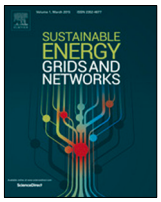
Sustainable Energy, Grids and Networks 34 (2023) 100999

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan



Stochastic optimization and Markov chain-based scenario generation for exploiting the underlying flexibilities of an active distribution network

Mohammad Rayati [a,*], Mokhtar Bozorg [a], Mauro Carpita [a], Rachid Cherkaoui [b]

[a] School of Engineering and Management Vaud, HES-SO University of Applied Sciences and Arts Western Switzerland, Yverdon-les-Bains, Switzerland
[b] EPFL, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 21 June 2022
Received in revised form 23 December 2022
Accepted 8 January 2023
Available online 11 January 2023

Keywords:
Active distribution network (ADN)
Battery energy storage (BES)
Flexibilities
Photovoltaic (PV)
Stochastic optimization

ABSTRACT

This paper proposes a scalable stochastic optimization model and a Markov chain-based scenario generation method to benefit from an active distribution network's (ADN's) flexibility. The optimization variables are the dispatch plan, such as the active and reactive power of battery energy storage (BES) and photovoltaic (PV) systems, as well as the active and reactive power and flexibilities given to the transmission network at the point of common coupling (PCC). The uncertainty vector, on the other hand, is made up of the PV system's production capability, electricity demands, the flexibility request of the transmission system operator (TSO), and the voltage at the PCC. The resulting stochastic optimization problem is a second-order cone programming (SOCP) problem that is solved using freely available convex solvers. To validate the performance of the proposed stochastic optimization, the tests were carried out in a laboratory, where a flexible structure mimics different distribution network topologies, such as a real low-voltage radial one in Switzerland.

1. Introduction

With the integration and expansion of renewable energy into the electric power sector, network operators at both the distribution and transmission levels must optimize the active and reactive power imbalance settlement processes. The operators at the distribution level can take advantage of the potential of distributed energy resources (DERs) for imbalance settlement because active distribution networks (ADNs) can accept significant amounts of renewable energy [1]. In this context, we are referring to DERs including battery energy storage (BES) and photovoltaic (PV) systems. Furthermore, we are referring to the underlying flexibilities as the ability to regulate the active and reactive power of DERs and at the point of common coupling (PCC) of the ADN.

The ADN's operators face the following challenges to optimize the operation of DERs and utilize the underlying flexibilities:

(i) They must respect the constraints of ADN's components (such as lines and transformers), as well as the security and quality of the power supply (see [2]).
(ii) They must deal with the uncertainties associated with electricity demands and PV power production capabilities [3]. In this regard, they derive a number of realistic operation scenarios, in which they summarize the patterns of electricity demands and PV power production capabilities.
(iii) They would like to be participants in the wholesale flexibility markets. In such a case, they are obligated to provide planned flexibilities at the PCC in response to the request of the transmission system operator (TSO) [4]. On the other hand, operators at the distribution level cannot accurately model the operation of the transmission network. As a result, they must treat the request of TSO for deploying flexibilities as an uncertain parameter.

Recent studies have emphasized the importance of scheduling and exploiting ADN's underlying flexibility. To summarize, two types of research have been conducted: (i) those that have focused on modeling the components, constraints, and uncertainties of ADNs in order to provide flexibility; and (ii) those that have proposed a mechanism or an algorithm for solving the scheduling optimization problem and exploiting the ADNs' underlying flexibilities.

A deterministic model for optimizing the flexibilities provided by DERs in ADN is presented in [5]. A framework for dispatching the power of an ADN has been developed in [6], with a BES system serving as a flexible component. Ref. [7] proposed a framework for stochastic co-optimization of primary frequency control supply by a BES system. In [8], a control framework for BES systems in ADNs based on a data-driven model has been presented for

* Corresponding author.
E-mail address: mohammad.rayati@heig-vd.ch (M. Rayati).

https://doi.org/10.1016/j.segan.2023.100999

Nomenclature

The main notations are defined in the following. Other symbols are defined as needed throughout the text.

Indices and Sets

n	Index of buses
i	Index of PV systems
s	Index of BES systems
t	Index of time steps
k	Index of scenarios
up(n)	Upstream bus of bus n

Sets

$\mathcal{N}$	Set of buses
$\mathcal{T}$	Set of time steps
$\mathcal{K}$	Set of scenarios
$\mathcal{I}$	Set of PV systems
$\mathcal{S}$	Set of BES systems

Variables

v_{ntk}	Voltage square of the bus n
l_{ntk}	Square of the current flowing in the central element of the line n's model or square of current flowing in the transformer n.
p_{ntk}	Real power flow entering the bus n from bottom
q_{ntk}	Reactive power flow entering the bus n from bottom
$v_{ntk}^{(up)}$	Voltage square of bus "up(n)".
$p_{ntk}^{(up)}$	Real power flow entering the bus n from the top.
$q_{ntk}^{(up)}$	Reactive power flow entering the bus n from the top.
$p_{ntk}^{(DM)}$	Real power demand of the bus n.
$q_{ntk}^{(DM)}$	Reactive power demand of the bus n.
$p_{itk}^{(PV)}$	Real power production of the PV system i.
$q_{itk}^{(PV)}$	Reactive power production of the PV system i.
$p_{stk}^{(BS)}$	Real power production of the BES system s.
$q_{stk}^{(BS)}$	Reactive power production of the BES system s.
$p_{ntk}^{(net)}$	Net active power injection to the bus n.
$q_{ntk}^{(net)}$	Net reactive power injection to the bus n.
$p_{ntk}^{(shed)}$	Quantity of load shedding at the bus n.
e_{stk}	State-of-Charge of the BES system s.
$p_t^{(PL)}$	Planned active power at the PCC.
$q_t^{(PL)}$	Planned reactive power at the PCC.
$r_t^{(p\uparrow)}$	Planned upward active power flexibility at the PCC
$r_t^{(p\downarrow)}$	Planned downward active power flexibility at the PCC
$r_t^{(q\uparrow)}$	Planned upward reactive power flexibility at the PCC
$r_t^{(q\downarrow)}$	Planned downward reactive power flexibility at the PCC
$r_{tk}^{(RE,p\uparrow)}$	Proportion of requested upward active power in real-time
$r_{tk}^{(RE,p\downarrow)}$	Proportion of requested downward active power in real-time
$r_{tk}^{(RE,q\uparrow)}$	Proportion of requested upward reactive power in real-time
$r_{tk}^{(RE,q\downarrow)}$	Proportion of requested downward reactive power in real-time

Parameters

$u_{nn'}$	A binary parameter, where $u_{nn'} = 1$ if $n' = \text{up}(n)$ and $u_{nn'} = 0$ otherwise
r_n	Resistance of the line n or the transformer n
x_n	Reactance of the line n or the transformer n
$2 \cdot b_n$	Shunt impedance of the line n or the transformer n
δ	Maximum allowable deviation from the requested active and reactive power at the PCC
$\eta_s^{(C)}$	Charging efficiency of the BES system s
$\eta_s^{(D)}$	Discharging efficiency of the BES system s
Δt	Time-step duration

delivering the flexibilities. However, the constraints of ADN have not been included, and exploiting the underlying flexibilities of PV systems has not been addressed. In [9], the effects of BES system cycle-degradation and the value of lost load (VOLL) on an ADN's flexibility booking strategy have been evaluated under demand uncertainty.

A risk-averse scheduling optimization problem has been developed in [10] to completely leverage the controllability of multiple DERs in an ADN on both spatial and temporal scales. In [11], a solution for planning BES and PV systems to give flexibility based on chance-constrained (CC) optimization has been proposed. The support of DERs in providing operational flexibility has been addressed in [12] by developing a non-cooperative game among strategic aggregators that coordinates the DERs. The active and reactive power flexibility capability area of an ADN has been discovered using linear stochastic optimization in [13], where the DERs are limited to PV systems. Finally, a coordination mechanism has been developed in [14] based on stochastic optimization for controlling the BES and PV systems in ADN to provide flexibility. The proposed two-stage problem in [14] only deals with scheduling flexibilities, and the robustness of the first-stage problem for day-ahead scheduling of flexibilities has not been guaranteed.

To solve the ADN's stochastic optimization problem, a number of scenarios must be generated to account for the uncertainties in electricity demands and PV power production capabilities. The chosen scenarios have a direct impact on the outcome of stochastic optimization and, as a result, the robustness, as well as the optimality of the derived solution. However, in prior studies, either the scenarios were given or the significance of generating realistic scenarios was overlooked. By realistic scenarios, we mean that they must be brief but comprehensive, in the sense that the number of scenarios must be *limited* while still capturing

the *typical* operation (i.e., time-series pattern) as well as *particular* conditions (i.e., small forecast error in net).

Li et al. in [15] have carried out a survey of various classes of scenario generation methods for renewable energy systems. The classic approach, which has been used in [13] and [14], is based on auto-regressive integrated moving average (ARIMA) forecasting and Monte Carlo sampling. As shown in [16,17], the solution's robustness would be justified if a large enough number of scenarios were generated. However, this technique is not scalable for ADNs with a large number of buses and time steps. Based on historical data and K-Means clustering, [18] presents a method for generating a small number of representative scenarios.

1.1. Solution approach

This work provides a Markov chain-based technique for generating representative scenarios to solve the proposed stochastic optimization problem of an ADN while taking advantage of underlying flexibilities. The huge amounts of measured data in ADNs, which resulted from the power industry's digitalization via the use of digital technologies, may assist the ADNs' operators in overcoming the obstacles to exploiting the underlying flexibilities. In particular, the ADNs' operators have improved visibility into the ADN's component constraints and power supply quality and security. The forecast of PV power production and electricity demands also becomes more precise. As a result, the ADNs' operators can handle the planned flexibilities in an autonomous manner.

In this paper, we present a scalable and data-driven optimization problem for scheduling the underlying active and reactive power flexibilities of ADNs to address the shortcomings of previous studies. We framed a stochastic optimization problem for ADNs' operators. Because of the operational uncertainties, the proposed stochastic optimization algorithm considers all constraints of the ADN's components (i.e., lines, transformers, and DERs), and security, as well as quality limitations. Variations in PV system production capability, electricity demands, the request of TSO for deploying the planned flexibilities, and the voltage magnitude at PCC are examples of operational uncertainties. We anticipate a data-driven scenario based on measured data selection strategy for proposed stochastic optimization, resulting in lower computational costs.

1.2. Contributions

Given the foregoing setting, the paper's major contributions in terms of modeling and solution methodology are as follows:

- We use a formulation that works in both directions of power flow in radial networks to model the constraints of ADN's line/transformer flows, as well as the quality and security of power supply. We also model the constraints of BES systems using a relaxed convex formulation that does not require binary variables to determine whether BES systems are charging or discharging. Finally, the PV system capability constraints are incorporated into our formulation. Compared to previous studies [10–14], our paper uses a more exact model that includes the power flow constraints of the ADN's network, and capability curves of PV as well as BES systems.
- We present a scalable second-order cone programming (SOCP) solution for our stochastic optimization problem, which can be solved using commercial or freely available convex optimization solvers. The developed optimization problems cited in [13,14,16,17] are not scalable because they cannot be applied to a network with many buses considering large number of operational scenarios, whereas our developed optimization problem is scalable.
- We introduce a scenario selection strategy based on measured data. Because the proposed strategy takes into account the temporal correlation in forecasting errors and results in a small number of scenarios, the proposed optimization problem can be used for large-scale problems. In this regard, we use a Markov chain-based technique to generate representative PV power and electricity demands scenarios that can be used for stochastic co-optimization of energy and flexibility, as opposed to [19], which developed a method for generating a large number of scenarios for only PV power production. Furthermore, in contrast to [18], which generates the optimal scenarios for optimizing the operation of an ADN based on K-Means, we develop a new strategy that takes the Markov chain character of uncertain parameters into account and is trained on historical data.

1.3. Structure of the paper

The remainder of the paper is organized as follows: The constraints' model is presented in Section 2. The stochastic optimization problem is formulated in Section 3. The proposed scenario selection strategy is introduced in Section 4. A numerical case study is given in Section 5. Finally, the paper is concluded in Section 6.

2. Mathematical formulation

The sections that follow discuss various constraints on the targeted optimization problem. The general notation and uncertain parameters are described in Section 2.1. The power flow equations with the network's security constraints are presented in Section 2.2. Finally, the models for PV and BES systems are provided in Sections 2.3 and 2.4, respectively.

2.1. Preliminary and uncertain parameters

Time and scenario are represented by the indices $t \in \mathcal{T}$ and $k \in \mathcal{K}$, where $\mathcal{T} = \{1, \ldots, T\}$ and $\mathcal{K} = \{1, \ldots, K\}$. Other than the PCC, buses are denoted by $n \in \mathcal{N}$, where $\mathcal{N} = \{1, \ldots, N\}$. The PCC is represented by an index of 0. The PV and BES systems are indexed by $i \in \mathcal{I}$ and $s \in \mathcal{S}$, respectively. Finally, $\mathcal{S}_n = \{1, \ldots, S_n\}$ and $\mathcal{I}_n = \{1, \ldots, I_n\}$ denote the sets of PV and BES systems connected to bus n.

The following four uncertain parameters are modeled with scenarios $k \in \mathcal{K}$ to exploit the underlying flexibilities in ADNs: (i) the active and reactive power demand of users, i.e., $p_{ntk}^{(DM)}$ and $q_{ntk}^{(DM)}$, which are negative; (ii) the available active power $p_{itk}^{(PV,max)}$ of PV systems, which is positive; (iii) the voltage square magnitude v_{0tk} at the PCC; and (iv) the proportion of exploited active and reactive power flexibilities with respect to the planned values, i.e., $r_{tk}^{(RE,p\uparrow)}$, $r_{tk}^{(RE,p\downarrow)}$, $r_{tk}^{(RE,q\uparrow)}$, and $r_{tk}^{(RE,q\downarrow)}$.

Power flow variables in ADNs, as well as state and auxiliary variables in PV and BES systems, are written with an index of $k \in \mathcal{K}$ because they are dependent on uncertain parameters. Except for the proportion of exploited active and reactive power, for which we assume a random *Uniform* probability density function (PDF), we have a record of measurements for uncertain parameters. We calculated a number of scenarios based on such measurements and PDFs. The method for generating random scenarios from given measurements and PDFs is discussed further in the paper.

2.2. Power flow, security, and quality constraints

We assume a radial ADN. The label "up(n)" refers to a bus upstream of the bus n. The line or transformer ending in bus

n is also indexed with n. The binary parameter $u_{nn'}$ is set to 1 if $n' = \text{up}(n)$, otherwise $u_{nn'} = 0$. For clarity and extra notations, the Π-model of a distribution line or a transformer is shown in Fig. 1. Note that because all variables are expressed in per-units based on the voltage level of the transformers' sides, the transformer voltage ratio has been eliminated.

We denote the voltage square of bus "up(n)" with $v_{\text{ntk}}^{(\text{up})}$. The variables $p_{\text{ntk}}^{(\text{up})}$ and $q_{\text{ntk}}^{(\text{up})}$ represent the active and reactive power flows entering the line n or the transformer n, i.e., from the bus "up(n)". The variables p_{ntk} and q_{ntk} represent the active and reactive power flow entering bus n from the bottom of the line n (or the transformer n). The variable l_{ntk} represents the square of the current flowing in the central element of the line n (or the transformer n). We consider the line n's (or the transformer n's) resistance, reactance, and shunt impedance to be r_{n}, x_{n}, and $2 \cdot b_{\text{n}}$, respectively. The variable $v_{0\text{tk}}$ represents the PCC's voltage square magnitude at time t and is treated as an uncertain parameter because of the following fact: The magnitude of the PCC's voltage depends on the transmission network's operation conditions, which include, among other factors, the voltage set points at the power plants. Such data is not available to the distribution level operators.

To balance the injection and withdrawal of power from various buses, the power flow equations are given. The net active and reactive power injections to bus n at time t are represented by $p_{\text{ntk}}^{(\text{net})}$ and $q_{\text{ntk}}^{(\text{net})}$, respectively.

$$p_{\text{ntk}}^{(\text{net})} = \left(\sum\nolimits_{i\in\mathcal{I}_{\text{n}}} p_{\text{itk}}^{(\text{PV})}\right) + \left(\sum\nolimits_{s\in\mathcal{S}_{\text{n}}} p_{\text{stk}}^{(\text{BS})}\right) + p_{\text{ntk}}^{(\text{DM})} + p_{\text{ntk}}^{(\text{shed})}, \forall n, t, k, \tag{1}$$

$$q_{\text{ntk}}^{(\text{net})} = \left(\sum\nolimits_{i\in\mathcal{I}_{\text{n}}} q_{\text{itk}}^{(\text{PV})}\right) + \left(\sum\nolimits_{s\in\mathcal{S}_{\text{n}}} q_{\text{stk}}^{(\text{BS})}\right) + q_{\text{ntk}}^{(\text{DM})}, \forall n, t, k, \tag{2}$$

where $p_{\text{itk}}^{(\text{PV})}$, $q_{\text{itk}}^{(\text{PV})}$, $p_{\text{stk}}^{(\text{BS})}$, and $q_{\text{stk}}^{(\text{BS})}$ denote the active power production of the PV system i, reactive power of the PV system i, active power output of the BES system s, and the reactive power output of the BES system s. The parameters $p_{\text{ntk}}^{(\text{DM})}$ and $q_{\text{ntk}}^{(\text{DM})}$ represent the active and reactive power demand of the bus n, respectively. Finally, $p_{\text{ntk}}^{(\text{shed})}$ is the quantity of load shedding at bus n.

Referring to Fig. 1, the followings are power flow equations and security constraints for a given radial ADN:

$$p_{\text{ntk}}^{(\text{up})} = r_n \cdot l_{\text{ntk}} - p_{\text{ntk}}^{(\text{net})} + \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot p_{n't}^{(\text{up})}, \forall n, t, k, \tag{3}$$

$$q_{\text{ntk}}^{(\text{up})} = x_n \cdot l_{\text{ntk}} - q_{\text{ntk}}^{(\text{net})} + \sum_{n'\in\mathcal{N}} u_{nn'} \cdot q_{n't}^{(\text{up})} - \left(v_{\text{ntk}} + v_{\text{ntk}}^{(\text{up})}\right) \cdot b_n, \forall n, t, k, \tag{4}$$

$$v_{\text{ntk}}^{(\text{up})} = v_{\text{ntk}} + 2 \cdot \left(r_n \cdot p_{\text{ntk}}^{(\text{up})} + x_n \cdot (q_{\text{ntk}}^{(\text{up})} + v_{\text{ntk}}^{(\text{up})} \cdot b_n)\right) - (r_n^2 + x_n^2) \cdot l_{\text{ntk}}, \forall n, t, k, \tag{5}$$

$$l_{\text{ntk}} \cdot v_{\text{ntk}}^{(\text{up})} \geq \left(p_{\text{ntk}}^{(\text{up})}\right)^2 + \left(q_{\text{ntk}}^{(\text{up})}\right)^2, \forall n, t, k, \tag{6}$$

$$\hat{p}_{\text{ntk}}^{(\text{up})} = \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \hat{p}_{n't}^{(\text{up})} - p_{\text{ntk}}^{(\text{net})}, \forall n, t, k, \tag{7}$$

$$\hat{q}_{\text{ntk}}^{(\text{up})} = \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \hat{q}_{n't}^{(\text{up})} - \left(\overline{v}_{\text{ntk}}^{(\text{up})} + \overline{v}_{\text{ntk}}\right) \cdot b_n - q_{\text{ntk}}^{(\text{net})}, \forall n, t, k, \tag{8}$$

$$\overline{v}_{\text{ntk}}^{(\text{up})} = \overline{v}_{\text{ntk}} + 2 \cdot \left(r_n \cdot \hat{p}_{\text{ntk}}^{(\text{up})} + x_n \cdot (\hat{q}_{\text{ntk}}^{(\text{up})} + \overline{v}_{\text{ntk}}^{(\text{up})} \cdot b_n)\right), \forall n, t, k, \tag{9}$$

$$\overline{p}_{\text{ntk}}^{(\text{up})} = r_n \cdot \bar{l}_{\text{ntk}} - p_{\text{ntk}}^{(\text{net})} + \sum_{n'\in\mathcal{N}} u_{nn'} \cdot \overline{p}_{n't}^{(\text{up})}, \forall n, t, k, \tag{10}$$

$$\overline{q}_{\text{ntk}}^{(\text{up})} = x_n \cdot \bar{l}_{\text{ntk}} - q_{\text{ntk}}^{(\text{net})} + \sum_{n'\in\mathcal{N}} u_{nn'} \cdot \overline{q}_{n't}^{(\text{up})} - \left(v_{\text{ntk}} + v_{\text{ntk}}^{(\text{up})}\right) \cdot b_n, \forall n, t, k, \tag{11}$$

$$\bar{l}_{\text{ntk}} \cdot v_{\text{ntk}} \geq \max\{\hat{p}_{\text{ntk}}^2, \overline{p}_{\text{ntk}}^2\} + \max\{(\hat{q}_{\text{ntk}} - \overline{v}_{\text{ntk}} \cdot b_n)^2, (\overline{q}_{\text{ntk}} - v_{\text{ntk}} \cdot b_n)^2\}, \forall n, t, k, \tag{12}$$

$$\bar{l}_{\text{ntk}} \cdot v_{\text{ntk}}^{(\text{up})} \geq \max\{(\hat{p}_{\text{ntk}}^{(\text{up})})^2, (\overline{p}_{\text{ntk}}^{(\text{up})})^2\} + \max\{(\hat{q}_{\text{ntk}}^{(\text{up})} + \overline{v}_{\text{ntk}}^{(\text{up})} \cdot b_n)^2, (\overline{q}_{\text{ntk}}^{(\text{up})} + v_{\text{ntk}}^{(\text{up})} \cdot b_n)^2\}, \forall n, t, k, \tag{13}$$

$$\overline{p}_{\text{ntk}} = -p_{\text{ntk}}^{(\text{net})} + \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \overline{p}_{n't}^{(\text{up})}, \forall n, t, k, \tag{14}$$

$$\overline{q}_{\text{ntk}} = -q_{\text{ntk}}^{(\text{net})} + \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \overline{q}_{n't}^{(\text{up})}, \forall n, t, k, \tag{15}$$

$$\hat{p}_{\text{ntk}} = -p_{\text{ntk}}^{(\text{net})} + \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \hat{p}_{n't}^{(\text{up})}, \forall n, t, k, \tag{16}$$

$$\hat{q}_{\text{ntk}} = -q_{\text{ntk}}^{(\text{net})} + \sum\nolimits_{n'\in\mathcal{N}} u_{nn'} \cdot \hat{q}_{n't}^{(\text{up})}, \forall n, t, k, \tag{17}$$

$$l_n^{(\max)} \cdot v_{\text{ntk}} \geq \max\{\hat{p}_{\text{ntk}}, \overline{p}_{\text{ntk}}\}^2 + \max\{\hat{q}_{\text{ntk}}, \overline{q}_{\text{ntk}}\}^2, \forall n, t, k, \tag{18}$$

$$l_n^{(\max)} \cdot v_{\text{ntk}}^{(\text{up})} \geq \max\{\hat{p}_{\text{ntk}}^{(\text{up})}, \overline{p}_{\text{ntk}}^{(\text{up})}\}^2 + \max\{\hat{q}_{\text{ntk}}^{(\text{up})}, \overline{q}_{\text{ntk}}^{(\text{up})}\}^2, \forall n, t, k, \tag{19}$$

$$v_n^{(\min)} \leq v_{\text{ntk}} \leq v_n^{(\max)}, \forall n, t, k, \tag{20}$$

where $l_n^{(\max)}$ is the maximum allowable current square for line n or transformer n, $v_n^{(\min)}$ and $v_n^{(\max)}$ are the minimum and maximum permissible values of voltage square for bus n, respectively. The auxiliary variables $\hat{p}_{\text{ntk}}^{(\text{up})}$, $\hat{q}_{\text{ntk}}^{(\text{up})}$, $\overline{p}_{\text{ntk}}^{(\text{up})}$, $\overline{q}_{\text{ntk}}^{(\text{up})}$, $\hat{p}_{\text{ntk}}$, $\hat{q}_{\text{ntk}}$, $\overline{p}_{\text{ntk}}$, $\overline{q}_{\text{ntk}}$, $\overline{v}_{\text{ntk}}$, $\overline{v}_{\text{ntk}}^{(\text{up})}$, and $\bar{l}_{\text{ntk}}$ are added to ensure that the constraints are accurate in both directions of power flow. Including these auxiliary variables and the resulting constraints is based on the model presented in [20], which relaxes the power flow equations to make the feasible space convex.

Here, (3)–(5) are directly obtained by applying Kirchhoff's law to Fig. 1. (6) is the relaxed version of the relationship between the square of current flowing in the central part of the two-port Π model and the voltage square and the power coming from upward of line n (or transformer n). In the case of reverse power flow, the relaxation of constraint (6) may be inexact. To address this issue, the constraints (7)–(17) are supplemented with auxiliary variables that are upper bounds of power flow variables ($\overline{p}_{\text{ntk}}$, $\overline{q}_{\text{ntk}}$, $\overline{v}_{\text{ntk}}$, and $\bar{l}_{\text{ntk}}$) and variables that do not depend on the flowing current of the lines ($\hat{p}_{\text{ntk}}$ and $\hat{q}_{\text{ntk}}$). See [20] for a comprehensive discussion of these constraints.

2.3. Photovoltaic systems capability constraints

The capability curve of a PV system $i \in \mathcal{I}$ is defined by the intersection of the following constraints: (i) the voltage constraint of the PV system's power electronic converter; (ii) the current constraint of the PV system's converter; and (iii) the maximum power production due to solar irradiation and cell temperature of the PV system.

(i) As demonstrated in [21], the converter's voltage constraint of a PV system is modeled as follows:

$$\left(p_{\text{itk}}^{(\text{PV})}\right)^2 + \left(q_{\text{itk}}^{(\text{PV})} + \frac{3 \cdot (v_{\text{i}}^{(\text{net,PV})})^2}{x_{\text{i}}^{(\text{PV})}}\right)^2 \leq \left(\frac{3 \cdot v_{\text{i}}^{(\text{net,PV})} \cdot v_{\text{i}}^{(\text{con,PV})}}{x_{\text{i}}^{(\text{PV})}}\right)^2, \forall i, t, k, \tag{21}$$

where $v_{\text{i}}^{(\text{net,PV})}$, $v_{\text{i}}^{(\text{con,PV})}$, and $x_{\text{i}}^{(\text{PV})}$ are the ADN's voltage, the converter's voltage, and the Thévenin's reactance from the PV system's perspective, respectively.

(ii) The converter's current constraint of PV system i at time-step $t \in \mathcal{T}$ is

$$\left(p_{\text{itk}}^{(\text{PV})}\right)^2 + \left(q_{\text{itk}}^{(\text{PV})}\right)^2 \leq \left(S_{\text{i}}^{(\text{PV,max})}\right)^2, \forall i, t, k. \tag{22}$$

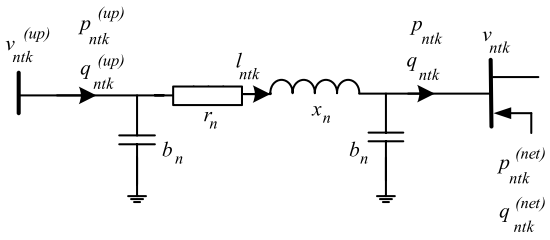


Fig. 1. Π-model of a distribution line or a transformer.

(iii) The PV system i's active power constraint at time-step $t \in \mathcal{T}$ is

$$0 \leq p_{\text{itk}}^{(\text{PV})} \leq p_{\text{itk}}^{(\text{PV,max})}, \forall i, t, k. \tag{23}$$

It should be noted that $p_{\text{itk}}^{(\text{PV,max})}$ is an uncertain parameter that depends on available solar irradiation and PV system i's cell temperature.

2.4. Battery energy storage systems constraints

Each BES $s \in \mathcal{S}$ is linked to the ADN by a power electronic converter. The converter's capability constraint must therefore be handled. The BES's state-of-charge (SoC), denoted by e_{stk}, must also be kept within the acceptable minimum and maximum values.

The model under consideration is based on the convex reformulation in [22], which does not use binary variables to force a BES to operate in just charging or discharging mode at each time-step. To generate a set of relaxed convex constraints, additional continuous variables and convex constraints are added.

The detailed equality and inequality constraints for each BES $s \in \mathcal{S}$ are

$$e_{s(t+1)k} = \frac{100}{e_{\text{s}}^{(\text{max})}} \cdot \left(\frac{p_{\text{st}}^{(\text{BS,D})} \cdot \Delta t}{\eta_{\text{s}}^{(\text{D})}} + \eta_{\text{s}}^{(\text{C})} \cdot p_{\text{st}}^{(\text{BS,C})} \cdot \Delta t \right) + e_{\text{stk}}, \forall s, t, k, \tag{24}$$

$$0 \leq p_{\text{stk}}^{(\text{BS,C})}, \forall s, t, k, \tag{25}$$

$$0 \geq p_{\text{stk}}^{(\text{BS,D})}, \forall s, t, k, \tag{26}$$

$$p_{\text{stk}}^{(\text{BS})} = p_{\text{stk}}^{(\text{BS,C})} + p_{\text{stk}}^{(\text{BS,D})}, \forall s, t, k, \tag{27}$$

$$(p_{\text{stk}}^{(\text{BS})})^2 + (q_{\text{stk}}^{(\text{BS})})^2 \leq (S_{\text{s}}^{(\text{BS,max})})^2, \forall s, t, k, \tag{28}$$

$$100 \cdot \frac{e_{\text{s}}^{(\text{min})}}{e_{\text{s}}^{(\text{max})}} \leq e_{\text{stk}} \leq 100, \forall s, t, k, \tag{29}$$

$$\tilde{e}_{s(t+1)k} = \tilde{e}_{\text{stk}} + \frac{100}{e_{\text{s}}^{(\text{max})}} \cdot \left(p_{\text{stk}}^{(\text{BS,D})} \cdot \Delta t + p_{\text{stk}}^{(\text{BS,C})} \cdot \Delta t \right), \forall s, t, k, \tag{30}$$

$$100 \cdot \frac{e_{\text{s}}^{(\text{min})}}{e_{\text{s}}^{(\text{max})}} \leq \tilde{e}_{\text{stk}} \leq 100, \forall s, t, k, \tag{31}$$

where $e_{\text{s}}^{(\text{min})}$ and $e_{\text{s}}^{(\text{max})}$ are the minimum and maximum amounts of energy that can be stored in the BES s, respectively; and $S_{\text{s}}^{(\text{BS,max})}$ is the maximum amount of apparent power that BES s can produce or consume.

Here, (24) models the classical energy balance for a lossy BES. The constraints (25) and (26) satisfy that the charge and discharge powers of the BES $s \in \mathcal{S}$ are positive variables at time t and scenario k. The net power of BES $s \in \mathcal{S}$ is determined in (27). The current limit of the power converter is given in (28). The maximum and minimum SoC limitations of BES $s \in \mathcal{S}$ are modeled by (29). The constraints (30) and (31) are also included, along with the auxiliary variable $\tilde{e}_{\text{stk}}$, which represents the SoC of the ideal BES $s \in \mathcal{S}$ without any loss, to prevent misuse of the relaxed model without binary variables. For a detailed explanation of the mentioned constraints, see [22].

3. Optimal dispatch plan based on stochastic optimization

First, we define the objective function of the ADN's operator in order to obtain the optimal dispatch plan for PV and BES systems by utilizing the underlying flexibilities. The PCC's planned active and reactive power are denoted as $p_{\text{t}}^{(\text{PL})}$ and $q_{\text{t}}^{(\text{PL})}$, respectively. The planned upward active and reactive power flexibilities are also denoted by $r_{\text{t}}^{(\text{p}\uparrow)}$ and $r_{\text{t}}^{(\text{q}\uparrow)}$, respectively. Furthermore, the planned downward active and reactive power flexibilities are denoted by $r_{\text{t}}^{(\text{p}\downarrow)}$ and $r_{\text{t}}^{(\text{q}\downarrow)}$, respectively.

The objective of ADN's operator is to maximize its profit from selling active and reactive power as well as corresponding flexibilities, as formulated in (32), while keeping the distribution network's and resources' security constraints in mind, as presented in (33). The optimization problem of ADN's operator based on stochastic programming is as follows, taking into account the given constraints in the previous section:

$$\begin{aligned} \text{objective} = \sum\nolimits_{t \in \mathcal{T}} \Big(& \lambda_{\text{t}}^{(\text{p,PL})} \cdot p_{\text{t}}^{(\text{PL})} + \lambda_{\text{t}}^{(\text{q,PL})} \cdot q_{\text{t}}^{(\text{PL})} \\ & + \lambda_{\text{t}}^{(\text{p}\uparrow)} \cdot r_{\text{t}}^{(\text{p}\uparrow)} + \lambda_{\text{t}}^{(\text{p}\downarrow)} \cdot r_{\text{t}}^{(\text{p}\downarrow)} \\ & + \lambda_{\text{t}}^{(\text{q}\uparrow)} \cdot r_{\text{t}}^{(\text{q}\uparrow)} + \lambda_{\text{t}}^{(\text{q}\downarrow)} \cdot r_{\text{t}}^{(\text{q}\downarrow)} - \text{VOLL} \cdot p_{\text{t}}^{(\text{shed})} \Big), \end{aligned} \tag{32}$$

$$\text{subject to:} \quad (1)\text{–}(31), \tag{33}$$

$$p_{\text{t}}^{(\text{shed})} := \frac{1}{K} \cdot \sum\nolimits_{k \in \mathcal{K}} \sum\nolimits_{n \in \mathcal{N}} p_{\text{ntk}}^{(\text{shed})}, \forall t, \tag{34}$$

$$\left| p_{\text{t}}^{(\text{PL})} + r_{\text{tk}}^{(\text{RE,p}\uparrow)} \cdot r_{\text{t}}^{(\text{p}\uparrow)} - r_{\text{tk}}^{(\text{RE,p}\downarrow)} \cdot r_{\text{t}}^{(\text{p}\downarrow)} - p_{0\text{tk}} \right| \leq \delta, \forall t, k, \tag{35}$$

$$\left| q_{\text{t}}^{(\text{PL})} + r_{\text{tk}}^{(\text{RE,q}\uparrow)} \cdot r_{\text{t}}^{(\text{q}\uparrow)} - r_{\text{tk}}^{(\text{RE,q}\downarrow)} \cdot r_{\text{t}}^{(\text{q}\downarrow)} - q_{0\text{tk}} \right| \leq \delta, \forall t, k, \tag{36}$$

where $\lambda_{\text{t}}^{(\text{p,PL})}$, $\lambda_{\text{t}}^{(\text{q,PL})}$, $\lambda_{\text{t}}^{(\text{p}\uparrow)}$, $\lambda_{\text{t}}^{(\text{p}\downarrow)}$, $\lambda_{\text{t}}^{(\text{q}\uparrow)}$, and $\lambda_{\text{t}}^{(\text{q}\downarrow)}$ are prices of planned active power, reactive power, upward active power flexibility, downward active power flexibility, upward reactive power flexibility, downward reactive power flexibility, respectively.[1] Furthermore, VOLL is the cost of load shedding and $p_{\text{t}}^{(\text{shed})}$ is the total expected amount of load shedding at time t as calculated in (34).

The TSO requests in real-time a proportion of the planned flexibilities, namely $r_{\text{tk}}^{(\text{RE,p}\uparrow)}$, $r_{\text{tk}}^{(\text{RE,p}\downarrow)}$, $r_{\text{tk}}^{(\text{RE,q}\uparrow)}$, and $r_{\text{tk}}^{(\text{RE,q}\downarrow)}$. The requested flexibilities are uncertain and depend on the behavior of transmission network's operation. To tighten the feasible space and balance the ADN for exploiting the underlying flexibilities in real-time, the constraints (35)–(36) are embedded in the problem, where $p_{0\text{tk}}$ and $q_{0\text{tk}}$ are the active and reactive power available at the PCC for scenario k. The parameter δ is a small positive number, e.g., $\delta = 0.01$ kW, that represents the maximum allowable deviation from the requested flexibilities in addition to the PCC dispatch plan.

The optimization problem (32)–(36) is a SOCP. The method for generating scenarios $\mathcal{K} = \{1, 2, \ldots, K\}$ based on the measured data is described in the following section.

4. Solution methodology

4.1. Benchmark scenario selection strategies

In order to evaluate the advantages of our proposed scenario selection strategy (which is described in the next subsection),

[1] This market assumption that there are real-time prices for reactive power, as well as upward and downward flexibility may not be valid in some cases or countries (see [23] for real-time reactive power pricing, as well as [24] for upward and downward flexibility pricing). Nonetheless, the given model and objective for scheduling reactive power and flexibility are general and may be adapted to various TSO-DSO coordination schemes.

we investigate a case where we use the data from previous days without any clustering as scenarios and the following two benchmark scenario selection procedures.

The first benchmark is based on data from a month ago, which is clustered into a limited number. The data from the preceding month is clustered using the K-Means approach, and the centers of the clusters are considered to be the scenarios [18]. A case without clustering and using the data from the most recent few days as scenarios directly are also taken into consideration in order to assess the benefit of clustering. The disadvantage of this approach is that the quality of the stochastic optimization-derived solution is strongly influenced by the scenarios that occurred in the preceding month. To evaluate the benefit

The second benchmark strategy is based on the ARIMA forecasting method, which is a classical approach to predicting the uncertain parameters [25]. We forecast each uncertain parameter for the following day. The standard deviation of forecasting errors can then be calculated. Using the forecasted value and the standard deviation of forecast errors, we generate a large number of scenarios, for example, 1000. Finally, we employ the K-Means method to reduce the number of generated scenarios, and the centers of the clusters are used to represent the reduced output scenarios. The disadvantage of this strategy is that the temporal correlation is not reflected in the output generated scenarios.

4.2. Proposed data-driven scenario selection strategy

A randomized sampling approach based on historical data or knowledge of the PDFs of the uncertain parameters is required for the proposed stochastic optimization problem. In order to accomplish this, we created a data-driven scenario selection strategy based on the model introduced in [19]. It is important to note that the solution to the stochastic optimization problem is robust if the number of scenarios, i.e., K, is large enough. Based on the results of [16,17], the minimum number of scenarios for establishing different confidence levels can be determined. However, the number of scenarios in the typical cases of our problem is so large that the resulting problem is not scalable.[2] Thus, we assume a fixed and small number of representative scenarios[3] based on historical data.

The following is the steps of the proposed scenario selection strategy, which is summarized in Fig. 2:

(i) *Pre-process data*: We aggregate the previous month's data on solar irradiance and electricity demand. Then, the data is normalized by removing outliers and standardizing it based on minimum and maximum values. The data is clustered into specific numbers using the K-Means approach, such as three clusters signifying cloudy, intermittently cloudy, or clear days. This feature, known as the "cluster number", will be added to the data.

(ii) *Train the model*: We train a model based on Markov chain for each cluster number. As a result, we get a specific number of, e.g., three, transition probability matrices (one for each cluster number).

(iii) *Forecast*: We use the *Random Forest* defined in [26] to forecast the uncertain parameters for the next day. The resulting forecast is referred to as the baseline scenario. The *Random Forest* is an ensemble learning method with outstanding computational performance that may be used for classification as well as regression. The *Random Forest* is a collection of many decision trees known as "forests". Each tree is based on an independent random sample, and the average of all the trees' outputs is regarded as the solution to the regression problem. We normalize both the input and output patterns before training the random forest model for each node. Outliers are also detected and removed from the time series to minimize over-fitting. To fit and validate these models, we use the python function `RandomForestRegressor` from the package `sklearn`.

(iv) *Generate scenarios*: Using the distance between the forecast and cluster centers, we identify which cluster number corresponds to the baseline scenario. We then generate a huge number of scenarios for that cluster number, using the corresponding transition probability matrix.

(v) *Reduce scenarios*: We apply the K-Means method to reduce the number of generated scenarios. Cluster centers are regarded as representative scenarios in addition to the previously forecasted baseline scenario.

[2] The required number of scenarios for our case study is more than 10,000, where the ADN has just five buses and the problem is formulated for 144 time steps (24 × 6 of 10 min time steps).

[3] In the sense that they exhibit the same behavior as a large number of scenarios, these scenarios are efficient.

Table 1
Parameters of the lines and transformer of the ADN.

Parameter	Value	Parameter	Value
r_{t1} in Ohm	0.0110	r_{l2} in Ohm	0.110
x_{t1} in Ohm	0.0144	x_{l2} in Ohm	0.252
$2 \cdot b_{t1}$ in Siemens	0.0000	$2 \cdot b_{l2}$ in Siemens	0.229
r_{l3} in Ohm	0.066	r_{l4} in Ohm	0.098
x_{l3} in Ohm	0.148	x_{l4} in Ohm	0.197
$2 \cdot b_{l3}$ in Siemens	0.0312	$2 \cdot b_{l4}$ in Siemens	0.241

5. Numerical case study

The test is carried out in a reconfigurable distribution network laboratory [27], mimicking a real low-voltage ADN in a rural area of Switzerland. The considered ADN is depicted in abstract form in Fig. 3, which includes an 8.5 kW PV system, a 69 kWh BES system, a transformer, and three distribution lines. Table 1 displays the parameters of case study.

The proposed stochastic optimization problem and data-driven scenario selection strategy are utilized to plan and exploit the underlying flexibilities for the last day of each month in 2021. The latest 30 days of data on active and reactive power electricity demand, as well as PV power production are used for each day.

5.1. Scenario generation by Markov chain model

The trained Markov chain tables for a sunny day for three variables of (a) active and (b) reactive power of electricity demands and (c) PV power production are shown in Fig. 4. We assumed that each variable has 21 states, so we discretized them linearly between their minimum and maximum. Each color code from blue to yellow represents a transition probability value between 0.0 to 1.0, respectively. We use the `quantecon` package in Python for producing the transition probability matrices form the input measured data. The sum of the probabilities in each column and row of a transition probability matrix must add up to unity based on the Markov requirement for transition probabilities. As shown in Fig. 4-c, the state of PV power production more than states 16 is persistent since the clear sky index is close to one in sunny days. Furthermore, Fig. 4-b shows that the reactive power is close to white noise without any pattern between its minimum and maximum values. Finally, Fig. 4-a shows that the active power of electricity demands is persistent at large values.

The scenarios generated by these trained Markov chain tables are also depicted in Fig. 4. In step (iv) of the proposed scenario selection strategy, we generate 100 scenarios. In the presented

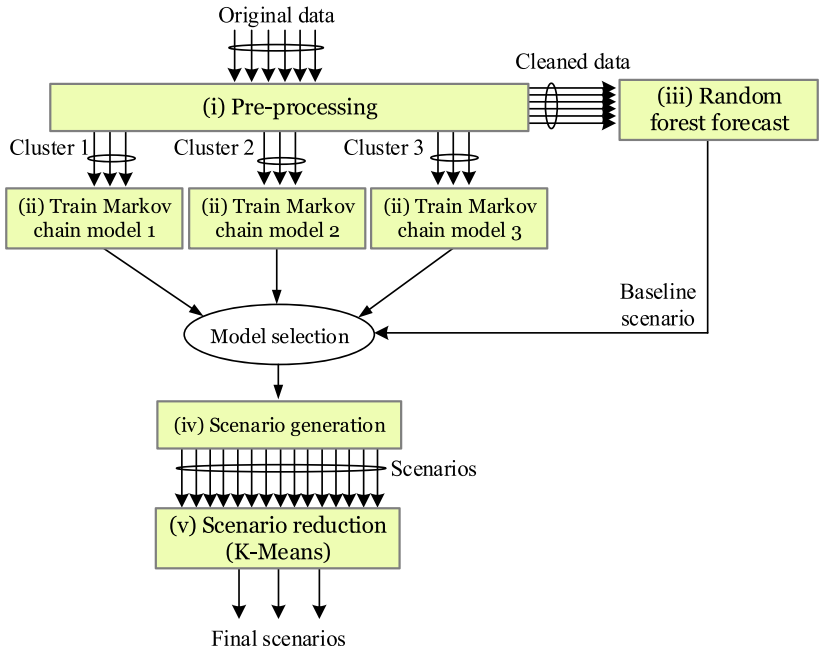


Fig. 2. The flowchart of the proposed scenario selection strategy.

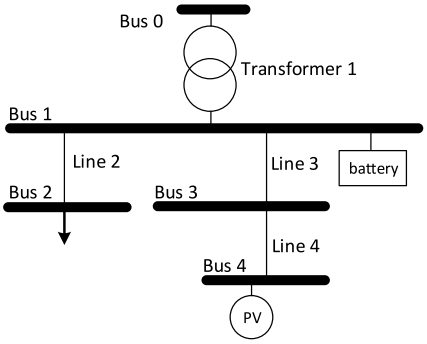


Fig. 3. Single-line diagram of the ADN for the case study.

result, the real data (red line) is between the minimum and maximum of the generated scenarios more than 85% of the time. More information about the usage of the Markov chain for generating a large number of scenarios may be found in [19].

5.2. Scenario selection results

We implement the proposed scenario selection strategy, no clustering approach, and the mentioned two benchmarks stated above in Python using the packages `sklearn`, `scipy`, and `quantecon`. To show that our proposed scenario selection is realistic, we used multiple metrics to demonstrate that generated scenarios are concise and comprehensive. This means that generated scenarios are *limited*, but still captures *typical* operation and *particular* conditions.

5.2.1. Inertia

First, we employ the Elbow technique (see [28]) to get the optimal (*limited*) number of scenarios, i.e., K, for the dataset, which means determining the point at which the declining trend in inertia is no longer worth the additional cost, i.e., a larger number of scenarios and time consumption. Fig. 5 depicts the dataset's inertia, where $K = 3$ is the graph's elbow. A dataset's inertia is measured by how well it was clustered using K-Means. It is calculated by squaring the distance between each data point and its centroid, and summing the squares across one cluster. A good model has low inertia and a small number of clusters (i.e., K). However, there is a trade-off since as K increases, inertia decreases.

5.2.2. Silhouette score

Fig. 6 depicts the outcomes of the first and second benchmarks, as well as the proposed scenario selection strategy with K=3 for active power of electricity demands on July 30th, 2021. Similar plots can be found in Figs. 7 and 8 for the reactive power of electricity demands and PV power production, respectively. Furthermore, the first subplots of Figs. 6–8 show the data of electricity demands and PV power production of the preceding 30 days for July 30th, 2021.

The silhouette score measures how closely the generated scenarios align with the centers of the clusters formed by the actual data from the previous 30 days. It indicates the degree to which the scenarios match the patterns and characteristics of the *typical* data. By comparing the silhouette scores in Figs. 6–8, we conclude that the first benchmark's outcome is more closely correlated to the actual data from the previous 30 days. The Silhouette score ranges from -1 for incorrect clustering to $+1$ for highly dense clustering, which is calculated as follows:

$$\text{Silhouette score} = \frac{b - a}{\max(a, b)}, \tag{37}$$

where a is the average distance between each point within a cluster and b is the average distance between all clusters.

The results of the second benchmark contain three nearly identical scenarios in which the temporal correlation is ignored, resulting in low Silhouette scores (0.1482, 0.0062, and 0.0922), indicating that the resulting scenarios are unlikely to be encountered in practice (by comparing to the data from the previous 30 days). The proposed scenario selection strategy in Figs. 6–8, on the other hand, gets poor scores for the active and reactive power of electricity demands and PV power production. We show in the following that, whereas the outputs of the proposed strategy are different (by comparing them to the data from the previous 30 days), they result in scenarios that are more *particular* and closer to what we expect the next day.

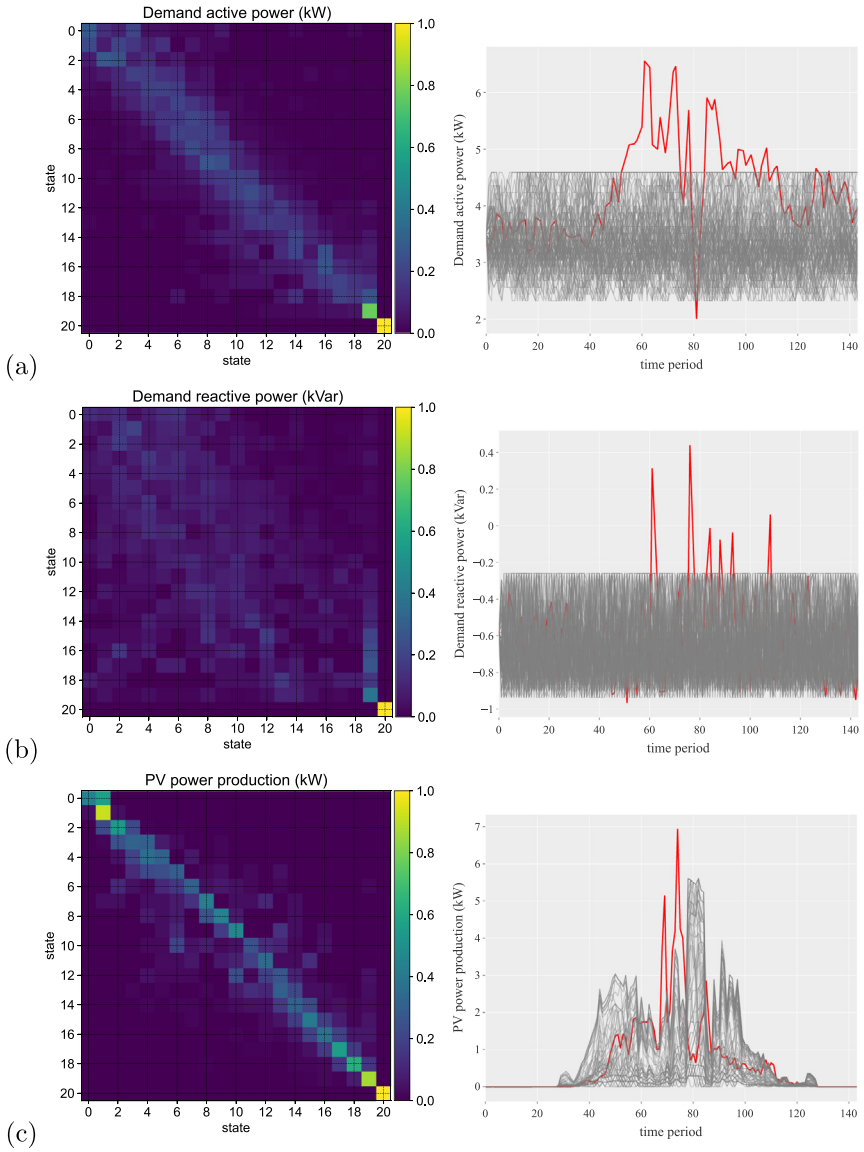


Fig. 4. The trained Markov chain and generated scenarios for June 30th, 2021 for (a) active power of electricity demand, (b) reactive power of electricity demand, and (c) PV power production; (each gray line represents a scenario, whereas the red line represents the actual data for June 30th, 2021). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

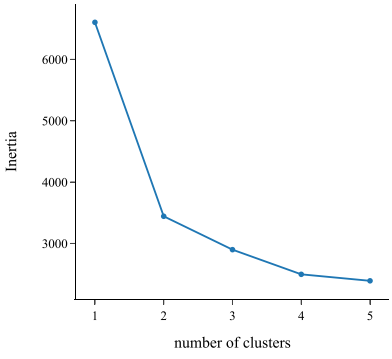


Fig. 5. Inertia comparison with different number of clusters.

5.2.3. P-value, mutual information, and log-likelihood

To finalize our analysis, we used three traditional metrics, i.e., *p*-value, mutual information, and log-likelihood, to assess the extent to which the generated scenarios accurately captured *typical* operations and *particular* conditions. These metrics allow us to evaluate the quality of the scenarios we generate.

P-value is a measure of the probability of obtaining a data that is at least as extreme as the observed data, given that the null hypothesis is true. In general, a small *p*-value (usually less than 0.05) suggests that the observed data is *non-typical* and provides strong evidence against the null hypothesis. On the other hand, a large *p*-value (usually greater than 0.05) suggests that the observed data is *typical* and provides weak evidence against the null hypothesis [29]. To calculate the *p*-value for two time series, you can use a t-test. To conduct a t-test in Python, you can use a t-test function from a statistical library, such as the `ttest_ind` function from the `scipy.stats` module.

Mutual information score is a measure of the mutual dependence between two variables. It quantifies the amount of information that scenarios contain about the *particular* data from the following day. A high mutual information score indicates a strong relationship between the two variables, while a low mutual information score indicates a weak relationship [29]. The

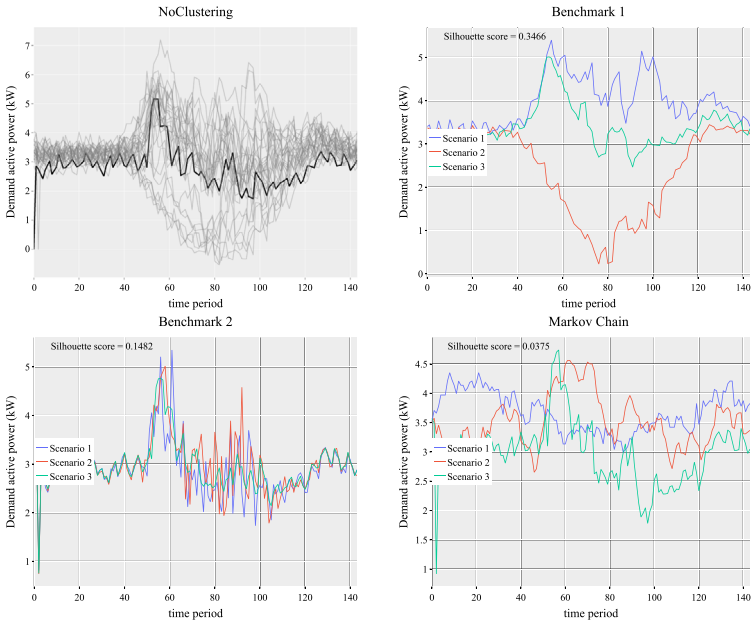


Fig. 6. The results for active power of electricity demands on July 30th, 2021.

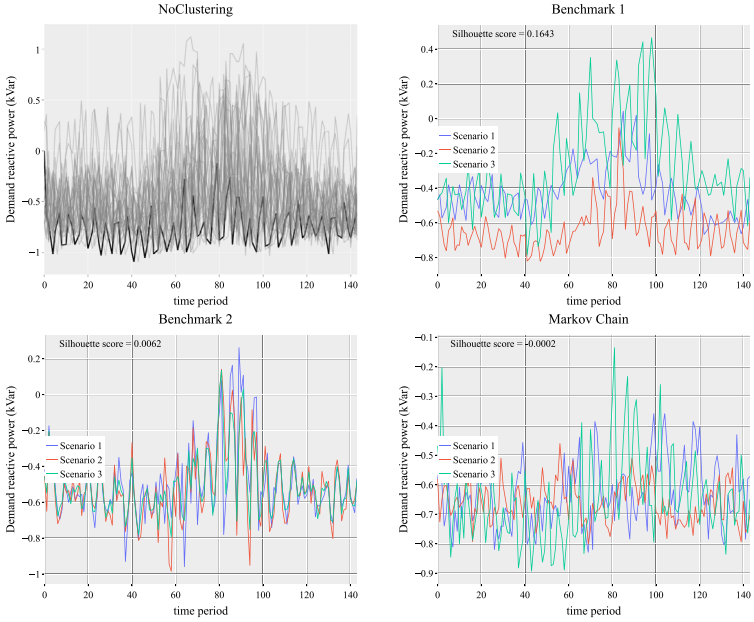


Fig. 7. The results for reactive power of electricity demands on July 30th, 2021.

mutual information score is calculated as the difference between the entropy of the joint distribution of two variables and the entropy of the marginal distribution of each variable. It is defined as:

$$\text{mutual information} = H(X) + H(Y) - H(X, Y), \tag{38}$$

where $H(X)$ is the entropy of the marginal distribution of X, $H(Y)$ is the entropy of the marginal distribution of Y, and $H(X, Y)$ is the entropy of the joint distribution of X and Y. Entropy is a measure of the uncertainty or randomness of a distribution. It is calculated as the sum of the probabilities of the events in the distribution multiplied by the logarithm of the probabilities. To calculate the mutual information score, we use a function from a statistical library, called `mutual_info_score` from the `sklearn.metrics` module.

Log-likelihood is a measure of the goodness of fit of a statistical model to the *particular* data. It is defined as the logarithm of the likelihood function, which is a measure of the probability of the data given the model. A high log-likelihood score indicates a good fit of the model to the actual data of the next day, while a low log-likelihood score indicates a poor fit. To calculate the log-likelihood score in Python, we use the `log_likelihood` function from the `sklearn.metrics`module [30].

In Fig. 9, we compare the p-value and mutual information score of various scenario selection strategies. Each point represents a scenario of a generated time series for PV power production, active power of electricity demands, and reactive power of electricity demands. From the distribution of mutual information scores and p-values for the mentioned approaches, we have the following observations:

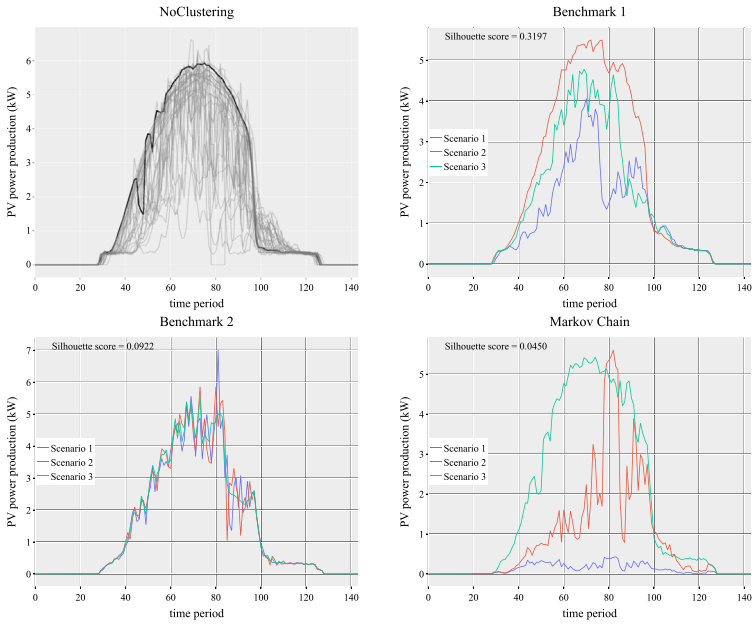


Fig. 8. The results for PV power production on July 30th, 2021.

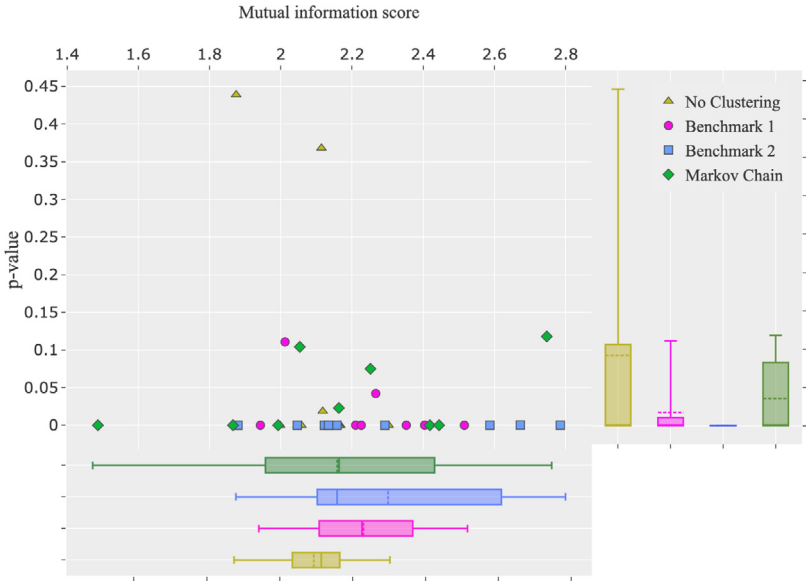


Fig. 9. Comparing *p*-value and mutual information for the scenarios generated by each scenario selection strategy.

- The *p*-value for the no clustering approach is higher than the others, indicating that what is expected to happen in the next day is not significantly different from the previous historical data. Thus, we will have more *typical* scenarios using this strategy.
- Although the *p*-value for the proposed Markov chain approach is lower than the no clustering approach, it has a higher mutual information score with the actual data of the next day. It is because the proposed approach forecasts the data of the next day, resulting in a higher mutual information score and scenarios closer to the *particular* condition of the next day.
- While the benchmark 2 has the highest mutual information, it has the lowest *p*-value. This indicates that the benchmark 2 forecasts better the *particular* conditions of the next day compared to other strategies; however, it generates a *non-typical* time series.

In Fig. 10, we show the log-likelihood of each scenario selection strategy's generated scenarios and the real data measured the next day. From this figure, we conclude that the proposed Markov chain strategy has a good fit between the model and the actual data as it has a greater log-likelihood.

5.3. Stochastic optimization results

Using the resulting representative scenarios, we solve the stochastic optimization problem. Although the prices might fluctuate over time, we assume fixed prices of $\lambda_t^{(p,PL)} = 1$ CHF/kWh, $\lambda_t^{(q,PL)} = 0$ CHF/kVARh, $\lambda_t^{(p\uparrow)} = 0.1$ CHF/kWh, $\lambda_t^{(p\downarrow)} = 0.1$ CHF/kWh, $\lambda_t^{(q\uparrow)} = 0.05$ CHF/kVARh, and $\lambda_t^{(q\downarrow)} = 0.05$ CHF/kVARh for all $t \in \mathcal{T}$. Furthermore, we considered VOLL = 6 CHF/kWh. Regardless, the selection of prices and VOLL will have no influence on findings of this paper.

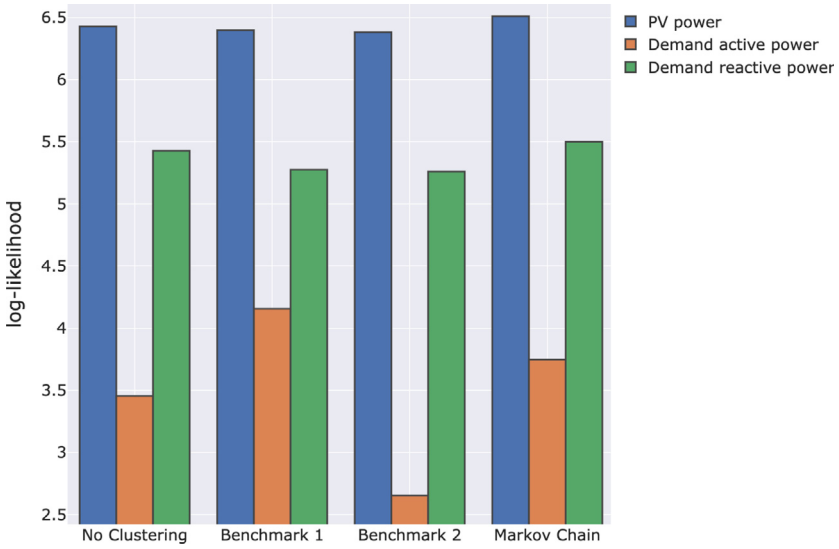


Fig. 10. Comparing the likelihood of the time series generated by each strategy with the measured data.

Table 2
Results of proposed stochastic optimization for different days of the year.

	Objective (CHF)				Deviation (%)			
	No Clustering	Benchmark 1	Benchmark 2	Markov Chain	No Clustering	Benchmark 1	Benchmark 2	Markov Chain
31/Jan.	36.343	3.104	39.987	40.507	6.670	0.165	3.815	3.737
28/Feb.	−0.306	18.999	24.823	17.598	10.236	3.591	16.707	6.698
31/Mar.	9.400	24.542	55.932	22.575	6.743	4.026	10.046	6.951
30/Apr.	12.831	40.625	52.527	56.025	7.271	7.037	13.630	10.265
31/May	32.390	26.139	57.004	40.264	3.463	3.415	0.076	3.860
30/Jun.	12.100	30.309	64.421	26.875	11.389	0.241	7.158	2.383
31/Jul.	14.173	16.583	54.397	27.551	3.473	0.067	0.628	4.131
31/Aug.	27.013	29.857	52.557	27.500	13.681	4.336	7.361	6.735
30/Sep.	−0.003	7.210	44.467	−0.094	0.070	3.433	3.382	0.321
31/Oct.	−0.623	4.075	28.311	11.096	0.165	0.324	6.711	0.048
30/Nov.	34.072	−0.310	43.442	−0.056	3.360	30.044	20.107	3.416
31/Dec.	−0.318	−0.049	−0.098	−0.063	0.102	6.679	13.359	10.050
Average	14.756	16.757	43.147	22.481	5.552	5.280	8.582	4.883

Table 2 and 11 show the resulting objectives for four cases, namely no clustering, benchmarks 1, 2, and proposed strategy,[4] and for different days of the year. In addition, the table and figure show the percentage deviation from the scheduled plan when data for the next day is available, as defined below.

$$\text{deviation} = \frac{ND}{T} \times 100\%, \tag{39}$$

where ND is the number of time instants that we would have deviation from the scheduled plan (considering the constraints of networks and resources).

The column of deviations demonstrates the robustness of the proposed solution in each case. According to Fig. 11, the Markov chain-based scenario selection strategy has the lowest deviation in average and distribution; however, the objective is not significantly greater than benchmark 1.

Figs. 12 also depict the planned flexibilities of different methods of scenario selection strategy for the 30th of July 2021.

We can deduce the following results from the Figs. 11–12 and Table 2:

- The proposed scenario selection strategy has the lowest average deviation in Table 2, which means that the results are more robust to the uncertainties. However, for the advantage of robustness against uncertainties, the objective of the proposed strategy is less than the benchmark 2.
- The first benchmark considers scenarios that are similar, and the second benchmark model is even more optimistic, and thus, we wrongly assumed that the average consumption could be less (see Figs. 12-b and 12-c).
- The green and read areas in Fig. 12 represent the flexibility capacity sold by the network operator to the TSO for each method. As one can see, the upward active power flexibility is set to 0 for all cases, as generating active power and selling the energy is worth more.
- Although the planned active and reactive power, as well as flexibility, for the proposed strategy and the first benchmark have similar patterns (with more average consumption in the proposed strategy), there is a difference in the distribution of deviation from the request of the TSO (0.4% less deviation in the case of the proposed strategy in average, as shown in Fig. 11). Note that the proposed scenario selection strategy leads to 7.2% less deviation by comparing the worst case of Markov chain and no clustering strategy (3.4% less deviation by comparing the worst case of Markov chain and benchmark 1). This is due to the fact that the scenarios generated by the proposed strategy results in a more realistic scenarios (e.g., 85% accuracy level for July 30th, 2021).

[4] The results of benchmarks 1, 2, and the proposed strategy are presented for K=3 clusters. If we increase the number of clusters to 4, the average of the objectives decreases by 2.6% while the average deviation does not change by more than 0.1% for the proposed strategy. It demonstrates that the number of clusters of 3 was a reasonable choice.

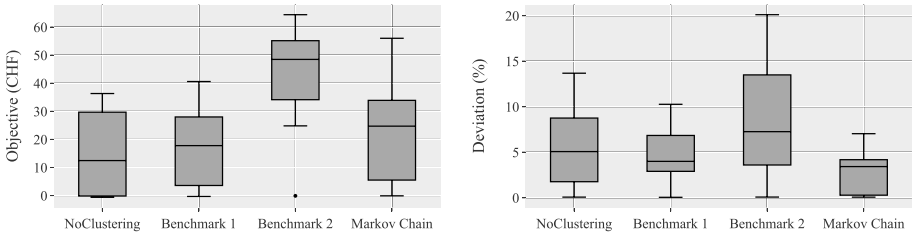


Fig. 11. Distribution of results for different days of the year.

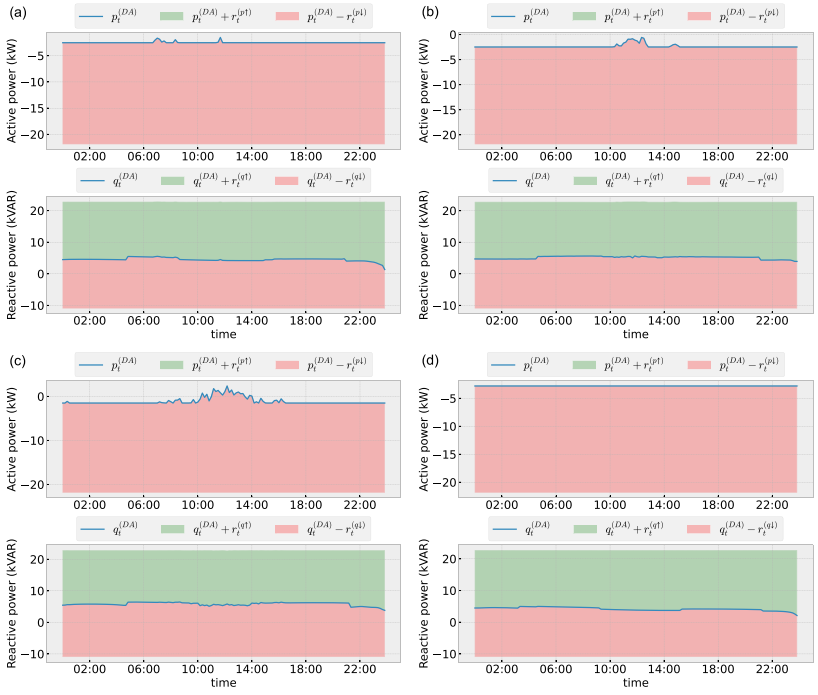


Fig. 12. The planned active and reactive power, and flexibilities of (a) no clustering, (b) benchmark 1, (c) benchmark 2, (d) and proposed scenario selection strategy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

The distributed energy resources (DERs) in an active distribution network (ADN) are used to plan active and reactive power as well as the flexibility, ensuring security and providing high-quality flexibility with less deviation in real-time. To achieve this, we have formulated a stochastic optimization problem. However, in order to solve this problem efficiently with limited computational resources, we have developed a scenario selection strategy based on historical data.

The proposed scenario selection strategy in the case study is compared to traditional methods, such as those that use real-world data and auto-regressive integrated moving average (ARIMA) forecasting. The proposed strategy has several advantages:

(i) It generates a large number of scenarios quickly, making it useful in situations where time and resources for data generation are limited.
(ii) It is able to generate a *limited* number of realistic scenarios that are close to the *particular* time series of the next day and *typical* information from the previous days.
(iii) It allows for the incorporation of uncertain parameters and constraints, while also allowing for the flexibility of an active distribution network.

Additionally, the proposed method has a lower computational complexity and includes robust constraints. Its scalability makes it suitable for use in active distribution networks with a large number of buses.

CRediT authorship contribution statement

Mohammad Rayati: Conceptualization, Methodology/Study design, Software, Formal analysis, Writing – original draft, Visualization. **Mokhtar Bozorg:** Conceptualization, Methodology/Study design, Validation, Formal analysis, Resources, Writing – review & editing, Project administration, Funding acquisition. **Mauro Carpita:** Methodology/Study design, Resources, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Rachid Cherkaoui:** Methodology/Study design, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

An attachment with a reference to the developed algorithm's code.

Appendix

All codes and simulations are available on a GitHub repository [31].

References

[1] R. Gupta, F. Sossan, M. Paolone, Countrywide PV hosting capacity and energy storage requirements for distribution networks: The case of Switzerland, Appl. Energy 281 (2021) 116010.

[2] M. Nick, R. Cherkaoui, M. Paolone, Optimal planning of distributed energy storage systems in ADNs embedding grid reconfiguration, IEEE Trans. Power Syst. 33 (2) (2017) 1577–1590.

[3] A. Sadeghi, A. Shahsavari, H. Haghighat, H. Mohsenian-Rad, Optimal market participation of distributed load resources under distribution network operational limits and renewable generation uncertainties, IEEE Trans. Smart Grid 10 (4) (2018) 3549–3561.

[4] G. Migliavacca, et al., SmartNet: H2020 project analysing TSO–DSO interaction to enable ancillary services provision from distribution networks, CIRED Open Access Proc. J. 2017 (1) (2017) 1998–2002.

[5] K. Oikonomou, M. Parvania, R. Khatami, Deliverable energy flexibility scheduling for ADNs, IEEE Trans. Smart Grid 11 (1) (2019) 655–664.

[6] F. Sossan, et al., Achieving the dispatchability of distribution feeders through prosumers data driven forecasting and MPC of electrochemical storage, IEEE Trans. Sustain. Energy 7 (4) (2016) 1762–1777.

[7] J. Engels, B. Claessens, G. Deconinck, Combined stochastic optimization of frequency control and self-consumption with a battery, IEEE Trans. Smart Grid 10 (2) (2017) 1971–1981.

[8] A. Zecchino, et al., Optimal provision of concurrent primary frequency and local voltage control from a BESS considering variable capability curves, Electr. Power Syst. Res. 190 (2021) 106643.

[9] S. Bjarghov, M. Kalantar, R. Cherkaoui, H. Farahmand, Battery degradation-aware congestion management in local flexibility markets, in: 2021 IEEE Madrid PowerTech, IEEE, 2021, pp. 1–6.

[10] X. Yang, et al., Flexibility provisions in ADNs with uncertainties, IEEE Trans. Sustain. Energy 12 (1) (2020) 553–567.

[11] F. Conte, S. Massucco, G.P. Schiapparelli, F. Silvestro, Day-ahead and intra-day planning of integrated BESS–PV systems providing frequency regulation, IEEE Trans. Sustain. Energy (2019).

[12] M. Rayati, M. Bozorg, R. Cherkaoui, Coordinating strategic aggregators in an ADN for providing operational flexibility, Electr. Power Syst. Res. 189 (2020) 106737.

[13] M. Kalantar, F. Sossan, M. Bozorg, R. Cherkaoui, Characterizing the reserve provision capability area of ADNs: A linear robust optimization, IEEE Trans. Smart Grid 11 (3) (2019) 2464–2475.

[14] M. Kalantar, R. Cherkaoui, Coordinating distributed energy resources and utility-scale battery energy storage system for power flexibility provision under uncertainty, IEEE Trans. Sustain. Energy (2021).

[15] J. Li, J. Zhou, B. Chen, Review of wind power scenario generation methods for optimal operation of renewable energy systems, Appl. Energy 280 (2020) 115992.

[16] M.C. Campi, S. Garatti, The exact feasibility of randomized solutions of uncertain convex programs, SIAM J. Optim. 19 (3) (2008) 1211–1230.

[17] S. Bolognani, F. Dörfler, Fast scenario-based decision making in unbalanced distribution networks, in: 2016 Power Syst. Computation Conf, PSCC, IEEE, 2016, pp. 1–7.

[18] L. Wu, L. Jiang, X. Hao, Optimal scenario generation algorithm for multi-objective optimization operation of ADN, in: 2017 36th Chinese Control Conf, CCC, IEEE, 2017, pp. 2680–2685.

[19] M. Rayati, P.D. Falco, D. Proto, M. Bozorg, M. Carpita, Generation data of synthetic high frequency solar irradiance for data-driven decision-making in electrical distribution grids, Energies 14 (16) (2021) 4734.

[20] M. Nick, R. Cherkaoui, J. Le Boudec, M. Paolone, An exact convex formulation of the optimal power flow in radial distribution networks including transverse components, IEEE Trans. Automat. Control 63 (3) (2017) 682–697.

[21] A. Cabrera-Tobar, et al., Capability curve analysis of photovoltaic generation systems, Sol. Energy 140 (2016) 255–264.

[22] M. Nick, M. Bozorg, R. Cherkaoui, M. Paolone, A robust optimization framework for the day-ahead scheduling of active distribution networks including energy storage systems, in: Milan PowerTech, IEEE, 2019.

[23] M.L. Baughman, S.N. Siddiqi, Real-time pricing of reactive power: Theory and case study results, IEEE Trans. Power Systems 6 (1) (1991) 23–29.

[24] W.W. Hogan, Independent system operator: Pricing and flexibility in a competitive electricity market, 1998, Center for Business and Government, JF Kennedy School of Government, Harvard University, MA.

[25] C. Lin, C. Fang, Y. Chen, S. Liu, Z. Bie, Scenario generation and reduction methods for power flow examination of transmission expansion planning, in: 7th Conf. Power and Energy Syst, ICPES, IEEE, 2017, pp. 90–95.

[26] M. Rayati, T. Pidancier, M. Carpita, M. Bozorg, State estimation for medium and low voltage distribution grids based on near real-time grid measurements and delayed smart meters data, in: 2020 22nd European Conf. Power Electronics and Applications, IEEE, 2020, pp. P–1.

[27] M. Carpita, J. Affolter, M. Bozorg, D. Houmard, S. Wasterlain, ReIne, a flexible laboratory for emulating and testing the distribution grid, in: 21st European Conf. Power Electronics and Applications, IEEE, 2019.

[28] T.M. Kodinariya, P.R. Makwana, Review on determining number of cluster in K-means clustering, Inter. J. 1 (6) (2013) 90–95.

[29] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, Inc., 2012.

[30] J. Grus, Data Science from Scratch: First Principles with Python, O'Reilly Media, 2019.

[31] M. Rayati, DiGriFlex, 2022, https://github.com/mohammadrayati/DiGriFlex/blob/master/tests/mc_paper.py.