conversion rate of current and voltage in edge $(i, j) \in E$ , representing the current and voltage adjustments between different voltage levels
$\Delta_{ijq}$	Voltage drop factor in edge $(i, j) \in E$ while using line $q \in Q_\lambda$ , representing the voltage drop caused by electrical resistance

**Table 4** Consideration of requirements by constraints

Requirement	I	II	III	IV	V	VI	VII	VIII
Constraint	6	7, 8, 9, 10, 11	3	3	4, 5	6	2	12, 13, 14

costs, utilization limits and other characteristics are dependent on selected line and transformer types. Therefore, we generate realistic distribution networks and describe them with network parameters in Sect. 5.1.1. In Sect. 5.1.2 we describe the correspondence between the network and model parameters.

### 5.1.1 Network parameter

As network parameters, we consider (1) the city network structure  $cn$ , (2) the number of buses  $bu$ , (3) transformers  $tr$  per 100 buses, (4) line and transformer types  $lt$ , (5) the load factor  $lf$ , and (6) the action scope  $as$ , see Table 5.

- (1) To construct the distribution network, we use data from the OpenStreetMap service [23] of ten different city networks  $cn$  in Germany. Since lines align along roads, we use the road network for locating the secondary network lines.
- (2) To allocate buses, we extract five different random subsections per city network with up to  $bu = 5000$  buildings and waypoints.
- (3) We distribute  $tr$  transformers per 100 buses evenly in the network and connect them with the secondary network;  $tr$  varies between 0.75 and 1.25. The primary network connects each transformer with one neighbor and an external grid in the center of the network.
- (4) The value of  $lt$  indicates how many line and transformer types are available for installation. Discussions with distribution network operators revealed that only a few standard types of lines and transformers are in use and available for purchase. We vary  $lt$  between two and four. They are taken from real world and presented in detail in Appendix C.
- (5) Regarding load behavior, we rely on empirical data of energy consumption. Studies predict a change of the future load by  $-17$  to  $34\%$  compared to today [9, 48]. Therefore, we vary a load factor  $lf$  between 0.83 and 1.34, and multiply it with the assigned loads for each bus.
- (6) When planning a real distribution network expansion, an infrastructure planner has to distinguish whether lines and transformer types must or may be replaced, or must remain in place: Assets must be replaced, e.g., when they are worn out. Some assets must not be replaced, e.g., because their costs are still being amortized. All other assets may be replaced. The action scope  $as$  reflects the relative amount of assets that must and may be modified and is varied between 20 and 50%.

To summarize, we distinguish ten city layouts with five network sizes. For each network size  $bu$ , we generate instances with three different load factors  $lf$ , three different transformer distributions  $tr$ , three settings of line and transformer types  $lt$  and three different sizes for the action scope  $as$ . That is a total of 4050 different instances.

### 5.1.2 Correspondence of network and model parameters

This section explains the relationship between network and model parameters. Therefore, we explain how the graph  $G = (V, E)$  as well as the sets  $S$ ,  $Q$  and  $K$  from the mathematical model result from the network parameters. Table 6 provides an overview of the relationships.

Table 5 Parameterization of distribution networks for data generation

Network Parameter	Values	Justification
City network	$cn \in \{\text{Paderborn, Detmold, Bielefeld, Gütersloh, Soest, Dortmund, Kassel, Hannover, Krefeld, Münster}\}$	Acquired German city layouts for the design of the line topology and load allocation
Buses	$bu \in \{1000, 2000, 3000, 4000, 5000\}$	Covering small to medium sized cities or districts of large cities
Transformers	$tr \in \{0.75, 1.00, 1.25\}$	Based on discussions with experts, covering good and bad initial situations
Line and transformer types	$lt \in \{2, 3, 4\}$	Based on discussions with experts
Load factor	$lf \in \{0.83, 1.00, 1.34\}$	Based on optimistic and pessimistic scenarios in literature
Action scope	$as \in \{0.2, 0.35, 0.5\}$	Assumption of different civil engineering capacities

**Table 6** Correspondence of network and model parameters

Correspondence	Description
$ V  = bu + tr + 1$	$V$ contains one node for each bus $bu$ , one associated node for each transformer $tr$ and one for the external grid
$ E  = bu + 2tr + \varepsilon$	$E$ contains one edge for each bus $bu$ and two edges for each transformer $tr$ . Due to the topology of the city network, $E$ can contain $\varepsilon$ further edges
$ Q  = Q_\lambda \cup Q_\tau$ , $ Q_\lambda  =  Q_\tau  = lt$	$Q$ is the union of $Q_\lambda$ and $Q_\tau$ , each containing $lt$ elements. Each element describes a specific line or transformer type
$ S  = tr$	$S$ contains all edges that transform current from the primary to the secondary network, i.e. $tr$ elements
$ K  \leq  E $	$K$ contains sections of edges for which the same line type must be selected

The set  $V$  contains one node for each bus  $bu$ , one associated node for each transformer  $tr$  and one for the external grid. The load of a bus (cf. Sect. 5.1) results in the node's demand  $d_i$  of the mathematical model. For the bounds of the voltage  $l_i^v$  and  $u_i^v$  of the mathematical model we follow German specifications: The voltage must be between  $l_i^v 18000$  and  $u_i^v 22000$  for nodes of the primary network and between  $l_i^v = 360$  and  $u_i^v = 440$  for nodes of the secondary network.

The set  $E$  contains one edge for each bus  $bu$ . Additionally, two edges for each transformer  $tr$  represent the transformer itself as well as the feeder for the transformer in the primary network. Depending on the topology (e.g., due to intersections or junctions) there are few more edges  $\varepsilon$  to model crossings and branches.

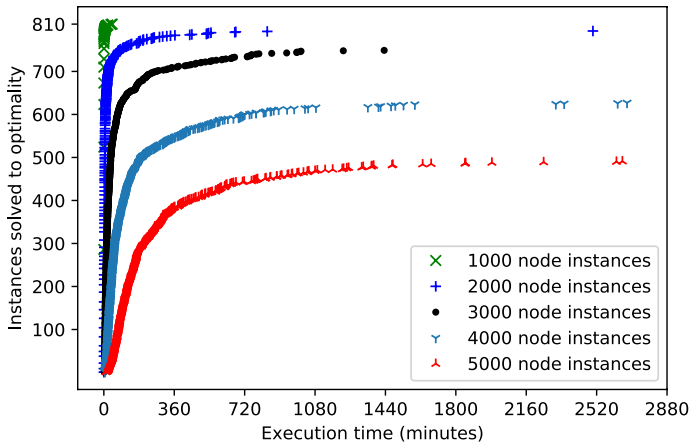
The set  $Q$  contains the qualifications for each edge. The edge qualifications  $Q$  depend on the selected number of line and transformer types  $lt$ . The associated parameters of a line or transformer type result in the model parameters  $c_{ijq}$ ,  $u_q^c$ ,  $l_q^\theta$ ,  $u_q^\theta$  and  $\Delta_{ijq}$ . Appendix C presents the specific model parameters and explains how they are connected to the different line and transformer types.

The set of failure scenarios  $S$  contains all edges that are transformers, as we require all transformers to be failure-safe.

The set  $K$  contains the sets of edges that must be assigned with the same line type. In our instances, we decide to group lines between two intersections or branches. Thus, the set  $K$  is smaller than  $E$ . The exact cardinality depends on the given topology of the city network  $cn$ .

## 5.2 Simulative validation of solutions

As emphasized in the introduction, it is particularly important that solutions can be implemented in practice. To ensure the feasibility of solutions in practice, we simulate all solutions using the current flow simulation library *pandapower*, which is described by Thurner et al. [54]. During simulation, we consider the failure scenarios to verify the reliability of the solution: we successively deactivate the distribution network's transformers to simulate their failure. In this case, the affected transformer cannot be used, and its load is redistributed to other transformers. Appendix E lists detailed simulation results. Overall, 3476 out of 4050 instances are



**Fig. 3** Cactus plot of solved instances

solved, 3473 of them optimal with a gap of 0.5% or less. The power flow calculation converges for all solutions (100.00%). When simulating the 3476 solutions, 0 load capacity violations (0.00%) and 53 voltage violations (1.52%) were detected. The 3423 successfully simulated networks contain 82,141 transformers, which failures were also simulated. The extended failure simulations detected load capacity violations in one case ( $< 0.01\%$ ) and voltage violations in 280 (0.34%) cases. These discrepancies can be explained by the fact that the optimization model designs the most cost-effective solution while exploiting the given bounds as much as possible. Since the simulation calculates current and voltage differently, few violations of the given bounds may occur in some cases. They are of no real consequence due to the fact that in practice, a decision maker could adjust the transformer settings, set the bounds for the optimization more narrowly or, if necessary, specifically strengthen the solution obtained at the affected network parts.

### 5.3 Results

In this section, we present the results of our computational experiments. First, we provide an overview of the optimization's computing time for the instances. Then, we consider the influence of the exogenous network characteristics on the computation time and solution quality using an ordinary least square regression. Finally, we investigate the effect size of the network size on the required computation time using a logistic regression.

Figure 3 presents a cactus plot of the optimization's computing time (wall time) of the instances. The instances are divided into five groups of 810 instances each, according to their number of buses. The cactus plot provides an overview about how efficiently the instances were solved depending on their number of buses. The x-axis shows the computing time in minutes, while the y-axis indicates the number of instances solved optimally in the respective time. We consider an instance to be solved optimally if its gap is 0.5% or less.

**Table 7** Logistic regression results for solvability

Effect		Coef	Std err	<i>t</i>	$P >  t $
Const		35.2395	2.575	13.687	< 0.0001
Buses	<i>bu</i>	− 0.0020	8.69 · 10 <sup>−5</sup>	− 23.401	< 0.0001
Lines per bus	$\frac{nl}{bu}$	− 17.8110	2.101	− 8.479	< 0.0001
Transformers	<i>tr</i>	2.6659	0.304	8.756	< 0.0001
Load per bus	$\frac{td}{bu}$	− 0.9525	0.260	− 3.665	< 0.0001
Line and transformer types	<i>lt</i>	− 0.2173	0.038	− 5.790	< 0.0001
Action scope	<i>as</i>	− 19.5555	1.230	− 15.895	< 0.0001

*N*: 4050 McFadden's Pseudo *R*-squared: 0.4125

To understand the influence of exogenous network characteristics on computing times and solution quality more precisely, we perform a sensitivity analysis using a logistic and an ordinary least square regression: first, for the logistic regression, we introduce the dependent variable *solved*, which is 1, if the solver could calculate an optimal solution during the specified solution time, and 0 otherwise. The logistic regression provides insights into which network characteristics caused an instance to be solved to optimality during a specified time window. We use the network characteristics *bu*, *tr*, *lt* and *as* as independent variables (cf. Table 5) and include the number of lines *nl* as well as the total demand *td* of the network. These values correspond with the mathematical model's sets and parameters as described in Sect. 5.1.2.

To avoid multicollinearities, *nl* and *td* are set in relation to the number of buses. The probability of *solved* given *bu*,  $\frac{nl}{bu}$ , *tr*,  $\frac{td}{bu}$ , *lt* and *as* is stated in Equation (16) with  $\sigma(z) = \frac{1}{1+e^{-z}}$  as the sigmoid function:

$$\begin{aligned} & \text{Prob}\left(\text{solved} = 1 \mid bu, \frac{nl}{bu}, tr, \frac{td}{bu}, lt, ds\right) \\ &= \sigma\left(\beta_0 + \beta_1 \cdot bu + \beta_2 \cdot \frac{nl}{bu} + \beta_3 \cdot tr + \beta_4 \cdot \frac{td}{bu} + \beta_5 \cdot lt + \beta_6 \cdot ds\right) \end{aligned} \quad (16)$$

Table 7 shows the results of the regression. To prevent distortion of the analysis, the instances were checked for outliers according to the Bonferroni correction (error < 5%) and no outliers could be detected.

Second, we use 3473 of 4050 instances which could be solved optimally and perform an OLS regression on the logarithmically transformed computing time *time* as the dependent variable as shown in Eq. (17). Complementary to logistic regression, OLS regression examines the effect size of network characteristics of solved instances on time required. We excluded 11 outliers according to the Bonferroni correction (error < 5%) to prevent distortion of the analysis.

$$\ln(\text{time}) = \beta_0 + \beta_1 \cdot bu + \beta_2 \cdot \frac{nl}{bu} + \beta_3 \cdot tr + \beta_4 \cdot \frac{td}{bu} + \beta_5 \cdot lt + \beta_6 \cdot ds + \varepsilon \quad (17)$$

Table 8 presents the results of the regression. Since the effect of transformers per 100 buses *tr* has no statistical significance, we refine the regression analysis

**Table 8** OLS regression results for computing time

Effect		Coef	Std err	<i>t</i>	<i>P</i> >   <i>t</i>
Const		− 16.3868	1.100	− 14.894	< 0.0001
Buses	<i>bu</i>	0.0017	2.32·10 <sup>−5</sup>	73.303	< 0.0001
Lines per bus	$\frac{nl}{bu}$	15.4091	0.995	15.481	< 0.0001
Transformer	<i>tr</i>	− 0.2110	0.108	− 1.953	0.051
Load per bus	$\frac{ld}{bu}$	0.2621	0.096	2.724	0.006
Line and transformer types	<i>lt</i>	0.1448	0.014	10.600	< 0.0001
Action scope	<i>as</i>	1.7801	0.370	4.817	< 0.0001

*N*: 3462 *R*-squared: 0.773

and group the instances according to the distribution density of transformers. The regression models generated in this process can be found in Appendix D.<sup>1</sup> Appendix E provides detailed results for every instance.

## 6 Discussion

In this section, we discuss the results of the computational experiments. First, we interpret our results and discuss the statistical significance and quality of our sensitivity analysis. Second, we derive implications for practitioners from our findings.

### 6.1 Result interpretations

In this section, we address our research question about how scenario characteristics do affect the model's computation time. We discuss the results of the computation time analysis, the logistic regression and the OLS regression.

In terms of computation time, Fig. 3 shows that as the instance size increases, the average computation time increases and fewer instances can be solved. At the same time, it demonstrates that other network characteristics are also decisive for computing a solution. Within the given computation time limit and hardware resources, all instances with 1000 buses and almost every (98.3%) instance with 2000 buses were solved optimally. With more buses, the number of optimally solved instances decreases to 77.4% and 60.7% for instances with 4000 and 5000 nodes, respectively. These results indicate that networks with more buses tend to be computationally more challenging. However, looking at instances with a computation time of less than 6 h, some instances with 5000 buses were solved in less time than instances with lower bus count. This means that other

<sup>1</sup> In addition to partitioning the data into three different groups, we also generated a regression model with dummy variables for the effect size of the transformers per 100 buses *tr*. Since the results did not yield any further insights, we refrain from presenting them in this study.

characteristics besides the number of buses influence the computational complexity of the model.

We find that the logistic regression results are suitable to guide the decision of whether to consider an instance in a scenario analysis. For example, the decision maker can decide to include an instance that is expected to be solved close to optimality with an expected gap of 0.5% or less. On the other hand, they can decide to omit instances of an insufficiently expected solution quality. The regression results show statistically significant influences on the solvability of an instance for all considered independent variables. To evaluate the model's quality, we calculated McFadden's pseudo  $R^2$ -value. Analogous to the  $R^2$ -index in OLS regressions, McFadden's pseudo  $R^2$  evaluates the predictive capacity of a logistic regression model [53]. A pseudo  $R^2$  of 0.4125 underlines the model's quality, since values between 0.2 and 0.4 are considered as an indicator of an excellent model fit [34, p.54]. The coefficients of the regression model in Table 7 provide insights into the effects of network characteristics on the probability of obtaining an optimal result within 48 hours. When interpreting the coefficients, they need to be considered as a parameter of the sigmoid function  $\sigma(z) = \frac{1}{1+e^{-z}}$ . For example, for a network with  $bu = 5000$  buses,  $\frac{nl}{bu} = 1$  line per bus, and an action scope of  $as = 0.3$ , the probability of solving the network within our experimental setting is  $\sigma(35.2395 - 0.002 \cdot 5000 - 17.811 \cdot 1.0 - 19.5555 \cdot 0.3) \approx 0.83$  (irrespective of the other network characteristics). If the number of lines per bus  $\frac{nl}{bu}$  increases to 1.2 or the action space  $as$  raises to 0.5, the probability reduces to 0.12 and 0.09, respectively. If both variables increase simultaneously, obtaining a solution is hardly expected with a probability of 0.002. Decreasing the network size to 2000 or 1000 would increase the probability to 0.52 and 0.89, respectively, and allows consideration of high line density and action scope again. This behavior can be explained by the fact that with a smaller number of lines, there are fewer options for reliably routing the current flows through the network. Similarly, a smaller action scope allows several variables to be fixed beforehand. Conversely, as the number of lines or action scope increases, the size of the solution space increases rapidly, requiring increased computation time and computing resources.

Looking at the OLS regression, we find that the combination of the number of buses  $bu$ , lines per bus  $\frac{nl}{bu}$ , and the action scope  $as$  have a considerable impact on the computation time. The OLS regression in Table 8 evaluates the influence of network characteristics on the expected computation time and finds statistically significant effects. The  $R^2 = 0.773$  indicates, that the regression model explains 77.3% of the variation in the dependent variable [15, p.18]. Its coefficients show the expected logarithmic computation time depending on the variables. Therefore, the actual computation time is obtained by using the exponential function: For example, if we consider a network with  $bu = 4000$  buses,  $\frac{nl}{bu} = 1$  line per bus,  $tr = 1$  transformer per 100 buses, a demand per bus of  $\frac{ld}{bu} = 0.8$ ,  $lt = 4$  line and transformer types, and an action scope of  $as = 0.4$ , the expected computation time is  $time = e^{-16.3868+0.0017 \cdot 4000+15.4091 \cdot 1.0-0.2110 \cdot 1.0+0.2621 \cdot 0.8+0.1448 \cdot 4+1.7801 \cdot 0.4} \approx 1227$  s ( $\hat{=}$  0.34h). If the action scope increases by 20% to  $as = 0.6$ , the expected

computation time increases by a factor of 1.15 ( $time \approx 1415 \text{ s} \hat{=} 0.39 \text{ h}$ ). An increase of the number of buses  $bu$  to 4800 results in an increase of the computation time to a total of  $time = 4780 \text{ s} (\hat{=} 51.33 \text{ h})$ , an increase by the factor of 3.9). An increase of the lines per bus by 20% to  $\frac{nl}{bu} = 1.2$  leads to an expected computation time of 26,743 s ( $\hat{=} 7.42 \text{ h}$ , an increase by the factor of 21.8). Thus, the example illustrates that the choice and combination of individual parameters can greatly change the computation time. The increase in line and transformer types have only a weak effect on the computation time due to a small effect size and a low coefficient, which is why we do not go into detail on this network characteristic. We also neglect the density of the transformers and the load per bus due to their weak significance. The results confirm the logistic regression in the sense that the lines per bus, number of buses, number of lines and transformer types as well as the action space influence the optimization model's complexity. However, for the share of transformers in the network and the load per bus, the OLS regression confirms only weakly significant influences.

Since the value  $P > |t|$  is least significant for transformers, we divided the data into instances with low, medium, and high density of transformers in the network and created three additional OLS regression models, which are presented in detail in Appendix D. These regression models show that no statistical significance can be found for the action scope when focusing on low transformer density ( $tr \leq 0.9$ ), while this feature significantly affects the computation time for instances with high density of transformers ( $tr > 1.1$ ). The opposite is found for the variable of the load per node, which is only weakly significant at low transformer densities and not significant otherwise. Compared to the OLS regression in Table 8, the coefficients reveal a higher effect size of the action scope (3.7135) and lines per bus (16.2187). The coefficients of the other variables do not change or show weaker effects. Overall, the results complement the logistic regression findings and provide further insight into the impact of individual network characteristics on computation time. Due to the high quality of the models and the statistical significance of their variables, we derive implications for practitioners in the following section.

## 6.2 Implications for practitioners

Our optimization model enables practitioners to calculate optimal distribution networks that can be deployed in practice. The results of our computational experiments provide important insights into the design of scenario analyses for distribution network planning. The implications are summarized in Table 9 and discussed in this section.

We find that decomposition should divide distribution networks into small contiguous areas (cf. Implication P1). When infrastructure planners use decomposition techniques, they are guided by geographical contexts, e.g., cities, city districts or neighborhoods. Although the regression results for the number of buses show the smallest coefficients (0.0017), due to the large effect sizes with up to 5000 buses, this network characteristic has a considerable influence on the computation time. Figure 3 additionally illustrates the increase in complexity and shows that with

**Table 9** Implications for practitioners

#	Implication for practitioners
P1	When network planners use decomposition techniques, an orientation towards geographical contexts is usually used. The results show that the number of buses has a considerable influence on the solution time. We recommend network planners to decompose the networks as far as possible into city districts or neighborhoods. This ensures that interdependencies between neighboring supply areas are respected and that complexity is kept low at the same time
P2	When network planners consider a scenario, they have to determine an action scope, which means that they need to decide which assets will be subject of expansion planning. Results show that the action scope disrupts the solvability of the model the most. Therefore, we recommend network planners to set the action scope cautiously
P3	When network planners analyze a network, the ratio of lines to buses is highly relevant to the computing time budget. Since the lines per bus are given exogenously, the network planner cannot reduce the number of lines per bus. We recommend network planners to choose a suitable decomposition or an appropriate scope for action to counteract an unacceptably high computing time, when there is a high amount of lines per bus
P4	When network planners assemble potential new investments, the results show a reciprocal relationship between the number of transformers and the expected solvability or computing time of the model. This means that additional transformers are conducive to solvability. Therefore, we recommend network planners to give extensive thought to possible locations for new transformers in order to benefit from the positive effects mentioned
P5	When network planners consider which asset types for lines and transformers to use in the future, the results of the regression analyzes show that the effect sizes of the bus load and the number of alternatives for asset types have only a minor impact on solvability and solving time. An increase in the load per node and the asset types would only extend the expected computing time slightly. We recommend network planners to consider the possible options for asset types comprehensively in order to create a suitable basis for determining expansion solutions

increasing network size, fewer instances were solved to optimality. We recommend infrastructure planners to decompose the networks as far as possible into city districts or neighborhoods. These small, but connected network areas have the advantage that the complexity is kept low without neglecting dependencies and synergy effects between neighboring distribution areas of different transformers.

We find that the action scope has a great influence on the solvability of the distribution network's model (cf. Implication P2). When infrastructure planners consider a scenario, they have to determine an action scope, which means that they need to decide which assets will be subject of expansion planning. Typically, ailing lines have to be replaced, while recent purchases are still being written off and therefore remain unchanged. For other assets, the infrastructure planners are free to decide whether they should be part of the action scope. According to the results, the action scope has the greatest influence on the solvability of a scenario. If the expected solvability is too improbable, we recommend reducing the scope of action if possible.

We find that the ratio of lines to buses is significant for the computation time and has the greatest influence on it (cf. Implication P3). For example, increasing the ratio of lines to buses by one percent would increase computation time by almost 17%. An increase in the ratio of 10% would result in an increase in computation time of almost 470%. Due to the strong effect size, the necessary computation time can quickly exceed the time budget. Since the lines per bus are given exogenously, the

infrastructure planner cannot reduce the number of lines per bus. To avoid an unacceptably high computation time in case of a too high density of lines, infrastructure planners may choose a suitable decomposition or an appropriate scope for action (cf. Implications P1 and P2).

We find that a higher number of alternatives for transformers had a reciprocal relationship between the number of transformers and the expected solvability or computation time of the model (cf. Implication P4). Contrary to the expectation that an increase in the number of alternatives for transformers and the accompanying increase in the solution space create a more complex model, it is shown here that additional transformers are conducive to solvability. The foremost cause for this rather contradictory result is likely the fact that additional transformers prevent expensive expansion of the lines, so the solver can exclude uneconomical solutions more quickly. We recommend infrastructure planners to give extensive thought to possible uses for new transformers to benefit from the positive effects mentioned.

We find that the number of asset types and the load per bus is of less importance to ensure solvability or fast computation time (cf. Implication P5). When infrastructure planners consider which asset types for lines and transformers to use in the future, the results of the regression analyzes show that the effect sizes of the bus load and the number of alternatives for asset types have only a minor impact on solvability and computation time. Due to the disproportionately low ratio, it is recommendable to not restrict the number of asset types and bus load in the scenario design to reduce complexity.

## 7 Conclusion

In this section, we summarize our work and provide an outlook on future research avenues.

### 7.1 Summary

This study addresses the challenge of planning future power distribution networks and designing scenario analyses in order to account for informational uncertainties subject to limited planning time and computing resources. We pursue the research question of how scenario characteristics affect the model's computation time to analyze correlations between parameter values of scenarios with their computation times. Based on requirements from literature and discussions with infrastructure planners, we propose a linear model for the distribution network optimization problem. Using an exact state-of-the-art solver, we conduct extensive computational experiments on distribution network instances based on empirical data, varying several parameters to adjust different network characteristics. We perform a simulative validation to confirm the solution's applicability in practice. We also examine the influence of network characteristics on the computation time in more detail with a sensitivity analysis. From the results, we derive implications for practitioners to guide them in how to design a scenario analysis with respect to limited time and

computing resources. This enables infrastructure planners to consider the uncertain development of the future energy consumption and to determine a cost-effective expansion strategy for their distribution networks.

## 7.2 Research agenda

We are aware of the limitations of our work and advise avenues for future research. As a first research path, we recommend integrating electrical storage systems (ESS) into the model and thereby contribute to identify potentials for smart distribution network expansion. For example, the optimization of distribution networks including ESS is considered in the studies of Asensio et al. [4] or Mariut and Helerea [35]. Since our approach is tailored for German distribution system operators, who are not yet authorized to use flexible assets due to German legislation, we neglected ESS in this study. However, ESS could be integrated as controllable loads in our model with their charging and discharging as a new decision variable. Load losses and voltage behavior would have to be taken into account in the constraints.

A second research path may extend our approach to include planning periods, thus open up the possibility of designing transformation plans for distribution networks in temporal context. The integration of time intervals can be found in the work of Arias et al. [3] or Shen et al. [51]. To model the temporal intervals, the decision variables can be extended by a temporal dimension. Adding another dimension would increase the number of variables and thus the complexity of the model; on the other hand, adding another dimension could result in a better handling of the uncertain future: Considering planning periods, a rolling horizon approach could be applied so that the optimization problem is solved for a specified forecast period and a solution is implemented. A new optimization for the subsequent planning period can then be performed at a later point in time with adjusted demands based on better knowledge about the change in load demand.

A third research path may deal with the computing resources to be used. From an experimental point of view, our study was limited to the effects of network characteristics. Supplementary, further research can determine how varying computational resources, e.g., memory and threads, affect the solver's performance. This approach would extend our computational experiments by varying not only the network characteristics, but also the resources available. An evaluation of solution times as well as solution quality in dependence on the computational resources could provide further information about which requirements should be placed on systems in order not to acquire expensive resources unnecessarily.

A fourth research path could shed light on the nonlinear current behavior. While we formulated a linear model and simulatively validated our results for it, other researchers like Wang et al. [56] or Aldarajee et al. [2] formulate nonlinear models. The nonlinear consideration accurately maps the current behavior, but worsens the performance of the solution calculation. More in-depth experiments can address the question to what extent the optimum of the linear model formulation deviates from the optimum of the nonlinear one and whether consequences for the decision recommendation arise from this.

Finally, as a fifth research path, our results should be further analyzed with real distribution networks. For example, in the work of Bagheri et al. [5] and Paiva et al. [40], a real distribution network is used for evaluation. In a case study, our approach could be used to analyze how an infrastructure decomposition can be used to enforce large distribution networks in a cost-minimizing way to meet the future needs. These insights can guide infrastructure planners on how to decompose their networks to enable coherent planning of different sub-networks that, when combined, represent the entire service area.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12667-023-00572-5>.

**Acknowledgements** The authors gratefully acknowledge the financial support provided by the European Regional Development Fund (ERDF) in the framework of FlexiEnergy (project number 0801186, EU-2-1-028) and the funding of this project by computing time provided by the Paderborn Center for Parallel Computing (PC<sup>2</sup>).

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data availability** The data that support the findings of this study are available from the corresponding author upon request.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ahmadian, A., Sedghi, M., Aliakbar-Golkar, M., et al.: Optimal probabilistic based storage planning in tap-changer equipped distribution network including pevs, capacitor banks and wdgs: a case study for Iran. *Energy* **112**, 984–997 (2016). <https://doi.org/10.1016/j.energy.2016.06.132>
2. Aldarajee, A.H., Hosseini, S.H., Vahidi, B., et al.: A coordinated planner-disaster-risk-averse-planner investment model for enhancing the resilience of integrated electric power and natural gas networks. *Int. J. Electr. Power Energy Syst.* **119**(105), 948 (2020). <https://doi.org/10.1016/j.ijepes.2020.105948>
3. Arias, N.B., Tabares, A., Franco, J.F., et al.: Robust joint expansion planning of electrical distribution systems and ev charging stations. *IEEE Trans. Sustain. Energy* **9**(2), 884–894 (2018). <https://doi.org/10.1109/TSTE.2017.2764080>
4. Asensio, M., de Quevedo, P.M., Muñoz-Delgado, G., et al.: Joint distribution network and renewable energy expansion planning considering demand response and energy storage-part i: Stochastic programming model. *IEEE Trans. Smart Grid* **9**(2), 655–666 (2018). <https://doi.org/10.1109/TSG.2016.2560339>
5. Bagheri, A., Monsef, H., Lesani, H.: Integrated distribution network expansion planning incorporating distributed generation considering uncertainties, reliability, and operational conditions. *Int. J. Electr. Power Energy Syst.* **73**, 56–70 (2015). <https://doi.org/10.1016/j.ijepes.2015.03.010>

6. Bennett, C.J., Stewart, R.A., Lu, J.W.: Development of a three-phase battery energy storage scheduling and operation system for low voltage distribution networks. *Appl. Energy* **146**, 122–134 (2015). <https://doi.org/10.1016/j.apenergy.2015.02.012>
7. Bin Humayd, A.S., Bhattacharya, K.: Comprehensive multi-year distribution system planning using back-propagation approach. *IET Gener. Trans. Distrib.* **7**(12), 1415–1425 (2013). <https://doi.org/10.1049/iet-gtd.2012.0706>
8. Brown, D., Hall, S., Davis, M.E.: Prosumers in the post subsidy era: an exploration of new prosumer business models in the uk. *Energy Policy* **135**(110), 984 (2019). <https://doi.org/10.1016/j.enpol.2019.110984>
9. Bründlinger, T., König, J.E., Frank, O., et al.: dena-leitstudie integrierte energiewende: Impulse für die gestaltung des energiesystems bis 2050. Deutsche Energie-Agentur GmbH (dena), ewi Energy Research & Scenarios gGmbH: Berlin/Köln, Germany (2018)
10. Büchner, J., Katzfey, J., Flörcken, O., et al.: Moderne Verteilernetze für Deutschland (Verteilernetzstudie). Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi), 1st edn, Bonn (2014)
11. Carpinelli, G., Celli, G., Mocci, S., et al.: Optimal integration of distributed energy storage devices in smart grids. *IEEE Trans. Smart Grid* **4**(2), 985–995 (2013). <https://doi.org/10.1109/TSG.2012.2231100>
12. Cossi, A.M., Romero, R., Mantovani, J.R.S.: Planning of secondary distribution circuits through evolutionary algorithms. *IEEE Trans. Power Deliv.* **20**(1), 205–213 (2005). <https://doi.org/10.1109/TPWRD.2004.839229>
13. Cristian, N., Al Ameri Ahmed, B.D.: Impact analysis of distributed generation on mesh and radial distribution network. *Overview State Art Fuel* **1**, MW 5 (2013)
14. Domínguez-Navarro, J., Dufo-López, R., Yusta-Loyo, J., et al.: Design of an electric vehicle fast-charging station with integration of renewable energy and storage systems. *Int. J. Electr. Power Energy Syst.* **105**, 46–58 (2019). <https://doi.org/10.1016/j.ijepes.2018.08.001>
15. Dougherty, C.: Introduction to Econometrics. Oxford University Press, Oxford (2011)
16. Ehsan, A., Yang, Q.: Optimal integration and planning of renewable distributed generation in the power distribution networks: a review of analytical techniques. *Appl. Energy* **210**, 44–59 (2018). <https://doi.org/10.1016/j.apenergy.2017.10.106>
17. El-Fouly, T., Zeineldin, H., El-Saadany, E., et al.: A new optimization model for distribution substation siting, sizing, and timing. *Int. J. Electr. Power Energy Syst.* **30**(5), 308–315 (2008). <https://doi.org/10.1016/j.ijepes.2007.10.002>
18. El-Zonkoly, A.M.: Multistage expansion planning for distribution networks including unit commitment. *IET Gener. Trans. Distrib.* **7**(7), 766–778 (2013). <https://doi.org/10.1049/iet-gtd.2012.0289>
19. García-Villalobos, J., Zamora, I., Knezović, K., et al.: Multi-objective optimization control of plug-in electric vehicles in low voltage distribution networks. *Appl. Energy* **180**, 155–168 (2016). <https://doi.org/10.1016/j.apenergy.2016.07.110>
20. Georgilakis, P.S., Hatziaargyriou, N.D.: A review of power distribution planning in the modern power systems era: models, methods and future research. *Electric Power Syst. Res.* **121**, 89–100 (2015). <https://doi.org/10.1016/j.epr.2014.12.010>
21. Gitizadeh, M., Vahed, A.A., Aghaei, J.: Multistage distribution system expansion planning considering distributed generation using hybrid evolutionary algorithms. *Appl. Energy* **101**, 655–666 (2013). <https://doi.org/10.1016/j.apenergy.2012.07.010>
22. Haffner, S., Pereira, L.F.A., Pereira, L.A., et al.: Multistage model for distribution expansion planning with distributed generation-part i: Problem formulation. *IEEE Trans. Power Deliv.* **23**(2), 915–923 (2008). <https://doi.org/10.1109/TPWRD.2008.917916>
23. Haklay, M., Weber, P.: Openstreetmap: user-generated street maps. *IEEE Pervas. Comput.* **7**(4), 12–18 (2008). <https://doi.org/10.1109/MPRV.2008.80>
24. Heathcote, M.: J & P Transformer Book. Elsevier, New York (2011)
25. Höflisch, B., Richard, P., Völker, J., et al.: Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030 (2012)
26. Jabr, R.A.: Polyhedral formulations and loop elimination constraints for distribution network expansion planning. *IEEE Trans. Power Syst.* **28**(2), 1888–1897 (2013). <https://doi.org/10.1109/TPWRS.2012.2230652>
27. Jooshaki, M., Abbaspour, A., Fotuhi-Firuzabad, M., et al.: Milp model of electricity distribution system expansion planning considering incentive reliability regulations. *IEEE Trans. Power Syst.* **34**(6), 4300–4316 (2019). <https://doi.org/10.1109/TPWRS.2019.2914516>

28. Khator, S.K., Leung, L.C.: Power distribution planning: a review of models and issues. *IEEE Trans. Power Syst.* **12**(3), 1151–1159 (1997). <https://doi.org/10.1109/59.630455>
29. Koutsoukis, N., Georgilakis, P., Hatziaargyriou, N.: 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, pp 1–6. <https://doi.org/10.1109/PMAPS.2014.6960627> (2014)
30. Kumar, D., Samantaray, S.: Design of an advanced electric power distribution systems using seeker optimization algorithm. *Int. J. Electr. Power Energy Syst.* **63**, 196–217 (2014). <https://doi.org/10.1016/j.ijepes.2014.05.073>
31. Lavorato, M., Rider, M.J., Garcia, A.V., et al.: A constructive heuristic algorithm for distribution system planning. *IEEE Trans. Power Syst.* **25**(3), 1734–1742 (2010). <https://doi.org/10.1109/TPWRS.2009.2038164>
32. Li, Z., Wu, W., Zhang, B., et al.: An milp model for urban distribution network planning considering street layout and block loads. In: 2019 IEEE Power and Energy Society General Meeting (PESGM), IEEE, pp. 1–5. <https://doi.org/10.1109/PESGM40551.2019.8973739> (2019)
33. Lotero, R.C., Contreras, J.: Distribution system planning with reliability. *IEEE Trans. Power Deliv.* **26**(4), 2552–2562 (2011). <https://doi.org/10.1109/TPWRD.2011.2167990>
34. Louviere, J.J., Hensher, D.A., Swait, J.D.: *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge (2000). <https://doi.org/10.1017/CBO9780511753831>
35. Mariut, L., Helerea, E.: Multiple stress life analysis on underground power cables from distribution networks. In: *Doctoral Conference on Computing, Electrical and Industrial Systems*. Springer, pp. 275–282. [https://doi.org/10.1007/978-3-642-28255-3\\_30](https://doi.org/10.1007/978-3-642-28255-3_30) (2012)
36. Mendoza, J.E., López, M.E., Pena, H.E., et al.: Low voltage distribution optimization: site, quantity and size of distribution transformers. *Electr. Power Syst. Res.* **91**, 52–60 (2012). <https://doi.org/10.1016/j.epsr.2012.05.004>
37. Mendoza, J.E., López, M.E., Fingerhuth, S.C., et al.: Low voltage distribution planning considering micro distributed generation. *Electr. Power Syst. Res.* **103**, 233–240 (2013). <https://doi.org/10.1016/j.epsr.2013.05.020>
38. Nazar, M.S., Haghifam, M.R., Nazar, M.: A scenario driven multiobjective primary-secondary distribution system expansion planning algorithm in the presence of wholesale-retail market. *Int. J. Electr. Power Energy Syst.* **40**(1), 29–45 (2012). <https://doi.org/10.1016/j.ijepes.2012.02.002>
39. Nicolli, F., Vona, F.: Energy market liberalization and renewable energy policies in oecd countries. *Energy Policy* **128**, 853–867 (2019). <https://doi.org/10.1016/j.enpol.2019.01.018>
40. Paiva, P., Khodr, H., Dominguez-Navarro, J., et al.: Integral planning of primary-secondary distribution systems using mixed integer linear programming. *IEEE Trans. Power Syst.* **20**(2), 1134–1143 (2005). <https://doi.org/10.1109/TPWRS.2005.846108>
41. Panwar, N., Kaushik, S., Kothari, S.: Role of renewable energy sources in environmental protection: a review. *Renew. Sustain. Energy Rev.* **15**(3), 1513–1524 (2011). <https://doi.org/10.1016/j.rser.2010.11.037>
42. Pereira Junior, B.R., Cossi, A.M., Contreras, J., et al.: Multiobjective multistage distribution system planning using tabu search. *IET Gener. Trans. Distrib.* **8**(1), 35–45 (2014). <https://doi.org/10.1049/iet-gtd.2013.0115>
43. Petrović, S.N., Karlsson, K.B.: Residential heat pumps in the future Danish energy system. *Energy* **114**, 787–797 (2016). <https://doi.org/10.1016/j.energy.2016.08.007>
44. Popović, Ž, Kerleta, V.D., Popović, D.: Hybrid simulated annealing and mixed integer linear programming algorithm for optimal planning of radial distribution networks with distributed generation. *Electr. Power Syst. Res.* **108**, 211–222 (2014). <https://doi.org/10.1016/j.epsr.2013.11.015>
45. Rodríguez, L.R., Ramos, J.S., Domínguez, S.A., et al.: Contributions of heat pumps to demand response: a case study of a plus-energy dwelling. *Appl. Energy* **214**, 191–204 (2018). <https://doi.org/10.1016/j.apenergy.2018.01.086>
46. Saboori, H., Hemmati, R., Ghiasi, S.M.S., et al.: Energy storage planning in electric power distribution networks—a state-of-the-art review. *Renew. Sustain. Energy Rev.* **79**, 1108–1121 (2017). <https://doi.org/10.1016/j.rser.2017.05.171>
47. Santos, S.F., Fitiwi, D.Z., Shafie-Khah, M., et al.: New multistage and stochastic mathematical model for maximizing res hosting capacity-part i: Problem formulation. *IEEE Trans. Sustain. Energy* **8**(1), 304–319 (2016). <https://doi.org/10.1109/TSTE.2016.2598400>
48. Schlesinger, M., Hofer, P., Kemmler, A., et al.: *Entwicklung der energiemärkte—energiereferenzprognose. Projekt Nr 57/12 Studie im Auftrag des Bundesministeriums für Wirtschaft und Technologie Basel/Köln/Osnabrück: ewi/gws/prognos* (2014)

49. Sedghi, M., Aliakbar-Golkar, M., Haghifam, M.R.: Distribution network expansion considering distributed generation and storage units using modified pso algorithm. *Int. J. Electr. Power Energy Syst.* **52**, 221–230 (2013). <https://doi.org/10.1016/j.ijepes.2013.03.041>
50. Sempértégui, R., Bautista, J., Cubero, R.G., et al.: Models and procedures for electric energy distribution planning: a review. *IFAC Proc. Vol.* **35**(1), 395–400 (2002). <https://doi.org/10.3182/20020721-6-ES-1901.01217>
51. Shen, X., Shahidehpour, M., Zhu, S., et al.: Multi-stage planning of active distribution networks considering the co-optimization of operation strategies. *IEEE Trans. Smart Grid* **9**(2), 1425–1433 (2018). <https://doi.org/10.1109/TSG.2016.2591586>
52. Shu, J., Wu, L., Li, Z., et al.: A new method for spatial power network planning in complicated environments. *IEEE Trans. Power Syst.* **27**(1), 381–389 (2011). <https://doi.org/10.1109/TPWRS.2011.2161351>
53. Smith, T.J., McKenna, C.M.: A comparison of logistic regression pseudo r2 indices. *Multiple Linear Regress. Viewpoints* **39**(2), 17–26 (2013)
54. Thurner, L., Scheidler, A., Schäfer, F., et al.: pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Trans. Power Syst.* **33**(6), 6510–6521 (2018). <https://doi.org/10.1109/TPWRS.2018.2829021>
55. Vahidinasab, V., Tabarzadi, M., Arasteh, H., et al.: Overview of electric energy distribution networks expansion planning. *IEEE Access* **8**, 34750–3469 (2020). <https://doi.org/10.1109/ACCESS.2020.2973455>
56. Wang, D., Qiu, J., Meng, K., et al.: Coordinated expansion co-planning of integrated gas and power systems. *J. Mod. Power Syst. Clean Energy* **5**(3), 314–325 (2017). <https://doi.org/10.1007/s40565-017-0286-z>
57. Wong, S., Bhattacharya, K., Fuller, J.D.: Electric power distribution system design and planning in a deregulated environment. *IET Gener. Trans. Distrib.* **3**(12), 1061 (2009). <https://doi.org/10.1049/iet-gtd.2008.0553>
58. Yang, H., Cui, P., Fang, Z.: Vertical-borehole ground-coupled heat pumps: a review of models and systems. *Appl. Energy* **87**(1), 16–27 (2010). <https://doi.org/10.1016/j.apenergy.2009.04.038>
59. Yu, K., Guo, W., Chen, X., et al.: Research on dynamic control of low-voltage distribution network with high penetration of electric heat pumps based on  $\mu$ pmu measurements and data-driven approach. *IET Gener. Trans. Distrib.* **14**(20), 4426–4434 (2020). <https://doi.org/10.1049/iet-gtd.2020.0864>
60. Zafar, R., Mahmood, A., Razzaq, S., et al.: Prosumer based energy management and sharing in smart grid. *Renew. Sustain. Energy Rev.* **82**, 1675–1684 (2018). <https://doi.org/10.1016/j.rser.2017.07.018>
61. Zheng, Y., Dong, Z.Y., Luo, F.J., et al.: Optimal allocation of energy storage system for risk mitigation of discos with high renewable penetrations. *IEEE Trans. Power Syst.* **29**(1), 212–220 (2013). <https://doi.org/10.1109/TPWRS.2013.2278850>
62. Ziari, I., Ledwich, G., Ghosh, A., et al.: Optimal distribution network reinforcement considering load growth, line loss, and reliability. *IEEE Trans. Power Syst.* **28**(2), 587–597 (2012). <https://doi.org/10.1109/TPWRS.2012.2211626>
63. Zidan, A., Shaaban, M.F., El-Saadany, E.F.: Long-term multi-objective distribution network planning by dg allocation and feeders' reconfiguration. *Electr. Power Syst. Res.* **105**, 95–104 (2013). <https://doi.org/10.1016/j.epsr.2013.07.016>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Sascha Christian Burmeister<sup>1</sup>  · Guido Schryen<sup>1</sup>

Guido Schryen  
guido.schryen@upb.de

<sup>1</sup> Paderborn University, Warburger Str. 100, 33098 Paderborn, NRW, Germany