Article:
Yin, M, Li, K orcid.org/0000-0001-6657-0522 and Yu, J (2022) A data-driven approach for microgrid distributed generation planning under uncertainties. Applied Energy. 118429. ISSN 0306-2619

https://doi.org/10.1016/j.apenergy.2021.118429

A Data-driven Approach for Microgrid Distributed Generation Planning under Uncertainties

Mingjia Yin[a], Kang Li[a,*], James Yu[b]

[a]*School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK*
[b]*SP Energy Networks, Glasgow G2 5AD, UK*

Abstract

The increasing demand for power system decarbonization and resilience raises the necessity of incorporating the renewable distributed generation (DG) into the microgird planning. The complexity of the microgrid renewable DG planning largely roots from the intermittent wind and solar energy and load variations throughout the planning period. This paper proposes a novel two-stage data-driven adaptive robust distributed generation planning (DDARDGP) framework considering both grid-connected and islanded modes of microgrids, wherein the overall system cost is minimized. By leveraging the spatio-temporal property of historical weather and grid information, a compact uncertainty set is developed based on a data-driven Bayesian nonparametric approach. The problem is further solved by a modified column and constraint generation (CC&G) algorithm. In the study, the effectiveness of the proposed framework is demonstrated using a modified IEEE 33-bus test system. The case study considers the optimal generation sizing, allocation and mixtures. The simulation results confirm that the proposed data-driven uncertainty set adapts well to the increase of data dimensions and solves the over-conservatism issue, leading to 34.14% reduction in uncertainty estimation compared with the traditional budget uncertainty set. Accordingly, the total cost can achieve a $23,185 reduction under the proposed DDARDGP framework.

Keywords: Distributed generation planning, Data-driven uncertainty set,

*Corresponding author
Email addresses: elmyi@leeds.ac.uk (Mingjia Yin), K.Li1@leeds.ac.uk (Kang Li)

Nomenclature

A. Sets and Indices

i	Index for buses.
j	Index for types of a certain DG category.
l	Index for incremental number of a certain DG type.
t	Index for time period.
k	Index for basic uncertainty set.
$\mathbb{F}$	Binary feasible set for binary indicators $f_{i,j,l}^{mt,wt,pv}$.
$\mathbb{Y}$	Feasible set for continues variables $P_{i,t}^{mt,wt,pv}$, $Q_{i,t}^{mt,wt,pv}$, $P_{1,t}^{+,-}$, $P_{i,t}$,$Q_{i,t}$ and $V_{i,t}$.
$\mathbb{U}$	Uncertainty Set for wind power output $\omega_{i,j,l,t}^{wt}$, solar power output $\omega_{i,j,l,t}^{wt}$ and load demand $P_{i,t}^{c}$, $Q_{i,t}^{c}$. It is an union of T basic uncertainty sets $\mathbb{U}_k$.

B. Parameters for DDARDGP framework

$\alpha^{fc,emi}$	Fuel price and emission penalty price of MT (\$/kWh).
ρ^{c}	Price for selling electricity to customers (\$).
$\rho^{s,b}$	Selling price and purchasing price at which the electricity is traded from microgrid to the main grid (\$).
σ	Emission factor of MT (kg/kWh).
τ	Minimum requirement for DG Capacity in microgrid for islanded operation.

$C^{mt,wt,pv}$	One discrete increment of capacity of MT, WT and PV (kW).
FF	Fill factor of solar cell.
I_{max}	Line current limit (kA).
$inv^{mt,wt,pv}$	Capital cost of MT,WT and PV ($/kW).
K_v/K_i	Voltage temperature coefficient ($V/°C$) and current temperature coefficient ($A/°C$).
$om^{mt,wt,pv}$	O&M cost of MT,WT and PV ($/kWh).
$pf_{min}^{mt,wt,pv}$	Lower limit of power factor of MT, WT and PV.
r_i/x_i	Resistance and reactance between bus i to bus $i+1$.
$s_{av,t}$	Average solar irradiance at time t (kW/m^2).
T_a	Ambient temperature of PV cells at time t ($°C$).
T_c/T_{no}	Cell temperature and nominal operating temperature of cell ($°C$).
$v_{av,t}$	Average wind speed (m/s) at time t.
$v_{ci}/v_{ci}/v_r$	Cut-in speed, cut-off speed and rated wind speed of WT (m/s).
V_{min}/V_{max}	Lower and upper limit of voltage (kV).
V_{MPP}/I_{MPP}	Voltage (V) and current (A) at maximum power point
V_{oc}/I_{sc}	Open-circuit voltage (V) and short circuit current (A).
$\mathbf{A}, \mathbf{J}$	Parameter matrices for binary variables $\mathbf{f}$ in the DDARDGP framework
$\mathbf{B}, \mathbf{G}$	Parameter matrices for continues variables $\mathbf{y}$ in the DDARDGP framework

$\mathbf{C}, \mathbf{E}, \mathbf{H}$	Parameter matrices for uncertain variables $\mathbf{u}$ in the DDARDGP framework
$\mathbf{m}, \mathbf{n}$	Vectors of scalars in the DDARDGP framework

C. Variables for DDARDGP framework

$\omega_{i,j,l,t}^{mt,wt,pv}$	lth incremental active power generation from jth type of MT,WT and PV of bus i at time t, respectively.
C^{EMI}	Emission penalty cost of MT ($).
C^{FC}	Fuel cost of MT ($).
C^{INV}	Investment cost of all DG units($).
C^{OM}	O&M cost of all DG units ($).
C^{REV}	Total revenue by selling electricity to customers and the main grid ($).
$f_{i,j,l}^{mt,wt,pv}$	Binary indicators that represent whether the lth increment of jth type of MT,WT or PV is installed at bus i. 1 if installed, otherwise 0.
$P_{i,t}^{c}/Q_{i,t}^{c}$	Active/reactive load demand of bus i at time t.
$P_{1,t}^{+,-}$	Power surplus and deficiency of microgrid at time t.
$P_{i,t}/Q_{i,t}$	Active/reactive power flow from bus i to bus $i+1$ at time t.
$P_{i,t}^{mt,wt,pv}$	Total Active power generation from MT,WT and PV of bus i at time t.
$Q_{i,t}^{mt,wt,pv}$	Total Reactive power generation from MT,WT and PV of bus i at time t.
$V_{i,t}$	Voltage of bus i at time t.

D. Other symbols for data-driven uncertainty set

λ_k	Weight of kth mixture component of DPMM.
λ_{th}	Truncation threshold value of the weight.
$\mu_k, \kappa_k, \nu_k, \Psi_k$	Parameters of kth normal Wishart pior distribution.
$\mu'_k, \kappa'_k, \nu'_k, \Psi'_k$	Parameters of kth normal Wishart posterior distribution.
$\overline{x_k}$	mean value of kth mixture component of DPMM.
ζ	Scaling factor to adjust the size of the basic uncertainty set.
a_k, b_k	Parameters of kth beta distribution.
n_k	Data size of kth mixture component of DPMM.

E. Abbreviation

BD	Benders decomposition
CC&G	Column and constraint generation
DDARDGP	Data-driven adaptive robust distributed generation planning
DG	Distributed generation
DPMM	Dirichlet process mixture model
MILP	Mixed-integer linear programming
MPPT	Maximum power point tracking
MT	Micro-turbine
O&M	Operation and maintenance
PCC	Point of common coupling
PV	Photovoltaic
RES	Renewable energy source
RO	Robust optimization

SO	Stochastic optimization
SVC	Support vector clustering
WT	Wind turbine

1. Introduction

The concept of microgrids arises from the proliferation of distributed resources. It is defined as a cluster of systematically grouped distributed generations (DGs), transmission facilities and interconnected loads served wihtin a clearly defined area. The small-scale and self-containing network can therefore operate in grid-connected mode or grid-off (islanded) mode, bringing significant techno-economical benefits to the power system such as the reduction of grid energy loss [1], the improvement of system efficiency [2] as well as the power supply flexibility [3], etc.

DG can generate energy on-site, rather than transmitting the bulk of energy from centralized power plants [4]. This advantage makes DG a great supporter for the development of microgrids, such as the network expansion in remote areas and the rural electrification in developing countries. In traditional regime, small-scale dispatchable DG units like micro-turbines (MTs) and diesel engines can seamlessly be integrated into the grid with the utility's permission. In recent years, renewable energy sources (RESs), e.g. wind turbines and solar panels, are reaching the top topics for the global awareness of decarbonization [5, 6, 7]. However, RESs such as wind and solar energy are naturally intermittent and variable due to geographical restrictions. It is significantly complex to predict their power generation [8]. Adding excessive RESs might cause system reliability problems and cascading blackouts. For DG planning, which aims to determine the optimal sizing and location of multiple DG units in response to the demand they serve, the problem becomes harder in order to maintain the balance of electricity supply and demand for a long period.

There exists a wide variety of studies to address the optimal sizing and sitting problem of DG units. In [9], a single-objective optimization problem is proposed

to minimize the system's annual power loss by improving the voltage profile. The study in [10] extends the problem to a multi-objective one considering the cost of facility's investment, DG sources and energy loss. However, all the above work did not consider the reactive capabilities of DG units and the impacts of the uncertainties of RES power production and load-side demand.

Over the past few years two most popular optimization modelling techniques have been proposed to address the uncertainties arising from the variable renewable generations and loads: stochastic optimization (SO) and robust optimization (RO). SO techniques model the uncertainties by a statistical model. To reflect the probability distribution of uncertainty accurately, methods such as scenario-based modeling [11], Monte Carlo simulation and point estimation have been widely used. In [12], the RES and load uncertainties have been divided into 16 sampled scenarios for a joint distribution network and renewable energy expansion planning. The same techniques have been adopted by [13]. A two-stage stochastic programming is developed to replace the full representation of the uncertainties of the hydro production by a subset of scenarios. Although the model adequately represents the time-dependent quantities, it incurs a considerably heavy computational burden while it can not guarantee that the construction of scenario samples is accurate enough to represent the future trends.

Recent studies [14, 15, 16] have shown that RO approaches could be successful applied to the power system operation and planning problems by improving the operation robustness against the uncertainties. Compared with SO, RO modelling method has several prominent advantages. First, RO can obtain an optimal solution against uncertainty that is deterministically formulated by uncertainty sets, which do not need accurate probability distribution in advance. Second, unlike the SO using probabilistic assumptions, the worst scenario reaches the top of the queue in RO. However, the solution of a single-stage RO is always criticized to be over conservative. To address the issue, in this paper we present a two-stage adaptive robust optimization framework for the DG planning in microgrids. The first-stage solves the DG allocation problem

while the second-stage considers other operation related variables and uncertainties. It not only reduces the computational burdens, but also immunizes the system against operational security issues. The applications of the adaptive RO framework in power industry could be found in unit commitment problem [17, 18] and DG planning problem [15]. In [19], various methods to construct uncertainty sets are discussed, including box uncertainty set, ellipsoidal uncertainty set, polyhedral uncertainty set and budget uncertainty set. The study in [15, 20] models the load consumption, wind and solar power output with budget uncertainty sets. This kind of uncertainty sets can reveal the fluctuation of variables. But it needs additional effort to identify appropriate boundaries.

In order to solve the over-conservatism of the aforementioned uncertainty sets, the data-driven uncertainty set based on historical data has been proposed in recent years. A parametric model of uncertainty sets for wind generation is proposed in [21]. Such kind of methods uses a fixed and finite number of parameters to model the existing data. But it can suffer from over- or underfitting of data when there is a misfit between the complexity of the model and the amount of data available. In [21, 18, 20], wind resource availability at different sites is considered to be the same, which is not true in practice. Only a few studies [22] have been done on large renewable generation and load datasets. It still remains a challenge to find an adaptive method to represent the temporal correlations and spatial heterogeneity of data.

Accordingly, in the paper we will propose a two-stage data-driven adaptive robust distributed generation planning (DDARDGP) framework for microgrids and the overall objective covers the investment cost, operation and maintenance (O&M) cost of DGs, fuel cost, and emission penalty cost of fossil fuel sourced generation units, and the costs of electricity exchange with the main grid. A novel Bayesian nonparametric solution will be proposed to construct the uncertainty set. The proposed method is flexible enough to capture the distribution of high-dimensional uncertain data. Outliers will be removed without incurring additional computation complexity. A modified column and constraint generation (CC&G) algorithm is sequentially developed to solve the problem.

The major contributions of this paper are summarized as follows:

- A multi-objective DG system configuration design is proposed for microgrids considering both the active and reactive capabilities of DG units;
- A novel data-driven uncertainty set construction method is developed for high-dimensional data with temporal and spatial information;
- A modified CC&G algorithm for DDARDGP framework is designed;
- The proposed data-driven uncertainty set is evaluated and results are compared with the budget uncertainty set.

The remainder of this paper is organized as follows. In Section 2, problem formulation including the objective function and constraints of two-stage DDARDGP framework is presented. The data-driven uncertainty set is developed in Section 3. In section 4, an effective solution algorithm - the modified CC&G is designed. Section 5 presents the simulation results of a case study. Finally, Section 6 summarizes the effectiveness of the proposed DDARDGP framework and concludes the paper.

2. Problem Formulation

Generally speaking, a DG planning design aims to find the best combinations from possible generation technologies, such as biomass generation, wind turbine (WT), solar photovoltaic (PV), diesel engine, internal combustion engine and gas turbine, etc.. Recently, wind and solar generations have been widely deployed in microgrids, especially for rural areas and islands with rich natural resources. High-efficiency WT and PV are considered as non-dispatchable DGs, which means that all the energy they produce will be fully utilized. Microturbines (MTs) have capabilities such as full controllability, high response speed and low emission [23, 24]. Thus, through suitable power electronic devices, i.e. inverter interfaces, these DG units could be connected to the microgrids.

2.1. Objective Function

The optimal integration of DGs into a microgrid has many benefits from the operational, environmental and techno-economical aspects. Each category of DG i.e. WT, PV and MT) has several different performance parameters to choose from. For example, if there are three types of WTs possibly installed at the same location, they can produce three different power generation profiles based on their parameters. The objective is to determine the optimal locations, sizes and mixtures of DGs with the purpose to maximize the long-term economic profits. Assume that the utility of the microgrid owns the all infrastructure, including the DG units deployment and operation, energy transactions with the main grid and regional customers, the total cost of the utility can be formulated as follows:

$$\min_{f,P,Q,V} C^{INV} + C^{OM} + C^{FC} + C^{EMI} - C^{REV} \tag{1}$$

where

$$C^{INV} = \sum_i \sum_j \sum_l inv^{mt} f^{mt}_{i,j,l} C^{mt}_j + \sum_i \sum_j \sum_l inv^{wt} f^{wt}_{i,j,l} C^{wt}_j + \sum_i \sum_j \sum_l inv^{pv} f^{pv}_{i,j,l} C^{pv}_j \tag{2}$$

$$C^{OM} = \sum_t \left[\sum_i \sum_j \sum_l om^{mt}_t f^{mt}_{i,j,l} C^{mt}_j + \sum_i \sum_j \sum_l om^{wt}_t f^{wt}_{i,j,l} C^{wt}_j + \sum_i \sum_j \sum_l om^{pv}_t f^{pv}_{i,j,l} C^{pv}_j \right] \tag{3}$$

$$C^{FC} = \alpha^{fc} \sum_t \sum_i P^{mt}_{i,t} \tag{4}$$

$$C^{EMI} = \alpha^{emi} \sigma \sum_t \sum_i P^{mt}_{i,t} \tag{5}$$

$$C^{REV} = \rho^c \sum_t \sum_i P^c_{i,t} + \rho^s \sum_t P^+_{1,t} - \rho^b \sum_t P^-_{1,t} \tag{6}$$

$$P_{1,t} = P^-_{1,t} - P^+_{1,t}, \forall P^+_{1,t}, P^-_{1,t} \geqslant 0 \tag{7}$$

The objective function (1) consists of five parts: investment cost (C^{INV}), operation and maintenance (O&M) cost of DGs (C^{OM}), fuel cost (C^{FC}) and emission penalty cost of fossil fuel sourced MTs (C^{EMI}) and the total revenue of the microgrid (C^{REV}). Equation (2) is the investment cost that represents the amortized one-time up-front costs of all categories and types of DGs, where inv^{mt}, inv^{wt} and inv^{pv} are capital investments for MT, WT and PV ($\$/kW$), respectively. C^{mt}, C^{wt} and C^{pv} are one discrete increment of capacity (kW) of MT, WT and PV, respectively. O&M cost in Equation (3) represents costs of operation and maintenance, consumables and replacement elements of all DGs during the planning horizon. om_t^{mt}, om_t^{wt} and om_t^{pv} are O&M costs for certain types of DG ($\$/kWh$) at time t. Fuel cost and emission penalty cost of MTs are described in Equation (4) and (5) where α^{fc}, α^{emi} and σ are fuel price ($\$/kWh$), emission price ($\$/kWh$) and emission factor of MT. Equation (6) describes the total revenues that are achieved by selling the electricity to customers in the microgrid at price ρ^c ($) and profits through the energy transactions with the main grid. The price ρ^b ($) is the sale price at which the mail grid sells electricity to the microgrid, and ρ^s ($) is called the feed-in tariff at which microgrid sells electricity back to the main grid. The electricity transaction is defined by Equation (7). $P_{1,t}^{+}$ and $P_{1,t}^{-}$ are power surplus and deficiency of the microgrid at time t respectively. The binary indicators $f_{i,j,l}^{mt}$, $f_{i,j,l}^{wt}$ and $f_{i,j,l}^{pv}$ are decision variables that represent whether the kth increment of jth type of DG (including microturbine, wind turbine, and PV unit) is deployed at bus i. The other continuous variables P, Q, V introduced in the following section will make the problem a mixed-integer linear programming (MILP) formulation.

2.2. Constraints

2.2.1. Power Flow Model

DistFlow Equations are the most popular conventional power flow formulation for the planning problem in a radial distribution system proposed by M. Baran and F. Wu in 1999 [25]. The calculations are based on the Ohm's law

and power balance as follows:

$$P_{i+1,t} = P_{i,t} - r_i \frac{P_{i,t}^2 + Q_{i,t}^2}{V_{i,t}^2} - P_{i+1,t}^c + P_{i+1,t}^{mt} + P_{i+1,t}^{wt} + P_{i+1,t}^{pv} \tag{8}$$

$$Q_{i+1,t} = Q_{i,t} - x_i \frac{P_{i,t}^2 + Q_{i,t}^2}{V_{i,t}^2} - Q_{i+1,t}^c + Q_{i+1,t}^{mt} + Q_{i+1,t}^{wt} + Q_{i+1,t}^{pv} \tag{9}$$

$$V_{i+1,t} = V_{i,t} - 2r_i P_{i,t} - 2x_i Q_{i,t} + (r_i^2 + x_i^2) \frac{P_{i,t}^2 + Q_{i,t}^2}{V_{i,t}^2} \tag{10}$$

where $P_{i,t}$ and $Q_{i,t}$ are active and reactive power flows from bus i to bus $i+1$ respectively. r_i and x_i are resistance and reactance between bus i and bus $i+1$. $P_{i,t}^c$ and $Q_{i,t}^c$ are load demand at bus i. $P_{i,t}^{mt}$, $P_{i,t}^{wt}$ and $P_{i,t}^{pv}$ are corresponding active power generations of a certain category of DG at bus i. $Q_{i,t}^{mt}$, $Q_{i,t}^{wt}$ and $Q_{i,t}^{pv}$ are reactive power output of DG at bus i. $V_{i,t}$ is the voltage at bus i, The subscript t indicates the corresponding time.

LinDistFlow is a linearized formulation of DistFlow that overcomes the complexity of quadratic equations and the non-convex feature of DistFlow, and has been widely adopted for the planning of radial distribution systems and microgrids [26, 27] because of its simplicity and availability. In the microgrid, LinDistFlow formulations among bus bars [27] are guaranteed by Constraints (11) - (13).

$$P_{i+1,t} = P_{i,t} - P_{i+1,t}^c + P_{i+1,t}^{mt} + P_{i+1,t}^{wt} + P_{i+1,t}^{pv} \tag{11}$$

$$Q_{i+1,t} = Q_{i,t} - Q_{i+1,t}^c + Q_{i+1,t}^{mt} + Q_{i+1,t}^{wt} + Q_{i+1,t}^{pv} \tag{12}$$

$$V_{i+1,t} = V_{i,t} - (r_i P_{i,t} + x_i Q_{i,t})/V_0 \tag{13}$$

where V_0 is the reference voltage at the point of common coupling (PCC).

The voltage at each bus should be within the utility's permissible limit, which is usually $\pm 5\%$ deviation from reference bus V_0

$$V_{min} \leqslant V_{i,t} \leqslant V_{max} \tag{14}$$

The current of each line $I_{i,t}$ is maintained within the limit

$$|I_{i,t}| \leqslant I_{max} \tag{15}$$

2.2.2. Operation Constraints

To achieve robust operation, operating constraints of each time segment must be taken into account in the planning horizon. We will have

$$0 \leq P_{i,t}^{mt} \leq \sum_{j} \sum_{l} f_{i,j,l}^{mt} C_{j}^{mt} \tag{16}$$

$$P_{i,t}^{wt} = \sum_{j} \sum_{l} f_{i,j,l}^{wt} \omega_{i,j,l,t}^{wt} \tag{17}$$

$$P_{i,t}^{pv} = \sum_{j} \sum_{l} f_{i,j,l}^{pv} \omega_{i,j,l,t}^{pv} \tag{18}$$

Constraints (16) - (18) represent the active power production requirements for all DGs. MT is a dispatchable generation while WT and PV are non-dispatchable units. It is assumed that every lth incremental active power produced by jth type WT ($\omega_{i,j,l,t}^{wt}$) is an uncertain variable and they will be consumed during the planning horizon. The same assumption holds for PV ($\omega_{i,j,l,t}^{pv}$). To deal with them, a specific technique will be introduced in the next section to determine the uncertainty set.

$$pf_{min}^{mt} \leqslant pf^{mt} \leqslant 1 \tag{19}$$

$$pf_{min}^{wt} \leqslant pf^{wt} \leqslant 1 \tag{20}$$

$$pf_{min}^{pv} \leqslant pf^{pv} \leqslant 1 \tag{21}$$

$$Q^{mt} = \pm P^{mt} \cdot tan\left(cos^{-1}(pf^{mt})\right) \tag{22}$$

$$Q^{wt} = \pm P^{wt} \cdot tan\left(cos^{-1}(pf^{wt})\right) \tag{23}$$

$$Q^{pv} = \pm P^{pv} \cdot tan\left(cos^{-1}(pf^{pv})\right) \tag{24}$$

Constraints (19) - (24) represent the requirements for inverter-based DGs. The reactive power Q^{mt},Q^{wt} and Q^{pv} are adjustable based on the power factor pf^{mt},pf^{wt} and pf^{pv} required by the utility.

Constraint (25) is the operational requirement for the microgrid at the islanded mode [15]. The installed capacity of DG units should be higher than or equal to a certain percentage of load within the microgrid. Therefore, τ represents the minimum requirement for DG capacity.

$$\tau \sum_{t}\sum_{i} P^{c}_{i,t} <= \sum_{t}\sum_{i}\sum_{j}\sum_{l} f^{mt}_{i,j,l} C^{mt}_{j} + \sum_{i}\sum_{j}\sum_{l} f^{wt}_{i,j,l} C^{wt}_{j} + \sum_{i}\sum_{j}\sum_{l} f^{pv}_{i,j,l} C^{pv}_{j} \tag{25}$$

2.3. Adaptive Robust DG Planning Model

The Equation (1) can be reformulated into a two-stage adaptive robust optimization modelling framework as shown below in Equation (26). The adaptive robust optimization framework has already been widely used in power system applications, such as unit commitment [28], economic dispatch [29] , reactive power optimization [30], etc., to immunize against the uncertainty of RES generation and load demand. The first-stage cost only covers the investment cost, O&M cost of DG units, and the second-stage costs will include the remaining fuel cost, emission penalty cost and electricity cost exchanged with the main grid and regional customers. The problem can be generalized as:

$$\underset{\mathbf{f}\in\mathbb{F}}{min}\ \mathbf{A}^{\mathsf{T}}\mathbf{f} + \underset{\mathbf{u}\in\mathbb{U}}{max}\ \underset{\mathbf{y}\in\mathbb{Y}(\mathbf{f},\mathbf{u})}{min}\ \mathbf{B}^{\mathsf{T}}\mathbf{y} + \mathbf{C}^{\mathsf{T}}\mathbf{u} \tag{26}$$

s.t.

$$\mathbf{Dy} + \mathbf{Eu} = \mathbf{m} \tag{27}$$

$$\mathbf{Jf} + \mathbf{Gy} + \mathbf{Hu} \leq \mathbf{n} \tag{28}$$

where $\mathbf{f}$ denotes the vector of all first-stage decisions including $\mathbf{f}^{mt} = [f^{mt}_{i,j,l}]$, $\mathbf{f}^{wt} = [f^{wt}_{i,j,l}]$ and $\mathbf{f}^{pv} = [f^{mt}_{i,j,l}]$ $\forall i, j, l$. The binary set $\mathbb{F} = \{0,1\}^{3*I_b*J_g*L_i}$ imposes

constraints on L_i increments of J_g types of MT, WT and PV in a microgrid with I_b candidate buses. $\mathbf{A}$ and $\mathbf{J}$ are corresponding parameter matrices for the binary variables $\mathbf{f}$. Meanwhile in the second stage, $\mathbb{U}$ represents the uncertainty set for the vector $\mathbf{u}$ that includes the wind and solar power output $\omega^{\mathbf{wt}} = [\omega^{wt}_{i,j,l,t}]$, $\omega^{\mathbf{pv}} = [\omega^{pv}_{i,j,l,t}]$ $\forall i, j, l, t$, and the active & reactive load demand $\mathbf{P^c} = [P^c_{i,t}]$, $\mathbf{Q^c} = [Q^c_{i,t}]$ $\forall i, t$. $\mathbf{C}$, $\mathbf{E}$ and $\mathbf{H}$ are parameter matrices for the uncertain variables $\mathbf{u}$. $\mathbb{Y}(\mathbf{f}, \mathbf{u}) = \{(\mathbf{f}, \mathbf{u}) | Eqns(27, 28)\}$ represents the feasible region for the vector of other continuous variables $\mathbf{y} = [\mathbf{P^{mt}}; \mathbf{P^{wt}}; \mathbf{P^{pv}}; \mathbf{Q^{mt}}; \mathbf{Q^{wt}}; \mathbf{Q^{pv}}; \mathbf{P}; \mathbf{Q}; \mathbf{V}]$ where $\mathbf{P^{mt}} = [P^{mt}_{i,t}]$, $\mathbf{P^{wt}} = [P^{wt}_{i,t}]$, $\mathbf{P^{pv}} = [P^{pv}_{i,t}]$, $\mathbf{Q^{mt}} = [Q^{mt}_{i,t}]$, $\mathbf{Q^{wt}} = [Q^{wt}_{i,t}]$, $\mathbf{Q^{pv}} = [Q^{pv}_{i,t}]$, $\mathbf{Q} = [Q_{i,t}]$, $\mathbf{V} = [V_{i,t}]$ $\forall i, t$ and $\mathbf{P} = [P^+_{1,t}, P^-_{1,t}, P_{i,t}]$ $\forall i, t, i \neq 1$. The corresponding parameter matrices in the objective function and inequality constraints are represented by $\mathbf{B}$ and $\mathbf{G}$. $\mathbf{m}$ and $\mathbf{n}$ are the vectors for the remaining scalars in equality constraints and inequality constraints, respectively.

3. Data-driven Uncertainty Set

3.1. Renewable Energy Resource Availability

Suppose the utility and related authorities have access to the empirical database that stores meteorological data from on-site measurements, i.e. solar irradiance and wind speed. One of the major work prior to the DG planing is the thorough assessment of the RES availability because of the intermittent nature of RES-based DGs.

3.1.1. Wind Turbine Output Power Model

Wind power generation models can be classified into three categories. The first one comprises of fluid dynamics that considers air density, swept area of wind turbine rotor and power coefficient, wind speed and electro-mechanical conversion efficiency. The second category is a non-liner model based on wind speed on-the-spot. The third category is the regression model or AI-based model developed from the average normalised power curve of historical data, which is not available for a new area without commercial MTs.

In the study, we will use the second type. Compared with the first one, the complexity of the model is simpler with fewer parameters at the cost of slightly reduced model accuracy. The output power of a wind turbine is dependent on the average wind speed $v_{av,t}$ at the site as well as the parameters of the power performance curve. Therefore, once the speed data is gathered for a specific time segment t, the output power of each WT type is calculated according to its corresponding characteristics:

$$\omega_t^{wt} = \begin{cases} 0 & 0 \leq v_{av,t} < v_{ci} \\ \frac{(v_{av,t} - v_{ci})}{(v_r - v_{ci})} * P_r^{wt} & v_{ci} \leq v_{av,t} < v_r \\ P_r^{wt} & v_r \leq v_{av,t} < v_{co} \\ 0 & v_{co} \leq v_{av,t} \end{cases} \tag{29}$$

where P_r^{wt} is the rated power generation of one wind turbine, and v_{ci}, v_{co}, v_r are the wind turbine cut-in, cut-off and rated wind speeds respectively.

3.1.2. PV Module Output Power Model

Similar to WT, there are also several popular generation models, such as the model related to solar irradiance, solar PV array area, and solar PV system efficiency and empirical regression models. Here, we will use a simplified model based on the current-voltage (I-V) relationship [31]. The output power of the PV module is assumed to be working under maximum power point tracking (MPPT) which is usually the case when they are connected to an inverter. The output power is dependent on the solar irradiance and ambient temperature on the site as well as the characteristics of the module itself. Therefore, once the solar irradiance and ambient temperature data is collected for a specific time period, the output power for different type of solar cells could be calculated using the following equation:

$$\omega_t^{pv} = FF * V * I \tag{30}$$

$$FF = \frac{V_{MPP} * I_{MPP}}{V_{oc} * I_{sc}} \tag{31}$$

$$T_c = T_a, t + s_{av,t}\left(\frac{T_{no} - 20}{0.8}\right) \tag{32}$$

$$V = V_{oc} - K_v * T_c \tag{33}$$

$$I = s_{av}\left[I_{sc} + K_i\left(T_c - 25\right)\right] \tag{34}$$

where FF is the fill factor of solar cell. T_c is the cell temperature ($^\circ C$). T_a is the ambient temperature ($^\circ C$). s_{av} is the average solar irradiance (kW/m^2). K_v and K_i represent voltage temperature coefficient ($V/^\circ C$) and current temperature coefficient ($A/^\circ C$). T_{no} is the nominal operating temperature of cell ($^\circ C$). I_{sc} and V_{oc} are short circuit current (A) and open-circuit voltage (V). I_{MPP} and V_{MPP} are current (A) and voltage (V) at the maximum power point.

3.2. Load Profile Analysis

Load profile analysis is another important prerequisites for DG planning. In order to satisfy the stable operation of the microgrid, the power generation must be matched with the load from hour to hour. Thanks to the smart meter technology, for the network expansion nowadays, it is more common for the utility to get access to the records of customers' load profiles. For instance, there will be a dataset shown as Figure 1. $D = \left\{D^1, D^2, ..., D^t\right\}$ is a vector of measurements of demand at different load points at time slot t. We will have $D^t = \{d_1^t, d_2^t, ..., d_i^t\}$ where d_i^t is the demand at ith location at time slot t. The temporal-spatial data is shown in Figure 1. The whole database will continuously store the data collected from n different locations for a period of time and mark them with specific location labels. Each day the smart meter will automatically record the power consumption where it is installed at reprogrammed time interval. Therefore, each day there will be nt sample loads in the database.

The time span of the database determines the reliability and accuracy of the uncertainty set of P_i^c. For instance, on the one-day basis it is assumed that there is no variation in the generation and profile from season to season and

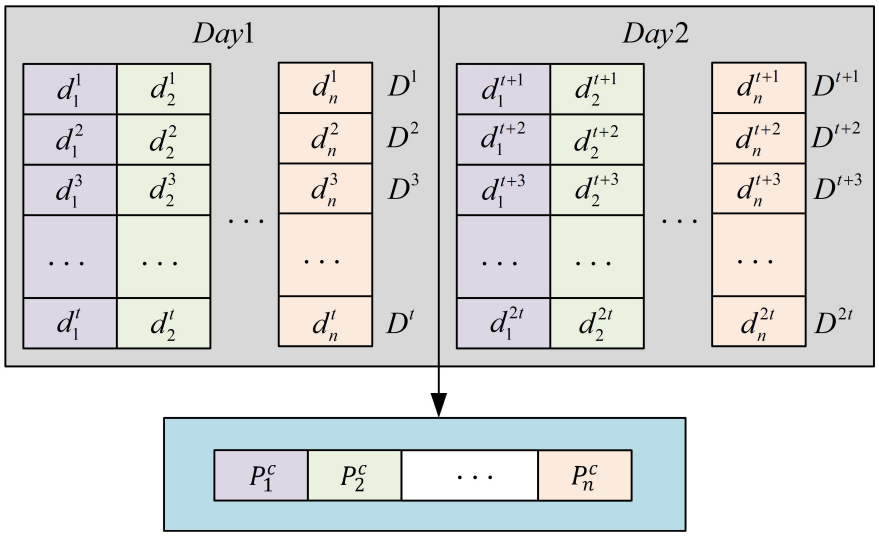


Figure 1: Schematic diagram of the database of a load profile.

from weekday to weekend, and from weekday to weekend. There is no doubt that longer time span will contain more information. For a long time span, it is attractive to find out an effective machine-learning method to cope with large amounts of high-dimensional data.

3.3. Dirichlet Process Mixture Model

Data-driven uncertainty set based on the Bayesian nonparametric approach is an alternative to parametric modeling and selection. It defines prior and posterior distributions on a single fixed parameter space. As opposed to parametric models, the dimension of the parameter space is adaptive to data size. Dirichlet process mixture model (DPMM) is a popular clustering implemention of Bayesian nonparamatirc models. Unlike other machine-learning algorithms, i.e. K-means, Gaussian mixure model and Support Vector Clustering (SVC) [32], it can naturally adapt the number of clusters to the complexity of the data to avoid overfitting.

The Dirichlet process is the fundamental prior on probability distributions for DPMM. It is a distribution over distributions, i.e. each draw from a Dirichlet

process is itself a distribution. Let G_0 be a base measure, and α be a positive scalar. A random measure G can be generated as a Dirichlet process $G \sim DP(\alpha, G_0)$. For any partition $\{A_1, ..., A_k\}$, it fits the Dirichlet distribution.

$$(G(A_1), ..., G(A_k)) \sim Dir(\alpha G_0(A_1), ..., \alpha G_0(A_k)) \tag{35}$$

The construction of measure G follows the stick-breaking procedures [33]. Let $\eta_k \sim G_0$ for $k = \{1, 2, ...\}$ and δ_{η_k} denotes the Dirac delta indicator function at η_k. G is given by the expression

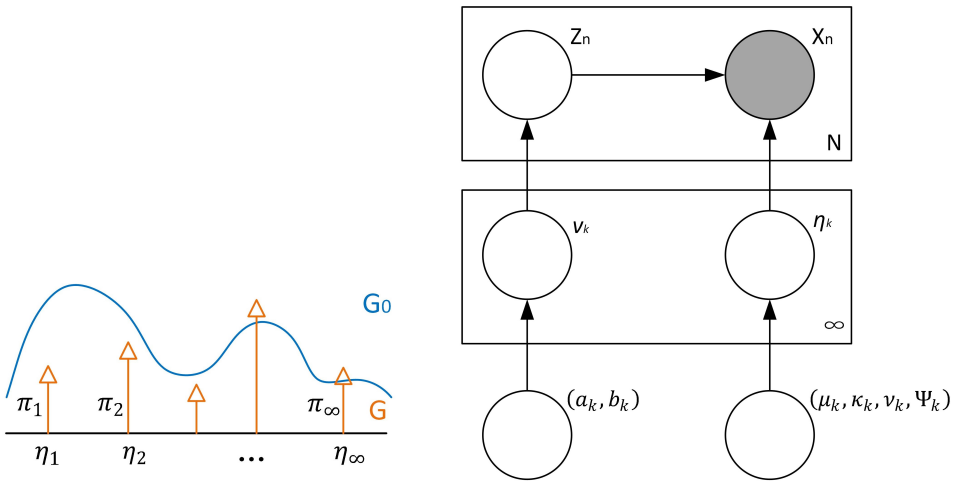
$$\pi_k(v) = v_k \prod_{j=1}^{k-1}(1 - v_j) \tag{36}$$

$$G = \sum_{k=1}^{\infty} \pi_k(v)\delta_{\eta_k} \tag{37}$$

where the stick lengths $\pi_k(v)$ are given by repeatedly breaking a stick of initial length 1 at points η_k. The length v_k draws from a beta distribution $Beta(a_k, b_k) \in (0, 1)$. Therefore, the stick could be broken into infinite number of pieces. Analogously, G is a mixture of infinite discrete components δ_{η_k} with proportions $\pi_k(v)$ as shown in Figure 2(a). We assign a new variable Z_n for corresponding nth sampled data x_n to identify the mixture component from a multinomial distribution. The generation of DPMM for observed dataset $X_n = \{x_1, x_2, ..., x_n\}$ can be summarized as

1. For each mixture component
   (a) Draw $v_k|(a_k, b_k) \sim Beta(a_k, b_k)$ for $k = \{1, 2, ...\}$
   (b) Draw $\eta_k|G_0 \sim G_0$ for $k = \{1, 2, ...\}$
2. For each data point
   (a) Draw $Z_n|\{v_1, v_2, ...\} \sim Mult(\pi(v))$
   (b) Draw $X_n|z_n \sim p(x_n|\eta_{z_n})$

Suppose each sampled data x_n is generated from a different distribution η_k with its own parameters. In the work, we choose a multivariate probability distribution- normal Wishart (NW) distribution that has four parameters



(a) (b)

Figure 2: (a) Construction of random measure G for DPMM; (b) Schematic diagram of model representation of DPMM.

$\mu_k, \kappa_k, \nu_k, \Psi_k$, that is, $\eta_k \sim NW(\mu_k, \kappa_k, \nu_k, \Psi_k)$. The distribution over distribution η_k follows G which is generated based on the length $v_k \sim Beta(a_k, b_k)$.

The graphical model is shown in Figure 2(b) where the latent parameter sets $\Xi = \{a_k, b_k, \mu_k, \kappa_k, \nu_k, \Psi_k\}$ for $k = \{1, 2, ...\}$. Due to its computational efficiency and deterministic nature, variational inference has been used to approximate the likelihood and posterior distributions in DPMM from observed data [33]. Therefore, we will have a group of parameter settings $\Xi = argmax\ p(a_k, b_k, \mu_k, \kappa_k, \nu_k, \Psi_k | X_n)$.

3.4. Formulation of Data-driven Uncertainty Set

The construction of the Bayesian nonparametric uncertainty set based on DPMM is to find out the most valuable mixtures that represent the posterior predictive distribution of $p(x_{n+1}|X_n, \Xi)$. Because of the great nature of conjugate distribution, the posterior distribution is in the same probability distribution family as the NW prior. The posterior predictive distribution of NW distribution follows Student's t-distribution with $\nu_k + 1 - dim(X_k)$ degrees of

freedom [34]. It can be written as

$$p(x_{n+1}|X_n,\Xi) \sim \sum_k \lambda_k \mathbf{St}_{\nu'_k+1-dim(X_k)}\left(\mu'_k, \frac{\kappa'_k+1}{\kappa'_k(\nu'_k+1-dim(X_k))}\Psi'^{-1}_k\right) \tag{38}$$

$$\lambda_k = \begin{cases} \dfrac{a_k}{a_k+b_k}\displaystyle\prod_{j=1}^{k-1}\frac{b_j}{a_j+b_j}, & k=1...,T-1 \\ 1-\displaystyle\sum_{j=1}^{T-1}\lambda_k, & k=T \end{cases} \tag{39}$$

$$\mu'_k = \frac{\kappa_k\mu_k + n_k\overline{x_k}}{\kappa_k+n_k} \tag{40}$$

$$\kappa'_k = \kappa_k + n_k \tag{41}$$

$$\nu'_k = \nu_k + n_k \tag{42}$$

$$\Psi'^{-1}_k = \Psi^{-1}_k + S_k + \frac{\kappa_k n_k}{\kappa_k+n_k}(\mu_k-\overline{x_k})(\mu_k-\overline{x_k})^{\mathsf{T}} \tag{43}$$

$$S_k = \sum_{i=1}^{n}(x_i-\overline{x_k})(x_i-\overline{x_k})^{\mathsf{T}} \tag{44}$$

where λ_k is the weight of kth mixture component of DPMM. The weight is calculated based on the stick length in Equation (36), indicating the importance of the kth mixture component. $\mu'_k, \kappa'_k, \nu'_k, \Psi'_k$ are the parameters of NW posterior distribution. n_k and $\overline{x_k}$ are the data size and the mean value belongs to mixture component k.

Suggest a λ_{th} as the truncation threshold value, only mixture components with weights larger than λ_{th} will be considered. Therefore, the Bayesian nonparametric uncertainty set $\mathbb{U}$ is an union of T basic uncertainty sets as follows

$$\mathbb{U} = \mathbb{U}_1 \cup \mathbb{U}_2 ... \cup \mathbb{U}_T \tag{45}$$

The method determines the number of clusters systematically and automatically rather than specifying this number as a priori. Each basic uncertainty set $\mathbb{U}_k$ for component k is defined based on t value

$$\mathbb{U}_k = \{x_{n+1} | x_{n+1} = \mu'_k + \zeta l_k \Lambda, \qquad \|\Lambda\|_n \leq 1\} \tag{46}$$

where $l_k = \sqrt{\frac{\kappa'_k+1}{\kappa'_k(\nu'_k+1-dim(X_k))}} \Psi'^{-1/2}_k$. Λ is the latent uncertainty and can choose norm 1,2 or ∞. ζ is a scaling factor that adjusts the confidence interval of the uncertainty set.

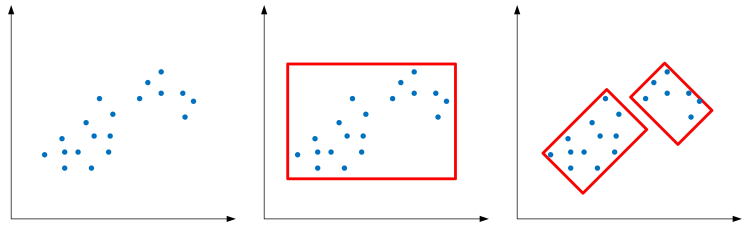


Figure 3: (a) Scatter plot of 2-dimensional data; (b) budget uncertainty set; (c) proposed data-driven uncertainty set.

To better explain the proposed data-driven uncertainty set, an example is presented below. Figure 3(a) shows the schematic of dimensional data. Traditional budget uncertainty set could find the worst case by making it fall in the upper or lower bounds as shown in Figure 3(b). However, the set is easy to be overly conservative. As shown in in Figure 3(c), the data-driven uncertainty set is tighter than the budget uncertainty set. It captures the correlated, asymmetric characteristics of the underlying 2-dimensional data.

The proposed construction of uncertainty set is cast as a union of multiple basic uncertainty sets. Its geometric shape is flexible enough to capture the distribution of high-dimensional uncertain data. Outliers will be removed through a self-adaptive truncation process without additional computation complexity. One of the significant highlights is that the proposed uncertainty sets is suitable for multi-dimensional data processing. As the latent uncertainty Λ establishes the underlying relationship between multidimensional data, a polyhedral set is

devised for each basic uncertainty set, which benefit the resulting DDARDGP problem with enormous computational efficiency.

4. Solution Methodologies

The proposed two-stage DDARDGP model (26)-(28) is very difficult to compute. In [35], two kinds of solving methods have been discussed. One is the Benders decomposition (BD) technique and the other is the CC&G algorithm. Compared with the Benders-style cutting plane methods, the CC&G algorithm is a general procedure with a unified approach to deal with optimality and feasibility. It is proved that the latter performs an order of magnitude faster on a two-stage robust problem [35].

To apply the CC&G in the proposed linear formulation model, the uncertainty is chosen to be a polyhedron. The optimal solution for the maximization problem over u is an extreme point of one of the basic polyhedral uncertainty set $\mathbb{U}_k$. In view of this, we can solve the optimization problem by enumerating all the extreme points. However from the aspect of computational efficiency, just think about the master problem will generate 2^{n_f} possibilities where n_f is the number of first-stage variables. It is prohibitively challenging to solve such a problem with external enumeration of uncertain variables. Therefore, the problem is spilt to two stages. In the first-stage, the master problem (MP) is formulated to generate significant scenarios.

(MP)

$$\min_{\mathbf{f}\in\mathbb{F}} \ \mathbf{A}^{\mathsf{T}}\mathbf{f} + \xi \tag{47}$$

s.t.

$$\xi \geq \mathbf{B}^{\mathsf{T}}\mathbf{f} + \mathbf{C}^{\mathsf{T}}\mathbf{u} \tag{48}$$

$$\mathbf{Df} + \mathbf{Ef} = \mathbf{m} \tag{49}$$

$$\mathbf{Jf} + \mathbf{Gy} + \mathbf{Hu} \leq \mathbf{n} \tag{50}$$

The result of the master problem is a relaxation of the original one as the constraints of continuous and uncertain variables are not defined. Thus, it yields a lower bound LB and the result of the master problem is recorded as $\hat{\mathbf{f}}$. Further in the second-stage, the remaining part that does not involve binary variables is decomposed into a set of subproblems $(SP_k, k = 1, ..., T)$ for later stage. Each subproblem SP_k corresponds to a basic uncertainty set $\mathbb{U}_k$ as introduced in Equation (46). The subproblem can be defined as:

(SP_k)

$$\max_{\mathbf{u}\in\mathbb{U}_k} \min_{\mathbf{y}\in\mathbb{Y}\{\mathbf{f},\mathbf{u}\}} \mathbf{B}^\mathsf{T}\mathbf{y} + \mathbf{C}^\mathsf{T}\mathbf{u} \tag{51}$$

s.t.

$$\mathbf{D}\mathbf{y} + \mathbf{E}\mathbf{u} = \mathbf{m} \tag{52}$$

$$\mathbf{J}\mathbf{f} + \mathbf{G}\mathbf{y} + \mathbf{H}\mathbf{u} \leq \mathbf{n} \tag{53}$$

The subproblem leads to a series of upper bound UB_k to identify the worst-case scenario of uncertainties. The optimality cuts which is associated with worst-case uncertain variables are then added to the master problem. To solve the bi-level "max-min" subproblem which is difficult to be solved with open-shelf solvers, KKT optimality conditions [35] of the problem is formulated as follows:

$$\mathbf{D}\mathbf{y} + \mathbf{E}\mathbf{u} = \mathbf{m} \tag{54}$$

$$\mathbf{B}^\mathsf{T} + \gamma^\mathsf{T}\mathbf{G} + \delta^\mathsf{T}\mathbf{D} = 0 \tag{55}$$

$$\mathbf{J}\mathbf{f} + \mathbf{G}\mathbf{y} + \mathbf{H}\mathbf{u} \leq \mathbf{n} \tag{56}$$

$$\gamma^\mathsf{T}(\mathbf{n} - \mathbf{J}\mathbf{f} - \mathbf{G}\mathbf{y} - \mathbf{H}\mathbf{u}) = 0 \tag{57}$$

Constraint (57) is the complementary slackness condition, which means only one of γ and $\mathbf{n} - \mathbf{J}\mathbf{f} - \mathbf{G}\mathbf{y} - \mathbf{H}\mathbf{u}$ is 0. It can be transformed together with

Constraint (56) by the Big-M method [15] and yields

$$0 \leq \mathbf{n} - \mathbf{Jf} - \mathbf{Gy} - \mathbf{Hu} \leq M\sigma \tag{58}$$

$$0 \leq \gamma \leq M(1-\sigma) \tag{59}$$

where M is a large value and σ is a binary variable. KKT condition is sufficient when the subproblem problem is convex. The constraints becomes(54), (55), (58) and (59).

Unlike the conventional RO problem, the proposed DDARDGP model has combined multiple basic uncertainty sets, which cannot be solved by the original CC&G algorithm. The flow chart of a modified CC&G [14] algorithm for DDARDGP is shown in Figure 4. In each SP, the corresponding extreme points are solved iteratively for a certain MP result, which generates optimality cuts. In the case where SP is an infeasible problem, the infeasible cuts will be generated. Both infeasible cuts and worse-case optimality cuts will be added to the master problem.

The comprehensive procedure of the solving method can be summarized as follows:

Step 0: Initialization. Set the lower bound $LB = -\infty$, the upper bound $UB = +\infty$, the tolerance of the optimality gap ϵ and iteration number $g = 1$.
Step 1: Solve the master problem (MP) and obtain the optimal solution $(\mathbf{f}^g, \xi^g)$. Update the lower bound $LB = max\left\{LB, \mathbf{A}^\mathsf{T}\mathbf{f}^g + \xi^g\right\}$.
Step 2: Solve all the subproblems SP_k^g using $\mathbf{f}^g$ and obtain the optimal solution $\mathbf{Q}_k^g = (\mathbf{y}_k^g, \mathbf{u}_k^g)$.
Step 3: Compare $\mathbf{Q}_k^i$ for all $k = 1, 2, ...T$ and identify the worst-case scenario $\mathbf{Q}_*^g = (\mathbf{y}_*^g, \mathbf{u}_*^g)$ and infeasible scenario set $\mathbf{Q}_{if}^g = (\mathbf{y}_{if}^g, \mathbf{u}_{if}^g)$. Update the upper bound $UB = min\left\{UB, \mathbf{A}^\mathsf{T}\mathbf{f}^g + \mathbf{B}^\mathsf{T}\mathbf{y}_*^g + \mathbf{C}^\mathsf{T}\mathbf{u}_*^g\right\}$.
Step 4: Denote the optimality gap as $Gap = \left|\frac{UB-LB}{UB}\right|$. If the $Gap < \epsilon$, the algorithm terminates and output the optimal decision. Otherwise update the MP by adding new variables $\mathbf{y}_*^g$,$\mathbf{y}_{if}^g$ and corresponding optimality cuts

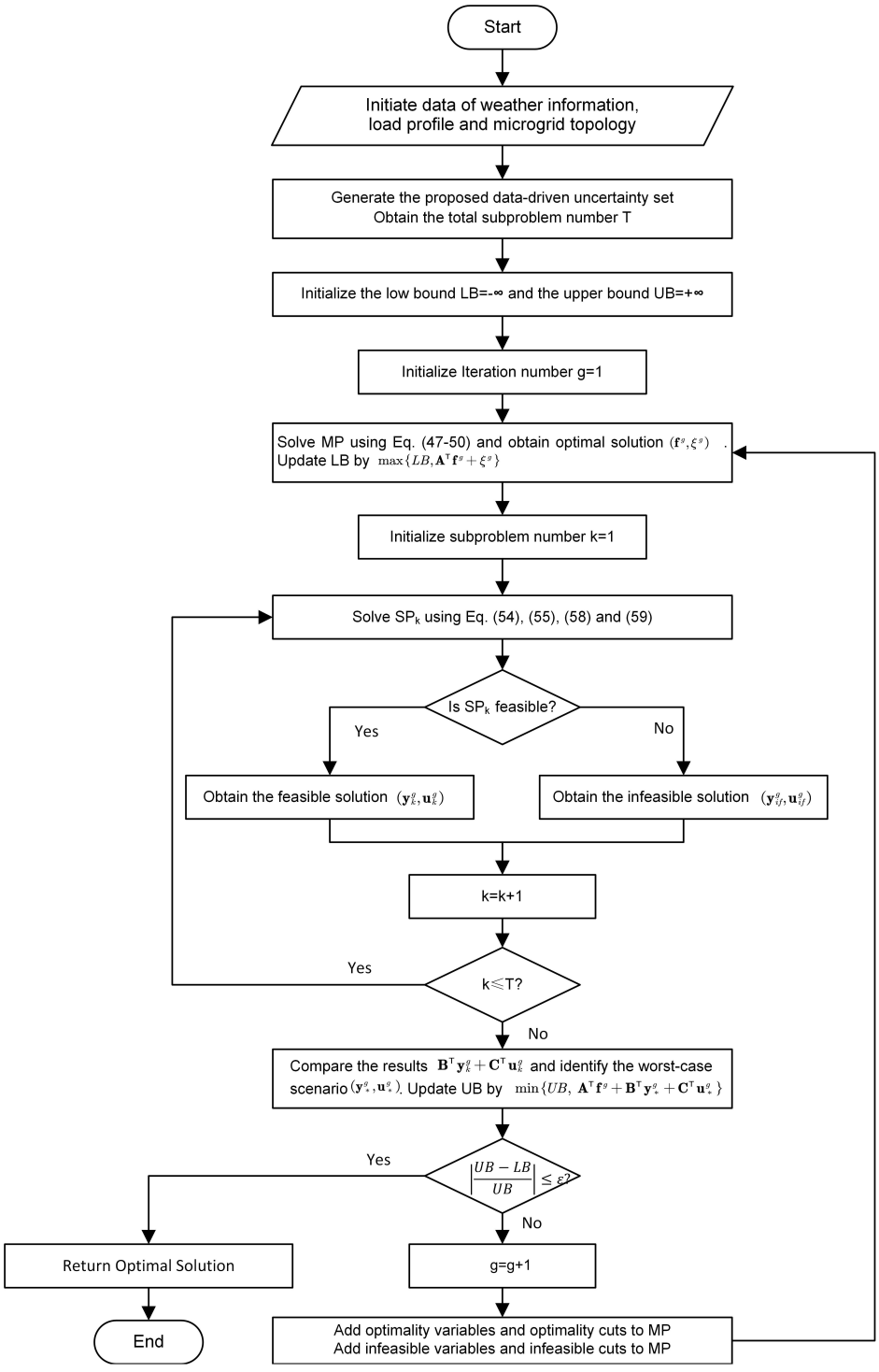


Figure 4: Flowchart of the solution algorithm for DDARDGP framework.

($\mathbf{Dy}_*^g + \mathbf{Eu}_*^g = \mathbf{m}$ and $\mathbf{Ff} + \mathbf{Gy}_*^g + \mathbf{Hu}_*^g \leq \mathbf{n}$) and infeasible cuts ($\mathbf{Dy}_{if}^g + \mathbf{Eu}_{if}^g = \mathbf{m}$ and $\mathbf{Ff} + \mathbf{Gy}_{if*}^g + \mathbf{Hu}_{if}^g \leq \mathbf{n}$). Update $g = g + 1$ and go back to Step 1.

5. Case Study: Distributed Generation Planning of a Microgrid

In the section, a modified IEEE 33-bus 12.66 kV distribution system is used as the test microgrid to validate the DG planning design with the DDARDGP framework. To find the optimal DG mix, the proposed methodologies will be analysed under various conditions.

5.1. System Under Study

The single line topology of the system is shown in Figure 5 and line data can be found in [36]. The total basic load is 3715 kW and 2300 kVAR. In the figure, the blue nodes are denoted as normal buses where load demands are connected. The red nodes are three possible sites (bus 4, 19 and 26) that not only have load demands connected but also can be chosen to install DG units due to its geographic advantages. Each possible site could install either wind generation, PV panels or fuel-based generation. The capital costs for MT, WT and PV generations are \$2293/kVA, \$1882/kVA and \$4004/kVA, respectively, while the O&M costs of the three DG units are \$0.012/kWh, \$0.01/kWh and \$0.01/kWh, respectively. For MT, the fuel cost and emission penalty cost are \$0.63/kWh and \$0.02kg/kWh, respectively. The emission factor is \$0.003 kg/kWh. All the aforementioned parameters are taken from [15].

For each different category of DG units, there are three different types. The related parameters and characteristics for WT and PV modules to calculate the hourly power generation are listed in Table 1. It is noted that the PV modules are working in a 15,000-series configuration, thus the peak power output of the three type are 750kW, 795kW and 900kW, respectively.

All DG units are assumed to work with 0.95 power factor according to grid code. The voltage range for all buses are set to [0.95, 1.05]. More than one type of individual DG categories are available, but there is a upper limit, for instance 2 in the case, because of land limitations.

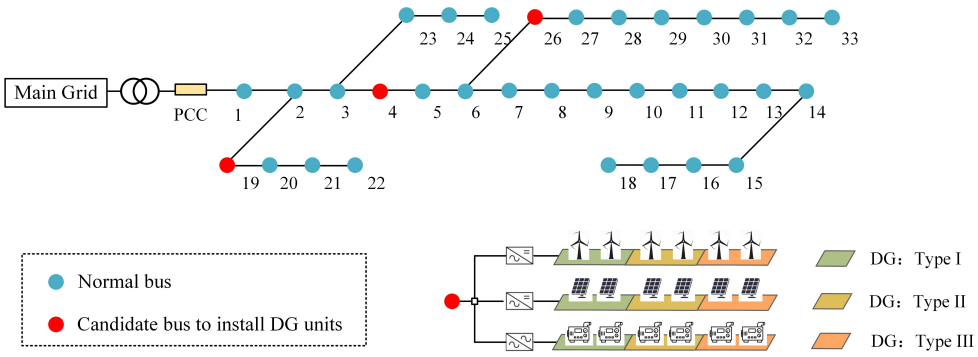


Figure 5: The schematic test microgrid system.

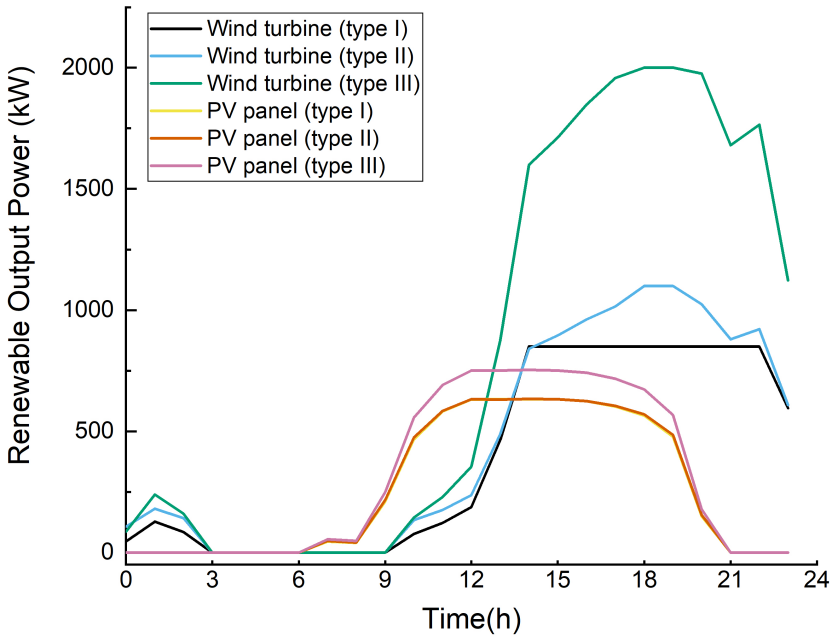


Figure 6: Output Power of different types of renewable generations.

Table 1: Characteristics of available Micro-turbines, wind turbines and PV modules.

Micro-turbine Parameters			
Parameters	Type I	Type II	Type III
Rated Power (kW)	1000	1500	2000
Wind Turbine Parameters			
Parameters	Type I	Type II	Type III
Rated Power(kW)	850	1100	2000
Cut-in speed (m/s)	4	3.5	4
Rated speed (m/s)	10	12	11.5
Cut-out speed (m/s)	25	24	25
PV Module Parameters [31]			
Parameters	Type I	Type II	Type III
Peak Power (W)	50	53	60
Open circuit voltage (V)	55.50	21.70	21.10
Short circuit current(A)	1.80	3.40	3.80
Voltage at maximum power point (V)	38.00	17.40	17.10
Current at maximum power point (A)	1.32	3.05	3.50
Voltage temperature coefficient ($mV/°C$)	194.00	88.00	75.00
Current temperature coefficient ($mA/°C$)	1.40	1.50	3.10
Nominal cell operating temperature ($°C$)	43.00	43.00	43.00

As defined in the previous section, fuel-based MT generations will operate as dispatchable units. Their performance is only related to their rated power outputs. In the meanwhile, all the power generated by WT and PV panels will be consumed by the microgrid, which will strongly be influenced by the weather condition (i.e. wind speed, solar irradiance and ambient condition) and unit parameters. Figure 6 illustrates the power generation profiles of the 6 type renewable generations of a typical day in the winter time. The fluctuation of the results reveals that no certain category of generation outperforms the others in all weather conditions, which proves the importance of DG mix selection, despite its sizing and location.

5.2. Evaluation of Uncertainty Set

The uncertainties in the proposed DDARDGP problem come from three factors: the power outputs of WT, PV and load demend of the microgrid. To predict the hourly renewable generation output, solar irradiance, ambient temperature and wind speed figures are needed according to models discussed in section 3.1. Here one year data at three weather stations obtained from SAURN (Southern African Universities Radiometric Network) [37] is used for three candidate buses 4, 19 and 26 respectively. The load profile is obtained from NYISO for all connected loads. Although in practice every consumer has its own energy usage preference, the analysis of the topic is beyond the scope of work here. The planner can easily change the setting if there are other load profiles. Therefore, the dimension of the corresponding uncertainties is $N_{bus} * TYPE_{WT} + N_{bus} * TYPE_{PV} + TYPE_{load} = 3 * 3 + 3 * 3 + 1 = 19$.

The DPMM-based uncertainty sets are a combination of infinite sets. It is adaptive to data size with a predefined truncation threshold value ω_{th}, which is the minimal weight of mixture components. Table 2 shows the results of set number and data coverage under different threshold values. It is noted that the rate of data coverage is negative correlated with the threshold. Although a smaller number of sets will decrease the computational complexity in DDARDGP framework for planning, the data distribution is not fully captured.

To find the balance between computational complexity and set accuracy, ω_{th} is set to be 0.02 in the case study.

Table 2: Uncertainty sets under different truncation threshold value ω_{th}.

ω_{th}	Set number of uncertainties	Data coverage rate (%)
0.006	11	99.45
0.007	11	99.27
0.008	11	98.78
0.009	10	98.76
0.01	9	97.77
0.02	9	97.18
0.03	8	93.61
0.04	7	93.13
0.05	7	89.53
0.06	6	83.86
0.07	5	77.05
0.08	4	72.34

The set-up of the polyhedron for each uncertainty set is based on Equation (46). The scaling factor ζ for confidence level is determined by looking up the t-distribution table. The degrees of each uncertainty set will vary from several hundred to more than two thousand. As the t value doesn't change much for degrees over 100, we could choose the value with the infinite degree for convenience.

5.3. Numerical Results of DG Planning for the Microgrid

Once the uncertainty sets are ready, the proposed modified CC&G algorithm will be used to solve DDARDGP problem. All the simulation results are implemented with CPLEX 12.8.0 using a computer with an Inter(R) Core(TM) i5-6400 CPU at 2.7 GHz.

5.3.1. Comparisons under different confidence levels

Figure 7 illustrates the results of cost of the utility which is responsible for the DG planning and 15-year operation of the microgrid under different confidence level. As shown from the bar chart, the red bar is the total cost and the blue bar represents the first stage cost which includes investment and O&M costs of installed DG units. The first stage cost of DDARDGP goes up with the increase of the confidence level of uncertainty set. The decision is made to deploy more DG units and import a large amount of electricity from the main grid in order to maintain the safety operation of the microgrid in island mode.

The gap between the total cost and first stage cost for all instances are negative, which is the second stage cost. It implies that the proposed DG deployment plan can gain positive profits through electricity arbitrage with the main grid and its consumer demand. The gaps shrink moderately with the increase of confidence level. This is due to the reason that wider confidence intervals lead to more conservative results. For confidence level over 90%, the cost hardly grows, which implies that the boundary reaches the maximum range. This phenomenon represents no "worse" case will be discovered, and it might lead to the trap of over-conservatism. Therefore, prior to the planning, confidence level should be carefully selected to balance between the model accuracy and conservatism. In the case study, the confidence level is set to 80%.

The deployment results under different confidence levels of the uncertainty set are listed in Table 3. It can be seen that the category of installed DG units are the same in all instances, which is WT. It is understandable that WT is the most cost-effective one based on the its low capital cost and O&M cost. In instances 1, 2 and 3, the total installed capacity is 2800kW with two type I WT and one type II WT installed. The highest capacity 3050 kW appears in instances 5 and 6. The chosen locations are different in the same system topology under various confidence levels. This is because the first-stage decision is made based on the worst-case scenario generated in the second stage of the DDARDGP framework.

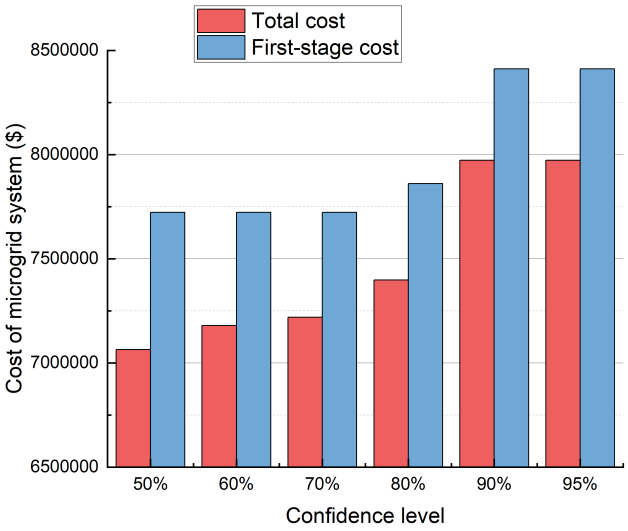


Figure 7: The second-stage cost of DDARDGP framework under different confidence levels.

Table 3: DG deployment under different confidence levels.

Instance	Confidence level	Location	Category	Capacity (kW)	Number	Total Capacity (kW)
1	50%	26	WT	850	2	2800
				1100	1	
2	60%	19	WT	850	2	2800
				1100	1	
3	70%	19	WT	850	2	2800
				1100	1	
6	80%	4	WT	850	1	2850
		26		2000	1	
5	90%	19	WT	850	1	3050
				1100	2	
6	95%	19	WT	850	1	3050
		26		1100	2	

Table 4 shows the computational results in each iteration of the modified CC&G algorithm. In all instances, the optimality gap converges in three iterations. It is noticeable that the gap narrows to less than 1% at the second iteration, which verifies the effectiveness of the proposed algorithm. For a large-scale system, the algorithm is guaranteed to work within a reasonable amount of time.

Table 4: Bounds and Gaps under different confidence levels under different confidence levels.

Instance	Confidence level	Iteration	LB	UB	Optimality Gap
		1	-42,075,758	56,140,273	174.95%
1	50%	2	7,047,322	7,064,629	0.24%
		3	7,064,629	7,064,629	0.00%
		1	-42,075,758	56,135,575	174.95%
2	60%	2	7,071,359	7,179,857	1.51%
		3	7,179,857	7,179,857	0.00%
		1	-42,075,758	56,293,283	174.74%
3	70%	2	7,200,853	7,218,691	0.25%
		3	7,218,691	7,218,691	0.00%
		1	-42,075,758	56,294,637	174.74%
4	80%	2	7,336,441	7,397,419	0.92%
		3	7,397,419	7,397,419	0.00%
		1	-42,075,758	56,301,486	174.73%
5	90%	2	7,898,586	7,972,174	0.92%
		3	7,972,174	7,972,174	0.00%
		1	-42,075,758	56,299,217	174.74%
6	95%	2	7,896,317	7,972,174	0.95%
		3	7,972,174	7,972,174	0.00%

5.4. Comparisons with budget uncertainty set

Budget uncertainty set is widely used in the literature because of its simple structure. However the method is not adaptive to data distribution as a certain budget interval should be set prior to the calculation. Table 5 lists the results of deployment plan under the same weather condition and load demand. Set the range of the box uncertainty set to be 1. In the extreme situation, the budget uncertainty set degenerates to the box uncertainty set. There will be one type III WT installed at bus 4 and one type I WT installed at bus 19. The deployment is different from the previous plan shown in Table 3. However, the total installed capacity is the same as the plan with 80% confidence level.

Table 5: DG deployment under budget uncertainty set.

Uncertainty set type	Location	Category	Capacity (kW)	Number	Total Capacity (kW)
Budget uncertainty set	4	WT	2000	1	2850
	19		850	1	

Table 6 shows the computational results of DDARDGP frame with budget uncertainty set and the proposed data-driven uncertainty set. The result indicates that the data-driven uncertainty set is 34.14% less conservative than the conventional one as the ranges of the budget uncertainty set and the data-driven uncertainty set are 0.6622 and 0.4361, respectively. The second-stage cost with budget uncertainty set is $-439,696 which is $23,185 larger than that with data-driven uncertainty set i.e. $-462,881. Hence, the proposed approach has achieved 0.3% reduction of the total cost than the budget uncertainty. Besides, it is also worth pointing out that the proposed data-driven uncertainty set can avoid constraint violation i.e. voltage violation while the occasions appear substantially as an infeasible problem at the second-stage subproblem solving phase. Although the construction of data-driven uncertainty set takes more time, the computational time for both uncertainty sets is small using the modified CC&G algorithm. The algorithm converges quickly, which means that it can work efficiently with the proposed DDARDGP framework.

Table 6: Comparison of computation results between budget uncertainty set and data-driven uncertainty set.

Uncertainty set type	Budget uncertainty set	Data-driven uncertainty set
Range	0.6622	0.4361
Total cost ($)	7,420,604	7,397,419
First-stage cost ($)	7,860,300	7,860,300
Second-stage cost ($)	-439,696	-462,881
Binary variables	54	54
Continuous variables	115	1035
Uncertain variables	64	576
Constraints	311	2799
CPU time	11.66s	112.56s

5.5. Comparisons under different cost weights

Compared with the first-stage cost, the effects of the second-stage cost are not significant. It is difficult for a system operator to recover the spend at the operation stage once the DG units are installed. To better understand the framework, we will compare the results of DDARDGP by adding a new cost weight factor θ as follows:

$$\underset{\mathbf{f}\in\mathbb{F}}{min}(\theta\mathbf{A}^{\mathsf{T}}\mathbf{f}) + \underset{\mathbf{u}\in\mathbb{U}}{max}\ \underset{\mathbf{y}\in\mathbb{Y}(\mathbf{f},\mathbf{u})}{min}\mathbf{B}^{\mathsf{T}}\mathbf{y} + \mathbf{C}^{\mathsf{T}}\mathbf{u} \tag{60}$$

The cost weight factor in the objective will influence the deployment plan. Table 7 shows the deployment plan with different weight factors ranging from 0.00001 to 1 using the same data-driven uncertainty set. It is observed that the identified worst case is the same. If the planner concerns more on the one-time investment cost, larger θ should be used. Otherwise, more RES-based DG units will be installed. This is because small θ in the objective will put more weights on the fuel cost and emission cost at the second-stage.

Table 7: DG deployment under different cost weight factor.

θ	Location	Category	Capacity (kW)	Number	Total Capacity (kW)
1	4	WT	850	1	2850
	26		2000	1	
0.1	4	WT	850	1	2850
	26		2000	1	
0.01	4	WT	850	1	2850
	26		2000	1	
0.001	4	WT	1100	1	3050
	19		850	1	
	26		1100	1	
0.0001	19	WT	850	1	3050
			1100	2	
0.00001	19	WT	850	1	3050
			1100	2	

6. Conclusions

This papre proposes a two-stage adaptive robust optimization framework DDARDGP for microgrid utility to investigate the DG planning strategy including category, type, number and location decisions. A novel data-driven based uncertainty set construction method has been used to handle the uncertainty of renewable power generations and load demand. As an independent step of the system planning, the construction of uncertainty sets incorporates spatio-temporal information from weather stations and smart meters, making the framework adaptive to new data sets. A modified CC&G algorithm is presented to solve the two-stage optimization problem within a reasonable computational time.

Experiments were conducted using a modified IEEE 33-bus system. Both dispatchable fuel-based microturbines and on-dispatchable RES-based DG units are considered. Through the analysis of the impact of under different truncation threshold values and confidence levels of data-driven uncertainty sets, the system planner could find the tradeoff between the conservatism and robustness of the framework. The effectiveness of the proposed DDARDGP framework is verified by the comparisons with the budget uncertainty set and under different cost weights of the objective. The simulation results show that the data distribution is well captured by a combination of 9 polyhedral sets, resulting 0.3% reduction of the total cost and 34.14% reduction on uncertainty estimation. The problem of over-conservatism is avoided. In addition, the total installed RES capacity will increase 200kW when cost weight changes from 0.01 to 0.001. Therefore, the proposed DDARDGP is suitable for the system planner to accommodate more renewable DG units due to public policy initiatives in practice.

Acknowledgement

The work is partially supported by SP Energy Networks on the project 'A holistic approach for power system monitoring to support DSO transition' and

EPSRC funded project on 'Creating Resilient Sustainable Microgrids through Hybrid Renewable Energy Systems' under grant EP/R030243/1.

References

[1] E. Karunarathne, J. Pasupuleti, J. Ekanayake, D. Almeida, The optimal placement and sizing of distributed generation in an active distribution network with several soft open points, Energies 14 (4). doi:10.3390/en14041084.
URL https://www.mdpi.com/1996-1073/14/4/1084

[2] S. Karki, M. Kulkarni, M. D. Mann, H. Salehfar, Efficiency improvements through combined heat and power for on-site distributed generation technologies, Cogeneration and distributed generation journal 22 (3) (2007) 19–34.

[3] M. A. Hossain, H. R. Pota, W. Issa, M. J. Hossain, Overview of ac microgrid controls with inverter-interfaced generations, Energies 10 (9). doi:10.3390/en10091300.
URL https://www.mdpi.com/1996-1073/10/9/1300

[4] F. Katiraei, M. Iravani, Power management strategies for a microgrid with multiple distributed generation units, IEEE Transactions on Power Systems 21 (4) (2006) 1821–1831. doi:10.1109/TPWRS.2006.879260.

[5] X. Xu, X. Jin, H. Jia, X. Yu, K. Li, Hierarchical management for integrated community energy systems, Applied Energy 160 (2015) 231–243.

[6] X. Tang, B. Fox, K. Li, Reserve from wind power potential in system economic loading, IET Renewable Power Generation 8 (5) (2014) 558–568.

[7] Y. Li, B. Feng, G. Li, J. Qi, D. Zhao, Y. Mu, Optimal distributed generation planning in active distribution networks considering integration of energy storage, Applied energy 210 (2018) 1073–1081.

[8] J. Yan, K. Li, E.-W. Bai, J. Deng, A. M. Foley, Hybrid probabilistic wind power forecasting using temporally local gaussian process, IEEE Transactions on Sustainable Energy 7 (1) (2016) 87–95. doi:10.1109/TSTE.2015.2472963.

[9] M. Zahedi Vahid, Z. M. Ali, E. Seifi Najmi, A. Ahmadi, F. H. Gandoman, S. H. Aleem, Optimal allocation and planning of distributed power generation resources in a smart distribution network using the manta ray foraging optimization algorithm, Energies 14 (16) (2021) 4856.

[10] H. Falaghi, C. Singh, M.-R. Haghifam, M. Ramezani, Dg integrated multistage distribution system expansion planning, International Journal of Electrical Power & Energy Systems 33 (8) (2011) 1489–1497.

[11] Y. Lei, D. Wang, H. Jia, J. Chen, J. Li, Y. Song, J. Li, Multi-objective stochastic expansion planning based on multi-dimensional correlation scenario generation method for regional integrated energy system integrated renewable energy, Applied Energy 276 (2020) 115395.

[12] M. Asensio, P. Meneses de Quevedo, G. Muñoz-Delgado, J. Contreras, Joint distribution network and renewable energy expansion planning considering demand response and energy storage—part i: Stochastic programming model, IEEE Transactions on Smart Grid 9 (2) (2018) 655–666. doi:10.1109/TSG.2016.2560339.

[13] E. Gil, I. Aravena, R. Cárdenas, Generation capacity expansion planning under hydro uncertainty using stochastic mixed integer programming and scenario reduction, IEEE Transactions on Power Systems 30 (4) (2014) 1838–1847.

[14] C. Zhang, Y. Xu, Z. Y. Dong, Probability-weighted robust optimization for distributed generation planning in microgrids, IEEE Transactions on Power Systems 33 (6) (2018) 7042–7051.

[15] Z. Wang, B. Chen, J. Wang, J. Kim, M. M. Begovic, Robust optimization based optimal dg placement in microgrids, IEEE Transactions on Smart Grid 5 (5) (2014) 2173–2182.

[16] B. Jeddi, V. Vahidinasab, P. Ramezanpour, J. Aghaei, M. Shafie-khah, J. P. Catalão, Robust optimization framework for dynamic distributed energy resources planning in distribution networks, International Journal of Electrical Power & Energy Systems 110 (2019) 419–433.

[17] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, T. Zheng, Adaptive robust optimization for the security constrained unit commitment problem, IEEE Transactions on Power Systems 28 (1) (2013) 52–63. doi:10.1109/TPWRS.2012.2205021.

[18] L. Zhao, B. Zeng, Robust unit commitment problem with demand response and wind energy, in: 2012 IEEE power and energy society general meeting, IEEE, 2012, pp. 1–8.

[19] Y. Guan, J. Wang, Uncertainty sets for robust unit commitment, IEEE Transactions on Power Systems 29 (3) (2013) 1439–1440.

[20] C. Shao, X. Wang, M. Shahidehpour, X. Wang, B. Wang, Security-constrained unit commitment with flexible uncertainty set for variable wind power, IEEE Transactions on Sustainable Energy 8 (3) (2017) 1237–1246. doi:10.1109/TSTE.2017.2673120.

[21] Y. Dvorkin, M. Lubin, S. Backhaus, M. Chertkov, Uncertainty sets for wind power generation, IEEE Transactions on Power Systems 31 (4) (2015) 3326–3327.

[22] E. Guevara, F. Babonneau, T. H. de Mello, S. Moret, A machine learning and distributionally robust optimization framework for strategic energy planning under uncertainty, Applied Energy 271 (2020) 115005. doi:https://doi.org/10.1016/j.apenergy.2020.115005.

URL https://www.sciencedirect.com/science/article/pii/S0306261920305171

[23] Q. Niu, H. Zhang, K. Li, G. W. Irwin, An efficient harmony search with new pitch adjustment for dynamic economic dispatch, Energy 65 (2014) 25–43.

[24] Q. Niu, H. Zhang, X. Wang, K. Li, G. W. Irwin, A hybrid harmony search with arithmetic crossover operation for economic dispatch, International journal of electrical power & energy systems 62 (2014) 237–257.

[25] M. E. Baran, F. F. Wu, Optimal capacitor placement on radial distribution systems, IEEE Transactions on Power Delivery 4 (1) (1989) 725–734. doi:10.1109/61.19265.

[26] K. Turitsyn, P. Šulc, S. Backhaus, M. Chertkov, Distributed control of reactive power flow in a radial distribution circuit with high photovoltaic penetration, in: IEEE PES General Meeting, 2010, pp. 1–6. doi:10.1109/PES.2010.5589663.

[27] M. E. Baran, F. F. Wu, Network reconfiguration in distribution systems for loss reduction and load balancing, IEEE Transactions on Power Delivery 4 (2) (1989) 1401–1407. doi:10.1109/61.25627.

[28] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, T. Zheng, Adaptive robust optimization for the security constrained unit commitment problem, IEEE transactions on power systems 28 (1) (2012) 52–63.

[29] A. Lorca, X. A. Sun, Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind, IEEE Transactions on Power Systems 30 (4) (2014) 1702–1713.

[30] T. Ding, S. Liu, W. Yuan, Z. Bie, B. Zeng, A two-stage robust reactive power optimization considering uncertain wind power integration in active distribution networks, IEEE Transactions on Sustainable Energy 7 (1) (2015) 301–311.

[31] Y. Atwa, E. El-Saadany, M. Salama, R. Seethapathy, Optimal renewable resources mix for distribution system energy loss minimization, IEEE Transactions on Power Systems 25 (1) (2009) 360–370.

[32] J. Cervantes, X. Li, W. Yu, K. Li, Support vector machine classification for large data sets via minimum enclosing ball clustering, Neurocomputing 71 (4-6) (2008) 611–619.

[33] D. M. Blei, M. I. Jordan, Variational inference for dirichlet process mixtures, Bayesian analysis 1 (1) (2006) 121–143.

[34] K. P. Murphy, Conjugate bayesian analysis of the gaussian distribution, def 1 ($2\sigma 2$) (2007) 16.

[35] B. Zeng, L. Zhao, Solving two-stage robust optimization problems using a column-and-constraint generation method, Operations Research Letters 41 (5) (2013) 457–461.

[36] M. E. Baran, F. F. Wu, Network reconfiguration in distribution systems for loss reduction and load balancing, IEEE Power Engineering Review 9 (4) (1989) 101–102.

[37] M. J. Brooks, S. Du Clou, W. L. Van Niekerk, P. Gauché, C. Leonard, M. J. Mouzouris, R. Meyer, N. Van der Westhuizen, E. E. Van Dyk, F. J. Vorster, Sauran: A new resource for solar radiometric data in southern africa, Journal of energy in Southern Africa 26 (1) (2015) 2–10.