Boundary Inference of Load Scenarios in Multi-energy Parks Based on Statistical Learning

Jiajia Huan
Power Grid Planning Research Center
China Southern Power Grid
Guangdong Power Grid Co. LTD
Guangzhou, China
winnie5983@126.com

Zhimeng Wang
Shenzhen International Graduate School
Tsinghua University
Shenzhen, China
zmwang0574@163.com

Yunfei Du
Shenzhen International Graduate School
Tsinghua University
Shenzhen, China
du.yunfei@sz.tsinghua.edu.cn

Xinwei Shen*
Shenzhen International Graduate School
Tsinghua University
Shenzhen, China
sxw.tbsi@sz.tsinghua.edu.cn
*Corresponding author

Baihao Qiao
School of Electronic and Information
Zhongyuan University of Technology
Zhengzhou, China
bhqiao@zut.edu.cn

Chungeng He
Power Grid Planning Research Center
China Southern Power Grid
Guangdong Power Grid Co. LTD
Guangzhou, China
980279904@qq.com

Xiaodong Lan
Power Grid Planning Research Center
China Southern Power Grid
Guangdong Power Grid Co. LTD
Guangzhou, China
Paullan112@163.com

Shuxin Luo
Power Grid Planning Research Center
China Southern Power Grid
Guangdong Power Grid Co. LTD
Guangzhou, China
luoshuxin2006@126.com

***Abstract*—Uncertainties of loads including but not limited to electricity, cold and heat have brought difficulties to safe and economical operation of multi-energy complementary parks. To deal with the uncertainties, system operators usually choose to set a fluctuation range for the loads based on their past experiences of operation of the system as well as their personal preferences and habits, making the fluctuation ranges, also known as uncertainty sets in optimization, less reliable and less explainable. In this paper, we proposed a method of determining uncertainty sets in multi-energy complementary parks based on statistical learning. By applying this method, the uncertainty sets can be derived more flexibly solely based on how much the safe operation of the system could be affected and the predicted values of the loads, with no need of past experiences of operation. Such advantages make the proposed method easier to be popularized and utilized in various areas and scenarios.**

Index Terms*—Statistical learning, multi-energy complementary park, operation optimization, uncertainty set, load demand scenarios

I. Introduction

Traditional fossil energy has the characteristics of non-renewable, and it may sometimes be pollutional when not being used properly. Nowadays, despite the depletion of fossil energy and the serious environmental pollution, the traditional energy supply mode is facing great challenges, and the comprehensive utilization efficiency of energy needs to be improved urgently. It is a method to solve the contradiction between energy demand growth, energy shortage and environmental protection pressure. Against such background, the concept of multi-energy complementary park came into being [1]–[6], which realizes the interconnection of various energy systems including electricity/gas/heat/cold, etc., and improves the comprehensive utilization efficiency of energy. However, various unexpected situations may occur during the operation of the multi-energy complementary park, resulting in various uncertainties. These uncertainties make the optimal operation strategy of the multi-energy complementary park face many difficulties. Modeling of uncertainties in multi-energy complementary park is an urgent problem to be solved so as to handle the difficulties of operating the system.

Methods using statistical learning to model uncertainties in multi-energy complementary park has been widely studied. Statistical learning is a machine learning framework, first proposed by Vladimir Vapnik [7]–[10], to get information from the perspectives of statistics. Classical and traditional statistical learning models include perceptron [11], K-nearest-neighbors [12], [13], naive Bayes model [14], decision tree [15], logistic regression [16], maximum entropy model [17], and support vector machine [18], [19].

Statistical learning has been widely used in multi-energy complementary park planning and uncertain description, where

This work is supported by Science and Technology Project of China Southern Power Grid Corporation (Energy Efficiency Data Mining Technology for Multi-energy Park Planning, No. 037700KK52200003).

Conditional Value-at-Risk (CVaR) is one of the most popular tools and has been utilized in various existing researches. In [20], a virtual power plant source-load-storage multiple reserve capacity system is constructed based on the characteristics of the constituent units of the virtual power plant, using the conditional value at risk to measure the risk loss brought by the uncertainty of wind power output to the virtual power plant, establishing a CVaR based on the cost-benefit analysis. The coordinated optimization model of power generation capacity and reserve capacity of multi-type distributed resources. In [21], uncertain factors in the objective functions and the constraints in virtual power plants are handled using the theory of CVaR and robust optimization theory, proposing a risk avoidance scheduling model. In [22], by introducing the theory of CVaR into integrated energy system (IES) operation problems, an economic dispatch model is proposed considering uncertainties of wind generation, solar generation as well as power demand and heat demand. To model uncertainties brought by production of renewable energy sources (RESs) and load variation, CVaR is introduced into park-level integrated energy system (PIES) in [23], and planning decisions with anti-risk ability can be derived with the proposed model.

Based on these researches, a boundary inference method for load demand, i.e., a method to determine the boundaries of load demand scenarios, is proposed. Different from most of the existing methods which rely on system operators' previous experiences to determine the boundaries, the proposed method relies on information of predicted load demand, making the method easier to be popularized and utilized in different areas and scenarios.

The paper is organized as follows. Section II formulates the problem and emphasize the goal of the proposed method. Section III elaborates on the proposed method by providing detailed descriptions of each step. Section IV gives a simple numerical example and integrates the results of the example into a real integrated energy system to demonstrate the effects of the proposed method. Section V concludes the paper.

II. Problem Formulation

Consider the mathematical optimization model of a multi-energy complementary park:

$$\min_{x_i, Q_b, P_g} \left(C^{inv} + C^{ope}\right) \tag{1}$$

$$\text{s.t. } C^{inv} = \sum_{i \in D} cc_i P_{i,g,\max} x_i \tag{2}$$

$$C^{ope} = \sum_{i=1}^{D} \sum_{t=1}^{K} c_i P_{i,g,t} \tag{3}$$

$$\sum_{i=1}^{D} P_{i,g,t} + \frac{Q_{b,t}}{\Delta t} + P_{\text{shed},t} = p_{0,t} \tag{4}$$

$$0 \le P_{i,g,t} \le p_{0,t} \quad i = 1, \cdots, D \tag{5}$$

$$0 \le \frac{Q_{b,t}}{\Delta t} \le p_{0,t} \tag{6}$$

$$0 \le P_{\text{shed},t} \le p_{0,t} \tag{7}$$

$$0 \le |P_{i,g,t+1} - P_{i,g,t}| \le \Delta P_{i,\max} \tag{8}$$

$$0 \le P_{i,g,t} \le P_{i,g,\max} \quad i = 1, \cdots, D \tag{9}$$

where C^{inv} and C^{ope} stand for the investment costs and operation costs of the multi-energy park, respectively, and are respectively formulated in (2) and (3). $x_i, \forall i \in D$ are the binary decision variables indicating whether the ith device should be installed, and D stands for both the total number of candidate devices as well as the set of the candidate devices. cc_i and c_i stand for the investment cost coefficient and the operation cost coefficient of the ith device in the system, respectively. $p_{0,t}$ stands for the total power consumption (MW) of all the devices at time t. Δt is the time change and is assumed as 1 hour in this study. t can be taken from $1, ..., K$, where K is the total number of data points. Q_b and P_g stand for the vector of electricity bought from the upper grid and the power generated by the devices in the multi-energy complementary park, respectively, i.e.,

$$Q_b = \begin{bmatrix} Q_{b,1} & Q_{b,2} & \dots & Q_{b,K} \end{bmatrix}^T \tag{10}$$

$$\begin{aligned} P_{1,g} &= \begin{bmatrix} P_{1,g,1} & P_{1,g,2} & \dots & P_{1,g,K} \end{bmatrix}^T \\ P_{2,g} &= \begin{bmatrix} P_{2,g,1} & P_{2,g,2} & \dots & P_{2,g,K} \end{bmatrix}^T \\ &\vdots \\ P_{D,g} &= \begin{bmatrix} P_{D,g,1} & P_{D,g,2} & \dots & P_{D,g,K} \end{bmatrix}^T \end{aligned} \tag{11}$$

$P_{i,g,t}$ is the generation (MW) of the ith device at time t, with $P_{i,g,\max}$ defined as the capacity (MW) of the corresponding device. $\Delta P_{i,\max}$ is the maximal ramping of the generation of device i, and the load shedding (MW) at time t is represented by $P_{shed,t}$. Equation (1) is the objective function of the optimization problem under the case when no uncertainty is considered, which is the total cost including investment cost C^{inv} and operation cost C^{ope}, and the goal of optimization is to minimize the costs by adjusting Q_b and P_g. Equations (4)-(9) are the constraints of the optimization problem. Equation (4), which is also known as power balance constraint, is an equality constraint requiring the sum of power bought from the upper level power grid, power generated by devices in the multi-energy complementary park, load shedding at time t to be equal to the predicted load demand at time t. As a result, all the three components on the left hand side of (4) should be non-negative numbers no greater than $p_{0,t}$, the right hand side of the constraint, leading to (5)-(7). Equation (8) is a ramping constraint to make sure the the deviation of the generation of the devices will not exceed the maximal allowable deviation, $\Delta P_{i,\max}$. Equation (9) is the generation constraint, making sure that the generation of all the devices at any time are non-negative numbers no greater than the capacity of the corresponding generator.

However, during the operation of the multi-energy complementary park, various unexpected situations and other uncertainties may occur and chances are high that loads of the devices cannot be predicted precisely. For example, the

sudden extremely high temperature weather might increase the power consumption of air conditioners, thereby increasing the electrical load, that is, the predicted data of the electrical load will be biased. These uncertainties make the optimal operation strategy of the multi-energy complementary park face many difficulties.

Fig. 1 depicts the case when there exists $[-10\%, +20\%]$ uncertainty in electricity load demand, where the red solid curve stands for the predicted values, and the two blue dotted curves are the lower bound and the upper bound of the electricity load demand, and the interval between is the range in which the actual load demand could fluctuate in. Hence to depict the uncertainties in the multi-energy complementary park, some uncertain variables should be taken into consideration. $\mathbf{u}_p \in \mathbb{R}^K$ is defined as the uncertain fluctuation vector of electricity, and $u_{p,t}$ is the tth term in $\mathbf{u}_p$ representing the uncertain fluctuation coefficient of load demand at time t, i.e., $\mathbf{u}_p = \begin{bmatrix} u_{p,1} & u_{p,2} & \dots & u_{p,K} \end{bmatrix}^T$. Elements in $\mathbf{u}_p$ are all randomly taken from a normal distribution with a mean of 1. Affected by the uncertain fluctuation coefficients, the electricity load demand at time t changes from the input predicted value $p_{0,t}$ to $p_t = p_{0,t}u_{p,t}$. Correspondingly, the input predicted vector of load demand $\mathbf{p}_0$ changes to the actual load demand $\mathbf{p} = \begin{bmatrix} p_1 & p_2 & \dots & p_K \end{bmatrix}^T$.

$$\min_{x_i, Q_b, P_g} \max_{\mathbf{u}_p} \left(C^{inv} + C^{ope} \right) \tag{12}$$

$$\text{s.t.} \sum_{i=1}^{D} P_{i,g,t} + \frac{Q_{b,t}}{\Delta t} + P_{\text{shed},t} = p_{0,t} \cdot u_{p,t} \tag{13}$$

$$0 \le \sum_{i=1}^{D} P_{i,g,t} \le p_{0,t} \cdot u_{p,t} \quad i = 1, \cdots, D \tag{14}$$

$$0 \le \frac{Q_{b,t}}{\Delta t} \le p_{0,t} \cdot u_{p,t} \tag{15}$$

$$0 \le P_{\text{shed},t} \le p_{0,t} \cdot u_{p,t} \tag{16}$$

$$0 \le |P_{i,g,t+1} - P_{i,g,t}| \le \Delta P_{i,\max} \tag{17}$$

$$0 \le P_{i,g,t} \le P_{i,g,\max} \quad i = 1, \cdots, D \tag{18}$$

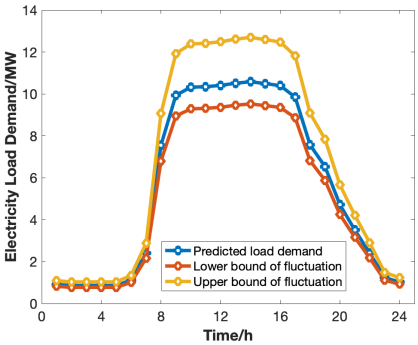

Fig. 1. Curves of predicted electricity load demand in 24 hours and the corresponding upper bound and lower bound when there exists $[-10\%, +20\%]$ uncertainty.

Taking uncertainty of electricity load demand into consideration, a robust model (12)-(18) can be formulated based on the deterministic one (1)-(9). The physical meaning of the constraints of the robust model is similar to the deterministic constraints, hence will not be elaborated here. Note that the objective function (12) of the new model is no longer a single-layer $\min$ problem but a bi-layer $\min\max$ problem, that is because the uncertain fluctuation coefficients would maximize the value of the objective function, while the other decision variables would still minimize the objective function as in (1). The model can be solved using some algorithms specifically designed for bilevel optimization, which is out of the scope of this paper and hence is not discussed here.

The three constraints in the first mathematical model (5)-(7) are both inequality constraints concerning the prediction of load demand, and can be written be into the following compact form:

$$g(\mathbf{p}_0) \le 0 \tag{19}$$

In the new mathematical model, the three corresponding constraints (14)-(16) can be reformulated into the following form:

$$g(\mathbf{p}_0, \mathbf{u}_p) \le 0 \tag{20}$$

In the problem of boundary inference, $\mathcal{U}_{p,\epsilon}$, the set of $\mathbf{u}_p$, should be found, and for $\forall \mathbf{u}_p \in \mathcal{U}_{p,\epsilon}$, there is

$$\mathbb{P}\left(g\left(\mathbf{p}_0, \mathbf{u}_p\right) \le 0\right) \ge 1 - \epsilon \tag{21}$$

In (21), $\mathbb{P}(\bullet)$ is the probability of the event, and $(1-\epsilon)$ is the confidence level of the set $\mathcal{U}_{p,\epsilon}$, $0 < \epsilon < 1$. ϵ is usually chosen as a relatively small number, with typical values 0.01, 0.05, etc, so as to make sure that $(1-\epsilon)$ is a value close to 1, i.e., the constraint will be satisfied with higher probability. Concretely speaking, (21) means that the constraint $g(\bullet) \le 0$ would still be satisfied with a probability no less than $(1-\epsilon)$ even when considering uncertain fluctuation vector $\mathbf{u}_p$.

To summarize, the goal of the problem of boundary inference is to find $\mathcal{U}_{p,\epsilon}$, a set of $\mathbf{u}_p$ satisfying (21). Different $\mathcal{U}_{p,\epsilon}$ should be derived based on different value of ϵ, so that the operator of the multi-energy complementary park could determine the value of ϵ based on how much the safety of the operation of the multi-energy park is affected and get different $\mathcal{U}_{p,\epsilon}$.

III. Proposed Method

To make the results of operation problems more reliable and more safe, many models and methods take uncertainty into consideration, while most of which set the scope of uncertainty set based on previous operation experiences and hardly ever construct uncertainty sets from the prospective of statistical learning [24]. The uncertainty sets determined based on previous experiences are not explainable, cannot guarantee optimality and cannot be popularized. To guarantee safety of operation more strictly, some operators may assume the loads fluctuating in a bigger range based on their previous experiences, while such a bigger fluctuation range may make the operation strategy to be too conservative. On the contrary,

some operators may assume a smaller fluctuation range, which may make the operation strategy too radical. These two scenarios would both result in the uncertainty sets becoming less effective, and may even affect the economy and safety of operation of the multi-energy parks. The proposed method can give an explainable boundary of uncertainty set based on the combination of the given predicted load demand and different degrees of influence on the safe operation of multi-energy parks, so as to make sure that the load demand following the uncertainty set would still satisfy the security constraints of the multi-energy complementary park. The steps of the method will be introduced as follows.

A. Step 1

The first step is to get the predicted load demand curve as input data. Electricity load demand is used as an example here:

$$\mathbf{p}_0 = \begin{bmatrix} p_{0,1} & p_{0,2} & \cdots & p_{0,K} \end{bmatrix}^T \tag{22}$$

K data points are input here as K continuous time points are considered at the same time. K is usually set as 24 to count for the fact that there are 24 hours in each day.

B. Step 2

Step 2 is to generate the alternative coefficients and choose among the coefficients based on calculation.

The alternative coefficients are generated by choosing K numbers randomly from a normal distribution with a mean of 1. These K numbers are denoted by $u^{(1)}, u^{(2)}, \cdots, u^{(K)}$. Then sort these numbers from the smallest to the largest, we get $\hat{u}^{(1)}, \hat{u}^{(2)}, \cdots, \hat{u}^{(K)}$. Among the sorted numbers, the $(K-s+1)$th and the sth, i.e., $\hat{u}^{(K-s+1)}$ and $\hat{u}^{(s)}$, would be chosen as the coefficients of the lower bound and the upper bound, respectively. s can be calculated based on the following optimization problem [25], [26]:

$$\min_{s\in\mathbb{N}} s \tag{23}$$

$$\text{s.t.} \sum_{j=k}^{N} \binom{N}{j} (\epsilon/d)^{N-j}(1-\epsilon/d)^{j} \le \frac{\alpha}{2d} \tag{24}$$

$$1 \le s \le K \tag{25}$$

$$0 \le \sum_{i=1}^{D} P_{i,g,t} \le p_{0,t} \cdot \hat{u}^{(K-s+1)} \tag{26}$$

$$0 \le \sum_{i=1}^{D} P_{i,g,t} \le p_{0,t} \cdot \hat{u}^{(s)} \tag{27}$$

$$0 \le \frac{Q_{b,t}}{\Delta t} \le p_{0,t} \cdot \hat{u}^{(K-s+1)} \tag{28}$$

$$0 \le \frac{Q_{b,t}}{\Delta t} \le p_{0,t} \cdot \hat{u}^{(s)} \tag{29}$$

$$0 \le P_{shed,t} \le p_{0,t} \cdot \hat{u}^{(K-s+1)} \tag{30}$$

$$0 \le P_{shed,t} \le p_{0,t} \cdot \hat{u}^{(s)} \tag{31}$$

where $\mathbb{N}$ denotes the set of natural numbers, ϵ denotes the risk level of the uncertainty set, and a greater ϵ leads to a larger uncertainty set, and higher risk of the operation of the multi-energy complementary park, while a smaller ϵ leads to a uncertainty set with smaller range and lower risk of the operation of the multi-energy complementary park.

Note that (26)-(31) are included as constraints to guarantee that the constraints of the multi-energy complementary park are still satisfied, where (26)-(27) are derived from (14), (28)-(29) are derived from (15), and (30)-(31) are derived from (16).

C. Step 3

The upper and lower boundaries of the uncertain fluctuation vector are obtained based on the coefficients corresponding to the upper and lower boundaries of the uncertain fluctuation vector in step 3. According to s calculated in step 2, $\hat{u}^{(K-s+1)}$ and $\hat{u}^{(s)}$ are chosen as the lower bound and upper bound of the uncertain fluctuation vector, i.e., the following uncertainty set $\mathcal{U}$ guarantees that more than $(1-\alpha)$ of the samples can satisfy the security constraints $g(\mathbf{p}, \mathbf{u})$ of the multi-energy complementary park at the level of ϵ:

$$\mathcal{U}_\epsilon = \left\{ \mathbf{u} \in \mathbb{R}^K : \hat{u}^{(K-s+1)} \le u_i \le \hat{u}^{(s)} \quad i = 1, \ldots, K \right\} \tag{32}$$

IV. Numerical Examples

In the numerical example, we assume all the data in the multi-energy complementary park are recorded each hour, hence $K = 24$ points in a single day. The aforementioned method is illustrated with a toy example, and results of the example is integrated in a multi-energy complementary park as defined in (12)-(18) to show the effect of the uncertainty set derived from the proposed method.

An array of electricity load demand data $\mathbf{p}_0$ (MW) is provided in TABLE I.

Following step 2, $K = 24$ uncertain fluctuation coefficients should be generated following a normal distribution with a mean of 1. The generated coefficients are listed in TABLE II.

By reordering the generated coefficients $u^{(1)}$, $u^{(2)}$, $\cdots$, $u^{(24)}$ from the smallest to the largest, the sorted list of uncertain fluctuation coefficients are listed in TABLE III.

Solving optimization problem defined by (23)-(31), the optimal s can be calculated as $s^* = 23$, hence the lower bound and upper bound of the uncertain fluctuation vector should be chosen as

$$\hat{u}^{(K-s^*+1)} = \hat{u}^{(24-23+1)} = \hat{u}^{(2)} = 0.261 \tag{33}$$

$$\hat{u}^{(s^*)} = \hat{u}^{(23)} = 1.882 \tag{34}$$

respectively. Applying the aforementioned bounds, the uncertainty set is shown in Fig. 2.

To show the effect of the derived uncertainty set, it is used in a specific multi-energy complementary park. The wind speed and solar radiation prediction data within 24 hours are shown in TABLE IV and V respectively, and parameters of the candidate devices are shown in TABLEs VI, VII and VIII.

TABLE I
PREDICTED ELECTRICITY LOAD DEMAND

Time	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
Load Demand (MW)	0.911	0.863	0.863	0.863	0.863	1.123	2.403	7.560	9.940	10.319	10.348	10.406
Time	$t=13$	$t=14$	$t=15$	$t=16$	$t=17$	$t=18$	$t=19$	$t=20$	$t=21$	$t=22$	$t=23$	$t=24$
Load Demand (MW)	10.512	10.581	10.493	10.396	9.844	7.567	6.534	4.717	3.508	2.415	1.241	1.028

TABLE II
THE GENERATED UNCERTAIN FLUCTUATION COEFFICIENTS

Number	$u^{(1)}$	$u^{(2)}$	$u^{(3)}$	$u^{(4)}$	$u^{(5)}$	$u^{(6)}$	$u^{(7)}$	$u^{(8)}$	$u^{(9)}$	$u^{(10)}$	$u^{(11)}$	$u^{(12)}$
Coefficients	1.317	1.162	1.128	0.883	0.352	0.962	0.801	1.882	0.944	0.159	0.845	0.585
Number	$u^{(13)}$	$u^{(14)}$	$u^{(15)}$	$u^{(16)}$	$u^{(17)}$	$u^{(18)}$	$u^{(19)}$	$u^{(20)}$	$u^{(21)}$	$u^{(22)}$	$u^{(23)}$	$u^{(24)}$
Coefficients	0.341	1.984	1.101	0.637	0.261	1.153	0.683	1.319	1.487	0.412	1.302	1.219

TABLE III
THE SORTED UNCERTAIN FLUCTUATION COEFFICIENTS

Number	$\hat{u}^{(1)}$	$\hat{u}^{(2)}$	$\hat{u}^{(3)}$	$\hat{u}^{(4)}$	$\hat{u}^{(5)}$	$\hat{u}^{(6)}$	$\hat{u}^{(7)}$	$\hat{u}^{(8)}$	$\hat{u}^{(9)}$	$\hat{u}^{(10)}$	$\hat{u}^{(11)}$	$\hat{u}^{(12)}$
Coefficients	0.159	0.261	0.341	0.352	0.412	0.585	0.637	0.683	0.801	0.845	0.883	0.944
Number	$\hat{u}^{(13)}$	$\hat{u}^{(14)}$	$\hat{u}^{(15)}$	$\hat{u}^{(16)}$	$\hat{u}^{(17)}$	$\hat{u}^{(18)}$	$\hat{u}^{(19)}$	$\hat{u}^{(20)}$	$\hat{u}^{(21)}$	$\hat{u}^{(22)}$	$\hat{u}^{(23)}$	$\hat{u}^{(24)}$
Coefficients	0.962	1.101	1.128	1.153	1.162	1.219	1.302	1.317	1.319	1.487	1.882	1.984

TABLE IV
PREDICTED WIND SPEED

Time	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
Wind Speed (m/s)	13.220	12.442	12.860	12.753	12.393	12.295	11.850	12.338	12.781	14.612	14.706	14.075
Time	$t=13$	$t=14$	$t=15$	$t=16$	$t=17$	$t=18$	$t=19$	$t=20$	$t=21$	$t=22$	$t=23$	$t=24$
Wind Speed (m/s)	13.431	12.710	12.460	11.771	10.785	8.978	8.420	7.342	7.556	8.041	6.933	6.408

TABLE V
PREDICTED SOLAR RADIATION

Time	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
Solar Radiation(W/m^2)	0	0	0	0	0	0	0	0	85.21	264.09	396.74	496.20
Time	$t=13$	$t=14$	$t=15$	$t=16$	$t=17$	$t=18$	$t=19$	$t=20$	$t=21$	$t=22$	$t=23$	$t=24$
Solar Radiation(W/m^2)	546.62	532.27	459.65	331.75	159.72	13.62	0	0	0	0	0	0

TABLE VI
PARAMETERS OF THE DISPATCHABLE UNITS

Unit No.	Capacity (MW)	Levelized Operation Cost ($/MWh)	Initial Investment Cost ($/MW)
1	6	120	50,000
2	5	110	60,000
3	4	100	70,000
4	3	90	80,000

TABLE VIII
PARAMETERS OF THE ENERGY STORAGE SYSTEMS

Unit No.	Rated Power (MW)	Rated Energy (MWh)	Initial Investment Cost - Power ($/MW)	Initial Investment Cost - Energy ($/MWh)
1	1	6	60,000	30,000
2	2	6	30,000	30,000

Solving the robust optimization model defined by (12)-(18), investment decisions and investment costs can be derived based on different uncertainty sets as in TABLE IX, where the results derived in the paper are highlighted with bold face. The uncertainty set derived by the method proposed in this paper achieves a balance between conservativeness and aggressiveness, showing the effectiveness of the method.

TABLE VII
PARAMETERS OF THE NONDISPATCHABLE UNITS

Unit No.	Capacity (MW)	Levelized Operation Cost ($/MWh)	Initial Investment Cost ($/MW)	Type of Energy
1	4	-	150,000	WT
2	2	-	90,000	PV

V. CONCLUSION

In this paper, we proposed a method of determining uncertainty sets. By applying this method, the uncertainty sets can be derived more flexibly since system operators can derive different uncertainty sets based on how much the safe operation of the system could be affected. Our method provides system

TABLE IX
PLANNING RESULTS AND TOTAL CONSTRUCTION COSTS GIVEN DIFFERENT UNCERTAINTY SETS

Percentage of Lower Bounds	Percentage of Upper Bounds	Investment Decisions			Costs (million $)
		Dispatchable Units (MW)	Nondispatchable Units (MW)	Energy Storage Systems (MW)	
0.5	1.5	15	2	0	1.0815
0.261	**1.882**	**18**	**4**	**0**	**1.7420**
0.1	1.9	18	4	0	1.7423

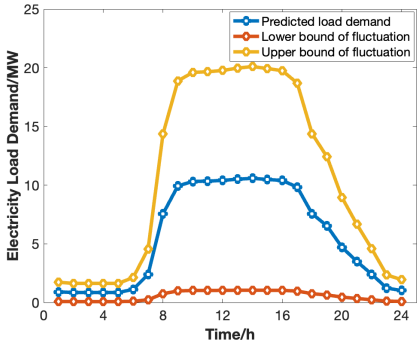

Fig. 2. Uncertainty Set

operators with a more reliable way to determine the uncertainty sets since no previous experience from system operators is needed, making the method easier to be popularized and utilized in various areas and scenarios. In the future, further comparison of this method and the other existing methods is necessary to further prove the significance and viability of the proposed method.

REFERENCES

[1] P. Li, J. Wang, C. Li, Z. Wang, Y. Yin, Z. Han, Y. Pan, and M. Wen, "Collaborative optimal scheduling of the community integrated energy system considering source-load uncertainty ad equipment off-design performance," *Proceeding of the CSEE*, pp. 1–14, 2023.

[2] W. Huang and W. Liu, "Multi-energy complementary based coordinated optimal planning of park integrated energy station-network," *Automation of Electric Power Systems.*, vol. 44, pp. 20–8, 2020.

[3] D. Wang, C. Wang, Y. Lei, Z. Zhang, and N. Zhang, "Prospects for key technologies of new-type urban integrated energy system," *Global Energy Interconnection*, vol. 2, no. 5, pp. 402–412, 2019.

[4] D. Wang, Z. Meng, H. Jia, J. Li, and J. Yu, "Research and prospect of key technologies for energy interconnection system planning for new-type towns," *Automation of Electric Power Systems*, vol. 43, no. 14, pp. 16–28, 2019.

[5] Z. Yuan, Y. Zhao, Z. Guo *et al.*, "Research summary of integrated energy systems planning for energy internet," *Southern Power System Technology*, vol. 13, no. 7, pp. 1–9, 2019.

[6] H. Cheng, X. Hu, L. Wang, Y. Liu, and Q. Yu, "Review on research of regional integrated energy system planning," *Automation of Electric Power Systems*, vol. 43, no. 7, pp. 2–13, 2019.

[7] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.

[8] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[9] R. Frost, B. C. Armstrong, and M. H. Christiansen, "Statistical learning research: A critical review and possible new directions." *Psychological Bulletin*, vol. 145, no. 12, p. 1128, 2019.

[10] Y. Tao, H. Chen, X. Qin, and Z. Meng, "A review of the short-term wind power forecasting theory, model and approach," *Electric Power Engineering Technology*, vol. 37, no. 5, pp. 7–13, 2018.

[11] D. Kumar, "Power system restoration using multilayer perceptron," *International Journal of Engineering, Science and Information Technology*, vol. 1, no. 1, pp. 10–14, 2021.

[12] M. P. Corso, F. L. Perez, S. F. Stefenon, K.-C. Yow, R. García Ovejero, and V. R. Q. Leithardt, "Classification of contaminated insulators using k-nearest neighbors based on computer vision," *Computers*, vol. 10, no. 9, p. 112, 2021.

[13] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Smart non-intrusive appliance identification using a novel local power histogramming descriptor with an improved k-nearest neighbors classifier," *Sustainable Cities and Society*, vol. 67, p. 102764, 2021.

[14] E. Aker, M. L. Othman, V. Veerasamy, I. b. Aris, N. I. A. Wahab, and H. Hizam, "Fault detection and classification of shunt compensated transmission line using discrete wavelet transform and naive bayes classifier," *Energies*, vol. 13, no. 1, p. 243, 2020.

[15] L. Vanfretti and V. N. Arava, "Decision tree-based classification of multiple operating conditions for power system voltage stability assessment," *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106251, 2020.

[16] M. Rajabdorri, E. Lobato, and L. Sigrist, "Robust frequency constrained uc using data driven logistic regression for island power systems," *IET Generation, Transmission & Distribution*, vol. 16, no. 24, pp. 5069–5083, 2022.

[17] B. Sui, K. Hou, H. Jia, Y. Mu, and X. Yu, "Maximum entropy based probabilistic load flow calculation for power system integrated with wind power generation," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 1042–1054, 2018.

[18] Z. Tan, G. De, M. Li, H. Lin, S. Yang, L. Huang, and Q. Tan, "Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine," *Journal of cleaner production*, vol. 248, p. 119252, 2020.

[19] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, and S. Islam, "Classifying transformer winding deformation fault types and degrees using fra based on support vector machine," *IEEE Access*, vol. 7, pp. 112 494–112 504, 2019.

[20] L. Mengxuan, L. Suhua, L. Jianqin *et al.*, "Coordinated optimization of multi-type reserve in virtual power plant accommodated high shares of wind power [j]," *Proceedings of the CSEE*, vol. 10, pp. 2874–2882, 2018.

[21] L. Ju, P. Li, Z. Tan, and W. Wang, "A dynamic risk aversion model for virtual energy plant considering uncertainties and demand response," *International Journal of Energy Research*, vol. 43, no. 3, pp. 1272–1293, 2019.

[22] N. Growe-Kuska, H. Heitsch, and W. Romisch, "Scenario reduction and scenario tree construction for power management problems," in *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 3. IEEE, 2003, pp. 7–pp.

[23] A. Xuan, X. Shen, Q. Guo, and H. Sun, "A conditional value-at-risk based planning model for integrated energy system with energy storage and renewables," *Applied Energy*, vol. 294, p. 116971, 2021.

[24] D. Yu, Y. Guo, J. Wu, J. Li, and C. Wang, "Flexibility improvement planning and evaluation of regional integrated energy system considering uncertainty," *Distribution & Utilization*, vol. 39, no. 4, pp. 84–92, 2022.

[25] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Mathematical Programming*, vol. 167, pp. 235–292, 2018.

[26] P. Embrechts, A. Höing, and A. Juri, "Using copulae to bound the value-at-risk for functions of dependent risks," *Finance and Stochastics*, vol. 7, pp. 145–167, 2003.