Stochastic Unit Commitment in Various System Sizes under High Uncertainty Photovoltaic Forecast

Muhammad Yasirroni[1], Lesnanto Multa Putranto[2], Sarjiya[3], Husni Rois Ali[4], Indra Triwibowo[5], Qiangqiang Xie[6]

[1,2,3,4,5] Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Jl. Grafika No. 2 Kampus UGM Yogyakarta, 55281 INDONESIA (tel.: 0274-552305; fax: 0274-552305; email: [1]muhammad.yasirroni@mail.ugm.ac.id, [2]lesnanto@ugm.ac.id, [3]sarjiya@ugm.ac.id, [4]husni.rois.ali@ugm.ac.id, [5]indra.triwibowo@mail.ugm.ac.id)
[6] College of Electronics and Information, Hangzhou Dianzi University, 115 Wenyi Rd, 310018 CHINA (tel.: +86 (571) 8691 50 76; fax: +86 (571) 8691 51 83; email: xqq@hdu.edu.cn)

[Received: 24 August 2022, Revised: 20 October 2022]
Corresponding Author: Lesnanto Multa Putranto

ABSTRACT — This paper proposes a stochastic unit commitment (SUC) approach to solve a day-ahead unit commitment (UC) problem in a system with high uncertainty net load which is caused by photovoltaic (PV) power plants. In contrast with robust unit commitment (RUC) which only considers the worst-case scenario, SUC considers every possible scenario with its probability. Multiple possible PV curves were obtained using k-means clustering on historical data. The proportion of cluster members was used as a weight factor representing the occurrence probability of PV curves. The test was separated into two-step tests, namely day-ahead and real-time markets, using IEEE 10 generating unit system and solved using CPLEX. The results showed that in a day-ahead UC, SUC ($539,896) had lower cost than RUC ($548,005). However, when the total energy generated was considered, the SUC (20.78 $/MWh) cost higher compared to RUC (20.75 $/MWh). It is because the solution proposed by SUC is as robust as the RUC, but the generation cost formulation also considers over-commitment. Thus, SUC produced a fairer price for the independent power producer and electric utility in the day-ahead calculation. The results also showed that in the test environment of the real-time market, SUC was able to produce a robust solution without going into over-commitment. It is clearly shown in a 30 units system test with 10 centroids, in which SUC had a cheaper solution (20.7253 $/MWh) compared to RUC (20.7285 $/MWh), without violating power balance or going to load shedding.

KEYWORDS — Intermittency, K-Means, Mixed-Integer Linear Programming, Stochastic Unit Commitment.

I. INTRODUCTION

The regulation of minimum renewable energy (RE) power share is getting higher each year. This regulation leads to a rise in solar photovoltaic (PV) installment as one of the most cost-efficient RE power plants. At the same time, the levelized cost of electricity (LCOE) of PV is decreasing every year, which further incentives independent power producers (IPPs) to choose the stochastic power source [1]. Whereas a high PV power share will be followed by a high uncertainty net load curve. This high uncertainty increases the difficulty of power system utility to schedule generating units while minimizing the total cost, resulting in an increase in operational costs.

The generating unit scheduling optimization problem is dubbed the unit commitment (UC). It is usually presented as a day-ahead deterministic unit commitment problem (DUC). It can be described as an optimization problem to determine the on/off schedule of generating units, including the power dispatched by each generating unit to meet the forecasts of the load demand profile of the next day, subjecting to various constraints [2], [3]. In considering the power produced by PV, the electrical power system utility can treat it as a negative load due to its characteristic as noncontrollable, noncontinuously available, and fluctuant power sources, simply called intermittent generators [4], [5]. Thus, the forecast of the load demand profile can be substituted with the forecast of a net load profile for the next day. One approach to solving UC in a system with high uncertainty net load caused by PV power plants is to implement the stochastic unit commitment (SUC). In contrast with robust unit commitment (RUC) [6] which only considers the worst-case scenario, the SUC considers every possible scenario with its probability. In systems with significant PV power share and high uncertainty, SUC is better suited to model the scheduling optimization problem. It is because RUC neglects all forecasts except the worst scenario forecast.

Various SUC formulations have been done in previous studies [7]–[10]. In [7], SUC was solved using an exchange market algorithm and fuzzy satisfying method by considering the uncertainty of forecast to schedule electricity and heat dispatch. This method could handle various scenarios but did not incorporate scenario-independent variables such as a unit on/off status. In [8], the stochastic solution included a unit of on/off status, but it was treated as a scenario-dependent variable where each scenario had an independent on/off status. In [9], a two-model approach was used for solving multiple possible PV curves in dividing the problem. In the first model, the whole grid on/off status was solved first and unchanged, while the loop inside the second stage further optimized the scenario-dependent variable. Further splitting has been done in [10] by incorporating multiple time resolutions where the nearest interval was binding, while the other was an advisory that will be recomputed in the future. In respect to the current solution, this approach has solved both scenario-dependent and scenario-independent variables at a time. However, there is an information gap for the electrical power system utility to determine whether the SUC model is required for their systems. Furthermore, the impacts of the number of PV curves forecast in various system sizes have not been investigated.

This paper aims to fill the gap from previous studies in SUC, namely the impacts of the number of PV forecast curves in various systems sizes. This paper also provides a step-by-step reformulation of DUC into SUC. Furthermore, a simple k-means clustering is also presented as an example to transform a high number of weightless forecasts into a weighted forecast sample that can represent the original forecast.

The paper is organized as follows: Section II presents the mathematical formulation of the DUC and SUC in the form of a mixed-integer linear problem (MILP) problem. Section III covers the methodologies to reproduce all the simulations used in this paper, namely electrical power system data and PV data.

Section IV presents the data of the hardware and software used to run the simulation. Section V analyses the result of the proposed method. Furthermore, this section also investigates the effect of the number of forecast points considered in various system sizes. Finally, Section VI is the conclusion.

II. MATHEMATICAL FORMULATION

A. DETERMINISTIC UNIT COMMITMENT

The objective of DUC is to minimize total cost. For each time $t \in T$ and generating unit $g \in G$, the objective to minimize total cost consisting of a startup, shutdown, and operation cost is as follows,

$$\min c = \sum_{t=1}^{T} \sum_{g=1}^{G} \begin{pmatrix} C_g^{su} v_{t,g} \\ +C_g^{sd} w_{t,g} \\ +c_{t,g}^{op} u_{t,g} \end{pmatrix} \quad (1)$$

where $v_{t,g}$, $w_{t,g}$, and $u_{t,g}$ denote the binary decision variable of startup, shutdown, and status of the generating unit, respectively. The value of C_g^{su} and C_g^{sd} are constant and they respectively denote the cost to turn on and off a generating unit g. Since the formulation of operation cost $c_{t,g}^{op}$ is a function of dispatched power $p_{t,g}$, it creates a multiplication between two decision variables. To avoid this multiplication, the operational cost can be expanded as follows,

$$c_{t,g}^{op} u_{t,g} = \alpha_{g,0} u_{t,g} + \sum_{l=1}^{L} (\alpha_{g,l} p_{t,g,l}) \quad (2)$$

where $\alpha_{g,0}$ and $\alpha_{g,l}$ denote the constant part of $c_{t,g}^{op}$ for every piecewise $l \in L$. Additionally, IEEE 10 generating units has \$0 shutdown cost C_g^{sd}. It means that the shutdown cost component $C_g^{sd} w_{t,g}$ can be removed. Thus, the formulation of the objective function in (1) for DUC can be rewritten into,

$$\min c = \sum_{t=1}^{T} \sum_{g=1}^{G} \begin{pmatrix} C_g^{su} v_{t,g} \\ +\alpha_{g,0} u_{t,g} \\ +\sum_{l=1}^{L} (\alpha_{g,l} p_{t,g,l}) \end{pmatrix} \quad (3)$$

The objective of DUC is subject to constraints,

$$p_{t,g} = \underline{P_g} u_{t,g} + \sum_{l=1}^{L} p_{t,g,l} \quad (4)$$

$$\underline{P_g} u_{t,g} \leqslant p_{t,g} \leqslant \overline{P_g} u_{t,g} \quad (5)$$

$$\sum_{g=1}^{G} p_{t,g} u_{g,t} = P_t^{net} = P_t^{d} - P_t^{sun} \quad (6)$$

$$\left(\overline{P_g} - \underline{P_g}\right) v_{t,g} + p_{t,g} \leq \overline{P_g} \quad (7)$$

$$\left(\overline{P_g} - \underline{P_g}\right) w_{t,g} + p_{t-1,g} \leq \overline{P_g} \quad (8)$$

$$\sum_{g=1}^{G} \left(\left(\overline{P}_g - p_{t,g} \right) u_{t,g} \right) \geq SR_{\%} P_t^{d} \quad (9)$$

$$\sum_{\tau=t-T_g^{up}}^{t-1} u_{\tau,g} \geq T_g^{up} w_{t,g}, \forall t \geq T_g^{up} \quad (10)$$

$$T_{g,0}^{up} + \sum_{\tau=1}^{t-1} u_{\tau,g} \geq T_g^{up} w_{t,g}, \forall t < T_g^{up} \quad (11)$$

$$\sum_{\tau=t-T_g^{dn}}^{t-1} (1 - u_{\tau,g}) \geq T_g^{dn} v_{t,g}, \forall t \geq T_g^{dn} \quad (12)$$

$$T_{g,0}^{dn} + \sum_{\tau=1}^{t-1} (1 - u_{\tau,g}) \geq T_g^{dn} v_{t,g}, \forall t < T_g^{dn} \quad (13)$$

where $p_{t,g,l}$ denotes power produced by generator g at time t in piecewise l. The $\underline{P_g}$ and $\overline{P_g}$ denote the constant value of the minimum and maximum power dispatched from generating unit g, respectively. The P_t^{net}, P_t^{d}, and P_t^{sun} denote net load, power demand or load, and PV power output profile curves respectively. The $SR_{\%}$ denotes spinning reserve requirements in the percentage of load demand at time t. T_g^{up} and T_g^{dn} denote minimum up-time and down-time for unit g in hours, while $T_{g,0}^{up}$ and $T_{g,0}^{dn}$ denote the duration of generator g has been turned on or off before the simulation starts, respectively. Constraints in (4) and (5) ensure that generating unit only produces power if the unit is turned on and within the unit operation limits. Constraints in (6) ensure that the total output power matches the required net load. Constraints in (7) and (8) ensure that each generator output power must equal the minimum power limit of the corresponding unit before being shut down and after being started up. Constraints in (9) ensure that there is a spinning reserve available. Constraints in (10)–(13) ensure that the operation follows the required minimum up-time and down-time of each generating unit.

The DUC objective and constraint formulation follow the MILP equation format as follows,

$$\min_{x} c^T x \quad (14)$$

$$Ax \preccurlyeq b \quad (15)$$

$$\underline{x} \preccurlyeq x \preccurlyeq \overline{x} \quad (16)$$

where c^T denotes constants to calculate the objective function in row vector format. The A, b, $\underline{x}$, and $\overline{x}$ denote constants for MILP formulation with decision variable x. The notation of $Ax \preccurlyeq b$ means that every row that resulted from Ax must be less than or equal to the corresponding row of b. The A is in matrix form, whereas b, $\underline{x}$, $\overline{x}$, and x are in column vector form.

B. STOCHASTIC UNIT COMMITMENT

In contrast to DUC that only considers a single net load profile or curve P_t^{net}, SUC considers various possibilities of net load profiles or curves $P_{s,t}^{net}$. To model those various net load profiles in a single MILP problem, the objective function and constraints of DUC need to be reformulated. All constraints are grouped into state-independent variables, marked with D subscript; and state-dependent variables, marked with S subscript. The state-independent variables are all variables that have the same values across all states, such as on/off status, while the state-dependent variables are all variables that have different values across all states, even when at the same time segment, such as power generation. In general, the MILP formulation can be written as follows,

$$\min_{x} (c_D^T x_D + c_S^T x_S) \quad (17)$$

$$\begin{bmatrix} A_D & 0 \\ A_{DS} & A_S \end{bmatrix} \begin{bmatrix} x_D \\ x_S \end{bmatrix} \preccurlyeq \begin{bmatrix} b_D \\ b_S \end{bmatrix} \quad (18)$$

$$\begin{bmatrix} \underline{x_D} \\ \underline{x_S} \end{bmatrix} \preccurlyeq \begin{bmatrix} x_D \\ x_S \end{bmatrix} \preccurlyeq \begin{bmatrix} \overline{x_D} \\ \overline{x_S} \end{bmatrix} \quad (19)$$

The notation of $Ax \preccurlyeq b$ is transformed into new constraint notations: $A_D x_D \preccurlyeq b_D$ and $A_{DS} x_D + A_S x_S \preccurlyeq b_S$. The state-independent constraints $A_D x_D \preccurlyeq b_D$ are identical and can be obtained directly from $Ax \preccurlyeq b$, whereas the new constraint equations $A_{DS} x_D + A_S x_S \preccurlyeq b_S$ are constraints that consist of state-dependent variable x_S. In this study, the state-independent variables x_D consisted of $v_{t,g}$, $w_{t,g}$, and $u_{t,g}$, whereas the state-dependent variables x_S consisted of $p_{s,t,g}$ and $p_{s,t,g,l}$. Thus, the

TABLE I
LINEARIZED GENERATING UNITS DATA FOR IEEE 10-UNIT UC PROBLEM

Parameters	Value				
	Unit 1	*Unit 2*	*Unit 3*	*Unit 4*	*Unit 5*
α_0 ($/h)	3,438.00	3,565.00	1,032.00	1,010.00	944.00
α_1 ($/MWh)	16.40	17.40	16.78	16.69	20.12
α_2 ($/MWh)	16.56	17.50	17.02	16.95	20.77
Parameters	**Value**				
	Unit 6	*Unit 7*	*Unit 8*	*Unit 9*	*Unit 10*
α_0 ($/h)	817.00	1,173.00	919.00	937.00	948.00
α_1 ($/MWh)	22.74	27.83	26.10	27.39	27.86
α_2 ($/MWh)	23.23	27.84	26.30	27.45	27.92

objective function of SUC in the form of MILP can be written as follows,

$$\min c = \sum_{t=1}^{T} \sum_{g=1}^{G} \begin{pmatrix} C_g^{su} v_{t,g} \\ +\alpha_{g,0} u_{t,g} \\ +c^s \end{pmatrix} \quad (20)$$

$$c^s = \sum_{s=1}^{S} \mu_s \left(\sum_{l=1}^{L} \alpha_{g,l} p_{s,t,g,l} \right) \quad (21)$$

where μ_s denotes the weight factor of scenario s based on the occurrence probability. Therefore, the sum of μ_s for all $s \in S$ must be equal to 1. The objective of SUC is subject to DUC constraints, with some modifications in constraints containing state-dependent variables. In this paper, the stochastic forecasts come from the net load, specifically caused by high uncertainty PV forecast $P_{s,t}^{sun}$. Hence, the DUC constraints in (4)–(9) are replaced by following SUC constraints,

$$p_{s,t,g} = \underline{P_g} u_{t,g} + \sum_{l=1}^{L} p_{s,t,g,l} \quad (22)$$

$$\overline{P_g} u_{t,g} \leqslant p_{s,t,g} \leqslant \underline{P_g} u_{t,g} \quad (23)$$

$$\sum_{g=1}^{G} p_{s,t,g,l} u_{g,t} = P_{s,t}^{net} = P_t^d - P_{s,t}^{sun} \quad (24)$$

$$\left(\overline{P_g} - \underline{P_g}\right) v_{t,g} + p_{s,t,g} \leq \overline{P_g} \quad (25)$$

$$\left(\overline{P_g} - \underline{P_g}\right) w_{t,g} + p_{s,t-1,g} \leq \overline{P_g} \quad (26)$$

$$\sum_{g=1}^{G} \left(\left(\overline{P}_g - p_{s,t,g}\right) u_{t,g} \right) \geq SR_{\%} P_t^d \quad (27)$$

The outputs of day-ahead SUC contain state-independent variables that must have the same value for all $s \in S$ ($v_{t,g}$, $w_{t,g}$, and $u_{t,g}$), and state-dependent variables that are allowed to be different across all scenarios ($p_{s,t,g}$ and $p_{s,t,g,l}$). In real-time operations, only the state-independent variable is fixed, while the state-dependent variable is allowed to be recalculated to satisfy operational constraints.

To test the performance of the SUC solution, a test in simulated real-time operation was conducted. The test used the DUC objective function and constraints; thus, no state-dependent variables were used. However, the DUC power balance constraints in (6) and (9) were replaced with the following equations,

$$\sum_{g=1}^{G} p_{t,g} u_{g,t} + p_t^{shed} - p_t^{dmy} = P_t^d - P_t^{sun} \quad (28)$$

$$\sum_{g=1}^{G} \left(\left(\overline{P}_g - p_{t,g}\right) u_{t,g} \right) + p_t^{SRviol} \geq SR_{\%} P_t^d \quad (29)$$

where p_t^{shed}, p_t^{dmy}, and p_t^{SRviol} denote load shedding, reserve violation, and dummy load at time t, respectively. To encourage the solver to maximize the value of load shedding and reserve violation, an arbitrary high cost was added as the value of lost load and reserve violation penalty. Thus, the objective function in simulated real-time operation replaced the objective in (3) with the following equation,

$$\min c = \sum_{t=1}^{T} \sum_{g=1}^{G} \begin{pmatrix} C_g^{su} v_{t,g} \\ +\alpha_{g,0} u_{t,g} \\ +\sum_{l=1}^{L} \left(\alpha_{g,l} p_{t,g,l} \right) \\ +\psi^{shed} p_t^{shed} \\ +\psi^{SRViol} p_t^{SRViol} \end{pmatrix} \quad (30)$$

where ψ^{shed} and ψ^{SRViol} denote the constant for load shedding penalty and reserve violation penalty. Then, a separate calculation outside the solver could be used to calculate only the total operational cost based on (3).

In the simulated real-time test, the commitment solution ($v_{t,g}$, $w_{t,g}$, and $u_{t,g}$) of every approach, both SUC and DUC, were fixed based on the outputs of the day-ahead solution of the corresponding approach. To evaluate the performance of every simulation, the following average generation cost c^{gen} was used.

$$c^{gen} = \frac{c}{\left(P^d - P^{sun} - p^{shed} + p^{dmy}\right)} \quad (31)$$

where P^d, P^{sun}, p^{shed}, and p^{dmy} are all in the cumulative sum of all time $t \in T$. The value of c here is based on (3).

III. TEST SYSTEM

A. ELECTRICAL POWER SYSTEMS DATA

The test system used to test both the DUC and SUC is the IEEE 10-unit UC problem [11]. The IEEE 10-unit UC problem was selected because it is commonly used by other algorithms to compare a novel method with an existing method. A small modification of the quadratic cost constant was applied to change the quadratic cost function into a linear piecewise cost function containing two piecewise each. The linearization method for the fuel cost was based on SciPy curve-fit [12], [13]. The resulting piecewise cost can be seen in Table I.

To simulate the system test with more than twenty generating units, the original ten generating units of the IEEE 10-unit UC problem were duplicated, and the load was multiplied accordingly. For simulating a PV power plant, a fixed PV power plant with a size equal to 20% of peak load was used, making the size of the PV power plant became 300 MW, 600 MW, 900 MW, 1.2 GW, and 1.5 GW for 10, 20, 30, 40, and 50 generating unit cases, respectively.

B. PHOTOVOLTAIC DATA

Forecast data used in this study were generated using scikit-learn k-means clustering [14], specifically the Llyod algorithm [15]. The clustering approach is chosen because of its simplicity compared to other forecast method such as Markov switching [16] and seasonal forecast [17], [18]. Furthermore, the clustering approach can generate flexible numbers of profiles based on the number of centroids used. In future research, other scenario reduction algorithm can be employed such as important sampling [19], and dynamic time warping clustering [20], [21]. The inputs for the clustering were data number 724030 of the National Renewable Energy Laboratory (NREL) [22]. The first 19 years' data were used as input data for k-means clustering, while the last year's data were used as test data. The output of k-means clustering was forecast data for the input of SUC. The proportion of the cluster member within the group to all data was used as the weight or probability of occurrence of the cluster centers. Thus, the input of SUC for day-ahead UC was multiple PV curves with the occurrence probability of each state.

The solution of UC was tested using data of 365 days. Each day in the test was treated as a single independent day, in which there was no relation between each day in the test. Thus, there were no between-day constraints such as minimum uptime and downtime between each day.

The raw output of NREL data number 724030 was in MW/m^2. To make it into percent energy produced by PV plants, normalization was done in the original data. The value of the highest PV solar radiation used for normalization was obtained from the highest value of the learning data set, which was 1,014 MW/m^2. This normalization was used in the day-ahead UC and the test. The normalized PV profile was used as the multiplier of PV size to create a PV power output profile.

C. TEST ENVIRONMENT

In this paper, the performances of various UC approaches in two different environments were tested. The first environment was day-ahead UC. In this environment, each approach was given several possibilities of the next-day PV profile forecast with the corresponding occurrence probability. In the second environment, a simulated real-time market was used. In this environment, the solution of generating units schedule was used whereas the power dispatch was recalculated based on the pseudo-real-time value of PV output.

IV. HARDWARE AND SOFTWARE INFORMATION

The simulation was conducted on HP Joy 2 laptop with Windows 10 operating system, AMD A4-9125 Dual-Core 2.3 GHz, and 4 GB RAM. The solver used to run the simulation was a MILP solver by CPLEX 12.9.0.0 [23] with the branch-and-cut algorithm. The CPLEX solver was run on Python 3.7.4.

V. RESULTS AND DISCUSSION

This section analyzes and compares various day-ahead UC solutions and parameters between five distinct UC perspectives or strategies. Two approaches were based on RUC, which considered only the worst-case scenarios, those were the cloudy worst-case scenario (WC) and no consideration of the PV scenario (NC). On the other hand, two optimistic perspectives were also considered, those were sunny best-case scenario (BC) and average irradiance case scenario (AIC). All those four perspectives used the DUC formulation. The last perspective was the stochastic approach that was based on the multiforecast of PV curves (MC). For WC, BC, and MC scenarios, the number following the name means the number of centroids used. For example, WC-10 means the worst-case scenario of a forecast containing 10 forecast curves is selected, whereas MC-10 means that ten scenarios are used for SUC inputs.

A. PHOTOVOLTAIC FORECAST

Ten separate forecast sets were generated using k-means clustering [14], [15] from the original data. The simplest set with one cluster center was used as the average irradiance forecast. The number of cluster centers was increasing for each forecast set, with the first set had only a single forecast. On the other hand, the tenth set had ten different forecast curves. The total daily energy in MWh generated from PV power plants with its occurrence probability, sorted from the highest probability, for the ten sets of forecasts can be seen in Table II.

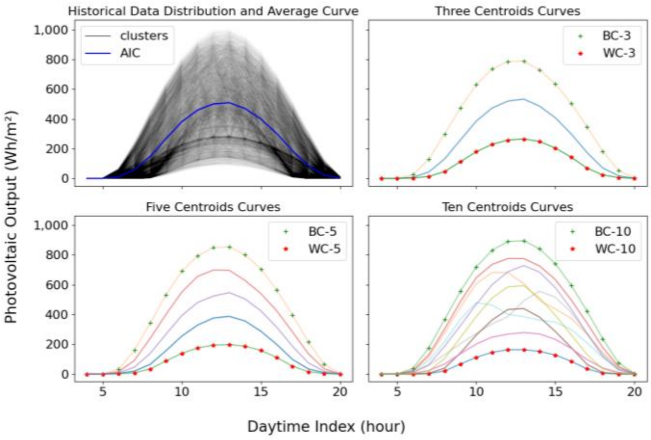

Figure 1. Photovoltaic historical data distribution and clustering results.

TABLE II
DAY-AHEAD TOTAL DAILY PV GENERATION (MWH) AND OCCURRENCE PROBABILITY

Cluster Numbers				
1	***2***	***3***	***4***	***5***
1,124/1.00	691/0.57	532/0.37	888/0.29	764/0.25
-/-	1,692/0.43	1,126/0.34	454/0.26	405/0.21
-/-	-/-	1,870/0.29	1,430/0.25	1,164/0.20
-/-	-/-	-/-	1,999/0.20	1,599/0.19
-/-	-/-	-/-	-/-	2,074/0.15
Cluster Numbers				
6	***7***	***8***	***9***	***10***
742/0.23	727/0.22	644/0.19	626/0.17	604/0.16
403/0.20	396/0.20	1,705/0.15	823/0.15	773/0.14
1,613/0.19	2,066/0.15	861/0.15	329/0.14	321/0.13
1,142/0.15	1,591/0.14	338/0.14	2,110/0.12	1,123/0.11
2,079/0.14	1,079/0.13	2,121/0.12	1,666/0.11	1,868/0.10
1,164/0.08	1,448/0.08	1,265/0.11	1,227/0.11	1,537/0.10
-/-	1,113/0.07	1,394/0.07	1,646/0.08	2,189/0.08
-/-	-/-	1,124/0.07	1,132/0.06	1,541/0.07
-/-	-/-	-/-	1,171/0.06	1,132/0.06
-/-	-/-	-/-	-/-	1,123/0.05

The total daily energies are shown on the left side of the slash, while the probabilities are shown on the right side of the slash.

From Table II, it can be seen that by increasing the number of forecast points, the gap in total daily energy produced by PV sources between the highest energy output (BC) and the lowest power energy output (WC) for that particular number of centroids also tended to increase. It means that if an electrical power system utility decides to use the RUC approach by considering only the worst-case scenario, it overlooks the opportunity to turn off some generating units if other scenarios appear in the real-time market. Thus, the RUC approach misses the chance to reduce overall operational costs. For example, in the ten centroids case, the RUC approach would only consider the PV profile with 321 MWh, neglecting the other profiles with a total of 87% occurrence probability. Choosing only the average profile or single centroid alone also might not help. It is because all forecast sets were slightly skewed towards the worst-case scenario. For example, in the ten centroids case, the AIC approach only used the single centroid with 1,124 MWh total energy produced while neglecting a worst-case scenario with only 321 MWh total energy output.

The shape of PV profile data and some of the clustering results can be seen in Figure 1. The darker the plot in Figure 1,

TABLE III
GENERATOR WORK HOURS ON IEEE 10- UNIT DAY-AHEAD UC

Case	Unit Work Hours (hours)									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
NC	24	24	17	19	20	9	9	5	2	1
AIC	24	24	17	19	20	8	3	4	1	0
BC-2	24	24	13	18	20	6	3	3	0	0
WC-2	24	24	17	19	20	9	7	3	0	0
MC-2	24	24	17	19	20	9	7	3	0	0

TABLE IV
TOTAL COST ($) AND GENERATION COST ($/MWH) ON IEEE 10- UNIT DAY-AHEAD UC

Case	Total Cost ($)	Generated Energy (MWh)	Generation Cost ($/MWh)
NC	567,145.6332	27,100	20.9279
AIC	537,596.8160	25,976	20.6960
BC-2	523,399.3931	25,408	20.5995
WC-2	548,005.1606	26,409	20.7507
MC-2	539,896.0258	25,976	20.7845

the higher the probability of occurrence. It can be seen that the shape of the historical data is in the form of a bell-shaped curve, with the lower area of the bell being darker compared to the top area. It confirms what is presented in Table II that most of the data are under the average value. However, it can be seen that the darkest area is not the one at the bottom of the plot. It further supports what has been presented in Table II that the most probable scenario is a low irradiance PV profile but does not necessarily mean the lowest area of the bell-shaped profile.

B. DAY-AHEAD UNIT COMMITMENT SOLUTION

In this study, the deterministic and stochastic solutions on the day-ahead UC problem were compared. The system used in this section was the base case of the IEEE 10-unit system that consisted of ten generating units. The sample of unit work hours from two centroids case scheduling solution of day-ahead UC can be seen in Table III. It can be seen that all the solution has their own schedule, except for the worst-case scenario (WC-2) and SUC solution (MC-2). The exact same solution between WC-2 and MC-2 is caused by the SUC that must meet all constraints caused by all given PV curves in the constraint declaration. Thus, the system in SUC was prepared for all the possible PV curves.

The total cost, generated energy, and generation cost of the day-ahead UC problem can be seen in Table IV. The value of generated energy for MC-2 was a weighted generated energy based on the probability of every possible scenario. It can be seen that the total cost of MC-2 is in the middle of the total cost of BC-2 and WC-2 because it considers both cases. This total cost can be seen as the expected total cost that the electrical power systems utility needs to be prepared for. Although MC-2 has a lower total cost compared to WC-2, it can be seen that MC-2 has a higher generation cost. It is because MC-2 has lower generated power. It means that the SUC approach of MC-2 requires a lower total cost compared to the RUC approach of WC-2, but resulted in higher generation cost due to the generating units in the SUC approach producing lower total energy, resulting in higher payment of generated energy to IPPs.

By looking at Table IV, it is argued that the SUC approach of MC-2 produces the best generation cost to be offered to IPPs. It is because it used all possible curves with their corresponding occurrence probability, unlike the others. The generation cost produced by NC gave a price that was too expensive for utility, while BC did the opposite. Although AIC and WC-2 produced a middle value between the two extremes, it did not consider whether the PV profile ended up having a higher value. Hence, the solution of MC-2 is the best compared to the others as it is prepared for the worst-case scenario yet produces a generation cost that considers the probability of high irradiance profile. It means that the generation cost formulation of SUC also considers the probability of over-commitment if the best-case scenario happens.

TABLE V
GENERATING COST IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Number of Centroids	Generation Cost ($/MWh)				
	NC	*AIC*	*BC*	*WC*	*MC*
0	20.9862	-	-	-	-
1	-	20.7041	-	-	-
2	-	-	20.5798	20.7864	20.7864
4	-	-	20.5029	20.8165	20.8165
6	-	-	20.5029	20.8443	20.8443
8	-	-	20.5037	20.8443	20.8443
10	-	-	20.4979	20.8443	20.8443

C. UNIT COMMITMENT PERFORMANCE ON TEST

A good solution for UC must be able to satisfy all constraints present in the actual next day's market. The solutions of day-ahead UC were tested using the independent PV profiles of 365 days. Furthermore, both the impact of the number of centroids and the size of the electrical power system on its performance were investigated. Both performance investigations were done in the test environment.

To test the performance of the resulting generating unit schedule by various UC solutions, a test scenario was used. The test scenario consists of ten generating units from IEEE 10-unit data. All solutions of day-ahead UC were tested in 365 independent tests. Each test had its PV profile that was not visible in the day-ahead process of making the generating unit schedule. The generation cost of each solution in the test environment can be seen in Table V. It needs to be noted that the NC approach was not using any PV profile, thus the value of the centroid was zero, while the average irradiance case approach of AIC used only single PV profile, that was the average profile. On the other hand, the RUC approach of WC and the SUC approach of MC were used when there were at least two PV profiles, and so did the optimistic approach of BC.

As presented in Table V, increasing the number of forecast PV curves will introduce more extreme PV curves. As centroids increase, the BC perspective becomes more optimistic about the next day's value and schedules fewer units, resulting in low generation costs. On the other hand, WC and MC become more careful in turning off a generating unit, resulting in higher generation costs.

Since different schedules result in different spinning reserve values, a UC solution impacts system reliability. The impact of centroid number on spinning reserve violation and load shedding can be seen in Table VI and Table VII. Even though BC gives the least operational cost, it comes with spinning reserve violations and even load shedding, as does AIC (one centroid case). Thus, neither BC nor AIC produced a robust schedule and should not be chosen as a UC solution approach. It can also be seen that MC produced a robust solution just like the RUC approach (NC and WC). Furthermore, although turning on lots of generators, no dummy load was used by all approaches.

TABLE VI
SPINNING RESERVE VIOLATION IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Number of Centroids	Spinning Reserve Violation (MWh/day)				
	NC	*AIC*	*BC*	*WC*	*MC*
0	0	-	-	-	-
1	-	76.08	-	-	-
2	-	-	243.76	10.96	10.96
4	-	-	329.56	2.38	2.38
6	-	-	329.56	1.02	1.02
8	-	-	378.04	1.02	1.02
10	-	-	384.75	1.02	1.02

TABLE VII
LOAD SHEDDING IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Number of Centroids	Load Shedding (MWh/day)				
	NC	*AIC*	*BC*	*WC*	*MC*
0	-	-	-	-	-
1	-	2.59	-	-	-
2	-	-	75.32	0	0
4	-	-	83.87	0	0
6	-	-	83.87	0	0
8	-	-	64.88	0	0
10	-	-	133.89	0	0

TABLE VIII
GENERATING COST IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Case	Generation Cost ($/MWh)				
	10	*20*	*30*	*40*	*50*
NC	20.9862	20.8822	20.8581	20.8472	20.8217
AIC	20.7041	20.5821	20.5424	20.5270	20.5251
BC-2	20.5798	20.4634	20.4412	20.4320	20.4279
WC-2	20.7864	20.6599	20.6330	20.6202	20.6126
MC-2	20.7864	20.6599	20.6325	20.6199	20.6128
BC-10	20.4979	20.4115	20.3913	20.3823	20.3719
WC-10	20.8443	20.7539	20.7285	20.7175	20.7094
MC-10	20.8443	20.7526	20.7253	20.7175	20.7092

The small number of the generating units in IEEE 10-unit restricts the number of possible solutions. This very limited possible solution limits the flexibility of SUC in making unit commitment schedules. That is why, in Table V, all SUC solutions ended up in the same solutions from the worst-case scenario of RUC. By increasing the number of available generating units, SUC was given more flexibility in making unit commitment schedules. The impact of the system size on the SUC solution can be seen in Table VIII, Table IX, and Table X. Only the lowest and highest centroids numbers are shown for the comparison, that is two and ten centroids respectively. The column title represents the number of generating units in the system.

In terms of the SUC solution compared to the RUC as shown in Table VIII, it can be seen that there is a slight difference between those two in some medium-sized systems. It means that SUC shifts its solution a little towards the possibility of a BC. Since the PV curve probability distribution used in this study was heavily leaning towards the WC, it is understandable that the shift of the SUC diverging from WC is only a little. If the probability of the occurrence of BC increases, this shift most likely also increases, producing a more distinctive result of the unit commitment schedule.

The impact of system size on the SUC solution in terms of spinning reserve violation and load shedding can be seen in Table IX and Table X, respectively. It can be seen from Table

TABLE IX
SPINNING RESERVE VIOLATION IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Case	Spinning Reserve Violation (MWh/day)				
	10	*20*	*30*	*40*	*50*
NC	0	0	0	0	0
AIC	76.08	199.02	425.03	640.64	810.47
BC-2	243.76	648.78	1,009.32	1,425.42	1,805.54
WC-2	10.96	52.19	114.19	146.07	194.97
MC-2	10.96	52.19	118.22	146.07	213.65
BC-10	384.75	922.29	1,584.04	2,191.45	2,758.69
WC-10	1.02	2.05	4.39	5.54	9.95
MC-10	1.02	2.47	3.93	5.54	9.95

TABLE X
LOAD SHEDDING IN TEST ENVIRONMENT BASED ON IEEE 10-UNIT

Case	Load Shedding (MWh/day)				
	10	*20*	*30*	*40*	*50*
NC	0	0	0	0	0
AIC	2.59	8.11	15.30	43.87	50.47
BC-2	75.32	96.45	178.39	287.70	411.81
WC-2	0.00	0.01	0.18	0.02	0.10
MC-2	0.00	0.01	0.18	0.02	0.10
BC-10	133.89	301.99	477.50	758.70	910.89
WC-10	0	0	0	0	0
MC-10	0	0	0	0	0

TABLE XI
COMPUTATION TIME COMPARISON BETWEEN SUC AND RUC IN SECOND

Case	Computation Time (second)				
	10	*20*	*30*	*40*	*50*
WC-10	0.50	10.30	111.10	39.00	17.40
MC-2	1.50	14.80	37.10	59.60	45.60
MC-4	1.20	15.70	455.00	84.90	110.60
MC-6	2.90	43.80	2497.50	674.40	125.00
MC-8	5.10	136.70	970.20	737.30	218.80
MC-10	4.20	310.80	1473.80	865.40	254.90

IX and Table X that SUC kept its robust solution in medium-sized systems while having a lower generation cost, as shown in Table VIII. Thus, it can be concluded that SUC solutions are cheaper than WC but have a similar result of spinning reserve violation without having load shedding. The best result in this study occurred in a 30 units system test in which SUC with 10 centroids had a cheaper solution (20.7253 $/MWh) compared to RUC (20.7285 $/MWh) without violating power balance or going to load shedding. Furthermore, the SUC solution even managed to have a lower spinning reserve violation on that system size, further exhibiting the superiority of MSUC.

D. COMPUTATION TIME OF UNIT COMMITMENT

A day-ahead deterministic unit commitment problem can be formulated as a mixed-integer programming (MIP) [2]. It is already well known that the general MIP optimization problem is an NP-hard (at least as hard as non-deterministic polynomial-time) problem [24]. In this study, mixed-integer linear programming was chosen instead of the original mixed-integer quadratic programming to reduce the computational time of UC.

A unit commitment problem in the form of a MILP optimization problem consisted of a set of decision variables and constraints. By adding multiple centroids as forecast curves in SUC, both the number of decision variables and constraints were also increased. The impact of the number of centroids in

SUC across various system sizes can be seen in Table XI. The computation time of the RUC solution of WC-10 is also presented as the baseline. It can be seen from Table XI that SUC generally increased the computation time of UC, but not in an exponential manner. Using up to ten centroids, the computational time increased near linear time to the number of centroids used. Thus, the SUC is a feasible approach compared to using a probability density function and iterating through thousands of DUC problems based on Monte Carlo simulation [25]. Furthermore, simulations using Monte Carlo did not have the advantage of SUC, that is having only single unit commitment schedule.

VI. CONCLUSION

An increase in PV penetration means an increase in disparity of the forecast of the next day's net load profile. Combined with the intermittent and uncertain nature of PV power plants, a robust solution of unit commitment is needed. This study has proposed an approach of stochastic unit commitment in solving a day-ahead unit commitment problem in a system with a high uncertainty net load. In contrast with robust unit commitment which only considers the worst-case scenario, the proposed approach considers every possible scenario of the next day's net load profile. The next day net load profile was acquired using k-means clustering on historical data. The proportion of cluster members is used as a weight factor that represents the occurrence probability of PV curves.

This study has shown that the proposed stochastic unit commitment using k-means clustering produced a fairer price of electricity generating cost compared to a robust unit commitment with the same robustness in the day-ahead unit commitment problem. It also has shown that the unit commitment schedule of the stochastic approach in the day-ahead unit commitment was able to outperform the robust unit commitment with the same robustness. All those advantages are also shown to only increase the computation time near linear time to the number of centroids used.

In future works, various approaches to generate the next day net load profile can be studied. It includes both the algorithm to forecast, such as seasonal forecast and Markov switching; and the scenario reduction algorithm, such as important sampling, and dynamic time warping clustering. Furthermore, a more in-depth study is needed to better determine the impact of the number of centroids used in stochastic unit commitment on the computational complexity based on big O notation.

CONFLICT OF INTEREST

The authors declare that they have no known conflicting interests that could influence the work of this paper.

AUTHOR CONTRIBUTION

Conceptualization, Muhammad Yasirroni and Sarjiya; methodology, Muhammad Yasirroni, Sarjiya, and Lesnanto Multa Putranto; software, Muhammad Yasirroni; validation, Sarjiya and Lesnanto Multa Putranto; formal analysis, Muhammad Yasirroni and Indra Triwibowo; investigation, Muhammad Yasirroni; resources, Sarjiya and Lesnanto Multa Putranto; data curation, Muhammad Yasirroni; writing—original draft preparation, Indra Triwibowo; writing—review and editing, Muhammad Yasirroni, Sarjiya, Husni Rois Ali, Indra Triwibowo, and Qiangqiang Xie; visualization, Muhammad Yasirroni; supervision, Sarjiya; project administration, Sarjiya; funding acquisition, Sarjiya and Lesnanto Multa Putranto.

ACKNOWLEDGMENT

The authors would like to thank the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada for the support.

REFERENCES

[1] Lazard, "Lazard's Levelized Cost of Energy Analysis—Version 13.0," Lazard, New York, United States, Rep., 2019.

[2] R.H. Kerr, A.J. Fontana, Jr., J.L. Scheidt, and J.K. Wiley, "Unit commitment," *IEEE Trans. Power Appar., Syst.*, Vol. PAS-85, No. 5, pp. 417–421, May 1966, doi: 10.1109/TPAS.1966.291678.

[3] N.P. Padhy, "Unit Commitment-A Bibliographical Survey," *IEEE Trans. Power Syst.*, Vol. 19, No. 2, pp. 1196–1205, May 2004, doi: 10.1109/TPWRS.2003.821611.

[4] E.C. Kern, E.M. Gulachenski, and G.A. Kern, "Cloud Effects on Distributed Photovoltaic Generation: Slow Transients at the Gardner, Massachusetts Photovoltaic Experiment," *IEEE Power Eng. Rev.*, Vol. 9, No. 6, pp. 43–44, Jun.1989, doi: 10.1109/MPER.1989.4310752.

[5] N. Nguyen and J. Mitra, "An Analysis of the Effects and Dependency of Wind Power Penetration on System Frequency Regulation," *IEEE Trans. Sustain. Energy*, Vol. 7, No. 1, pp. 354–363, Jan. 2016, doi: 10.1109/TSTE.2015.2496970.

[6] D. Bertsimas *et al.*, "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem," *IEEE Trans. Power Syst.*, Vol. 28, No. 1, pp. 52–63, Feb. 2013, doi: 10.1109/TPWRS.2012.2205021.

[7] P. Pourghasem *et al.*, "Stochastic Multi-Objective Dynamic Dispatch of Renewable and Chp-Based Islanded Microgrids," *Electr. Power Syst. Res.*, Vol. 173, pp. 193–201, Aug. 2019, doi: 10.1016/j.epsr.2019.04.021.

[8] A. Papavasiliou and S.S. Oren, "Multiarea Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network," *Oper. Res.*, Vol. 61, No. 3, pp. 578–592, May-Jun. 2013, doi: 10.1287/opre.2013.1174.

[9] R.B. Hytowitz and K.W. Hedman, "Managing Solar Uncertainty in Microgrid Systems with Stochastic Unit Commitment," *Electr. Power Syst. Res.*, Vol. 119, pp. 111–118, Feb. 2015, doi: 10.1016/j.epsr.2014.08.020.

[10] E.A. Bakirtzis, C.K. Simoglou, P.N. Biskas, A.G. Bakirtzis, "Storage Management by Rolling Stochastic Unit Commitment for High Renewable Energy Penetration," *Electr. Power Syst. Res.*, Vol. 158, pp. 240–249, May 2018, doi: 10.1016/j.epsr.2017.12.025.

[11] S. Kazarlis, A.G. Bakirtzis, and V. Petridis, "A Genetic Algorithm Solution to the Unit Commitment Problem," *IEEE Trans. Power Syst.*, Vol. 11, No. 1, pp. 83–92, Feb. 1996, doi: 10.1109/59.485989.

[12] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods*, Vol. 17, No. 3, pp. 261–272, Feb. 2020, doi: 10.1038/s41592-019-0686-2.

[13] M.A. Branch, T.F. Coleman, and Y. Li, "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems," *SIAM J. Sci. Comput.*, Vol. 21, No. 1, pp. 1–23, 1999, doi: 10.1137/S1064827595289108.

[14] F. Pedregosa *et al.*, "Scikit-Learn: Machine Learning in Python," 2012, *arXiv:1201.0490*.

[15] S.P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, Vol. 28, No. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.

[16] R.L.G. Latimier, E.L. Bouedec, and V. Monbet, "Markov Switching Autoregressive Modeling of Wind Power Forecast Errors," *Electr. Power Syst. Res.*, Vol. 189, pp. 1–7, Dec. 2020, doi: 10.1016/j.epsr.2020.10664.

[17] M.M.H. Shawon *et al.*, "Forecasting PV Panel Output Using Prophet Time Series Machine Learning Model," *2020 IEEE Region 10 Conf. (TENCON)*, 2020, pp. 1141–1144, doi: 10.1109/TENCON50793.2020.9293751.

[18] S.J. Taylor and B. Letham, "Forecasting at Scale," *The Am. Stat.*, Vol. 72, No. 1, pp. 37–45, Apr. 2017, doi: 10.1080/00031305.2017.1380080.

[19] A. Papavasiliou, S.S. Oren, and B. Rountree, "Applying High Performance Computing to Transmission-Constrained Stochastic Unit Commitment for Renewable Energy Integration," *IEEE Trans. Power Syst.*, Vol. 30, No. 3, pp. 1109–1120, May 2015, doi: 10.1109/TPWRS.2014.2341354.

[20] O. Yurdakul, F. Sivrikaya, and S. Albayrak, "A Distributionally Robust Optimization Approach for Unit Commitment in Microgrids," 2021, *arXiv: 2011.05314*.

[21] D.J. Berndt and J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," *AAAIWS'94: Proc. 3rd Int. Conf. Knowl. Discov., Data Min.*, 1994, pp. 359–370.

[22] S. Wilcox, "National Solar Radiation Database 1991-2010 Update: User's manual," National Renewable Energy Laboratory, Colorado, USA, Tech. Rep. TP-5500-54824, 2012.

[23] *IBM ILOG CPLEX Optimization Studio CPLEX User's Manual Version 12 Release 8*. IBM, New York, US, 2017.

[24] A.D. Pia, S.S. Dey, and M.S. Molinaro, "Mixed-Integer Quadratic Programming Is in NP," *Math. Program.*, Vol. 162, pp. 225–240, Mar. 2017, doi: 10.1007/s10107-016-1036-0.

[25] N. Petcharaks, "Optimal Spinning Reserve Under Load and Intermittent Generation Uncertainty Using Monte Carlo Simulation," *2015 IEEE Innov. Smart Grid Technol. - Asia (ISGT ASIA)*, 2015, pp. 1–6, doi: 10.1109/ISGT-Asia.2015.7387056.