Supplementary Notes on Robust Unit Commitment Model Based on Multi-state Uncertainty Sets

A. Two-stage of Robust Unit Commitment Model (RUCM)

$$\min_{x,y,z,p} \sum_t \Bigg(\sum_g C_g^{SU} y_{g,t} + C_g^0 x_{g,t} + C_g p_{g,t} + \max_{\tilde{w}\in U} \min_{\hat{p},\Delta w,\Delta d\in\Omega(x,y,z,p)} \Big(\sum_r C^{WS}\Delta w_{r,t} + \sum_j C^{LS}\Delta d_{j,t} \Big) \Bigg) \tag{1}$$

$$x_{g,t} - x_{g,t-1} = y_{g,t} - y_{g,t-1} \quad \forall g, \forall t \tag{2}$$

$$\sum_{\tau=t-TU_g+1}^{t} y_{g,\tau} \le x_{g,t}, \quad \forall g, \forall t \tag{3}$$

$$\sum_{\tau=t-TD_g+1}^{t} v_{g,\tau} \le 1 - x_{g,t}, \quad \forall g, \forall t \tag{4}$$

$$\sum_g p_{g,t} + \sum_r p_{r,t} = \sum_j D_{j,t}, \ \forall t \tag{5}$$

$$P_g^{\min} x_{g,t} \le p_{g,t} \le P_g^{\max} x_{g,t}, \ \forall g, \forall t \tag{6}$$

$$p_{g,t} - p_{g,t-1} \le RU_g x_{g,t-1} + SU_g(1 - x_{g,t-1}), \ \forall g, \forall t \tag{7}$$

$$p_{g,t-1} - p_{g,t} \le RD_g x_{g,t} + SD_g(1 - x_{g,t}), \ \forall g, \forall t \tag{8}$$

$$0 \le p_{r,t} \le \bar{w}_{r,t}, \ \forall r \in R, \forall t \tag{9}$$

$$-F_l^{\max} \le \sum_m SF_{l,m} \Big(\sum_{g\in \mathrm{G}(m)} p_{g,t} + \sum_{r\in R(m)} p_{r,t} - \sum_{j\in \mathrm{D}(m)} D_{j,t} \Big) \le F_l^{\max}, \ \forall l, \forall t \tag{10}$$

$$x_{g,t}, y_{g,t}, z_{g,t} \in \{0,1\} \tag{11}$$

i.e.,

$$\Omega(x,y,z,p) = \{(\hat{p}, \Delta w, \Delta d) : \tag{12}$$

$$\sum_g \hat{p}_{g,t} + \sum_r \hat{p}_{r,t} = \sum_j (D_{j,t} - \Delta d_{j,t}), \ \forall t \tag{13}$$

$$\max\{P_g^{\min}, p_{g,t} - R_g^{down}\} x_{g,t} \le \hat{p}_{g,t}, \quad \hat{p}_{g,t} \le \min\{P_g^{\max}, p_{g,t} + R_g^{up}\} x_{g,t} \ \forall g, \forall t \tag{14}$$

$$\hat{p}_{g,t} - \hat{p}_{g,t-1} \le (RU_g x_{g,t-1} + SU_g(1 - x_{g,t-1})), \ \forall g, \forall t \tag{15}$$

$$\hat{p}_{g,t-1} - \hat{p}_{g,t} - \le (RD_g x_{g,t} + SD_g(1 - x_{g,t})), \ \forall g, \forall t \tag{16}$$

$$\hat{p}_{r,t} + \Delta w_{r,t} = \tilde{w}_{r,t}, \ \forall r, \forall t : \lambda_{11,r,t} \tag{17}$$

$$0 \le \Delta d_{j,t} \le D_{j,t}, \ \forall j, \forall t \tag{18}$$

$$-F_l^{\max} \le \sum_m SF_{l,m} \Big(\sum_{g\in \mathrm{G}(m)} \hat{p}_{g,t} + \sum_{w\in \mathrm{W}(m)} \hat{p}_{r,t} - \sum_{j\in \mathrm{D}(m)} (D_{j,t} - \Delta d_{j,t}) \Big) \le F_l^{\max}, \quad \forall l, \forall t\} \tag{19}$$

The objective function (1) is to minimize the total cost, including the cost of unit commitment, the generation cost at the expectation scenario(or first stage(or base) scenario), and the cost of wind curtailment and load shedding at the worst scenario. Equation (2)-(11) are the constraints at the first stage scenario. Equation (2) is the logical constraint of the start-up and shut-down states of the unit. Equations (3) and (4) are constraints about the minimum start-up and shut-down time of unit. Equation (5) is the system power balance constraint. Equation (6) is the upper and lower limit of the output of the traditional unit. Equations (7)-(8) are the traditional unit ramp constraints. Equation (9) is the constraint of the accepted wind power in the first stage. Equation (10) is the line transmission power constraints. Equation (11) is the definition domain of decision variables. $\Omega(x,y,z,p)$ is the constraint set of the optimal redispatch (or adjustment) problem at the worst scenario based on the optimization results of first stage. Equation (13) is the system power balance constraint for redispatch. Equation (14) is the upper and lower limit constraints of the traditional unit output for redispatch, indicating that the output of the units are subject to both capacity constraints and their adjustment capability. Equations (15)-(16) are the traditional unit ramp constraints for redispatch, and (17) is the accepted wind power constraint for redispatch, meaning that the actual accepted wind power does not exceed the actual available power. Equation (18) is the load shedding constraint for redispatch. Equation (19) is the line transmission power constraint for redispatch.

B. Compact Form of RUC

$$\min_{\mathbf{x},\mathbf{p}} \left(\mathbf{c}^T\mathbf{x} + \mathbf{b}^T\mathbf{p} + \max_{\mathbf{w}\in U} \min_{\mathbf{k}\in\Omega(\mathbf{x},\mathbf{p},\mathbf{w})} \mathbf{d}^T\mathbf{k} \right) \tag{20}$$

$$\text{s.t. } \mathbf{Ax} + \mathbf{Bp} \le \mathbf{h} \tag{21}$$

$$\mathbf{Jx} + \mathbf{Np} = \mathbf{n} \tag{22}$$

$$\Omega(\mathbf{x},\mathbf{p},\mathbf{w}) = \{\mathbf{k} : \mathbf{Cx} + \mathbf{Dp} + \mathbf{Ek} \le \mathbf{g} \tag{23}$$

$$\mathbf{Op} + \mathbf{Rk} + \mathbf{Fw} = \mathbf{u}\} \tag{24}$$

$$\mathbf{x} \text{ binary}, \ \mathbf{p}, \mathbf{k}, \mathbf{w} \ge 0 \tag{25}$$

In which, $\mathbf{x}$ is UC variable vector at base state and composed of $x_{g,t}, y_{g,t}, v_{g,t}$; $\mathbf{p}$ is dispatch variable vector at first stage(expectation scenario) and composed of $p_{g,t}, p_{r,t}$; $\mathbf{w}$ is the uncertain variable vector; $\mathbf{k}$ is the vector of redispatch

variables vector and compensation variables vector at the second stage, which is composed of $\hat{p}_{g,t}, \hat{p}_{r,t}$, and $\Delta w_{r,t}, \Delta d_{j,t}$, respectively; $\mathbf{c}^T\mathbf{x}$ and $\mathbf{b}^T\mathbf{p}$ is UC cost and unit output cost at first stage, respectively; $\mathbf{d}^T\mathbf{k}$ is the cost of wind curtailment and load shedding responding to WP fluctuations at second stage; $\mathbf{A}$ and $\mathbf{J}$, $\mathbf{B}$ and $\mathbf{N}$ are the coefficient matrices of the first stage UC variables and dispatch variables, respectively; $\mathbf{C}$ and $\mathbf{D}$, $\mathbf{O}$ and $\mathbf{E}$, $\mathbf{R}$ and $\mathbf{F}$ is the coefficient matrix of UC variables, dispatch variables, second stage redispatch variables and compensation variables, respectively; $\mathbf{h}, \mathbf{n}, \mathbf{g}, \mathbf{u}$ are constants; $\Omega(\mathbf{x}, \mathbf{p}, \mathbf{w})$ is the set of second stage constraints.

C. C&CG Solution Method

1) Master Problem: its model is formulated as below.

$$\min_{\mathbf{x},\mathbf{p}} \left(\mathbf{c}^T\mathbf{x} + \mathbf{b}^T\mathbf{p} + \beta\right) \tag{26}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{p} \le \mathbf{h} \tag{27}$$

$$\mathbf{J}\mathbf{x} + \mathbf{N}\mathbf{p} = \mathbf{n} \tag{28}$$

$$\beta \ge \mathbf{d}^T\mathbf{s}_k, \forall k \in \mathcal{K} \tag{29}$$

$$\mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{p} + \mathbf{E}\mathbf{s}_k \le \mathbf{g}, \forall k \in \mathcal{K} \tag{30}$$

$$\mathbf{O}\mathbf{p} + \mathbf{R}\mathbf{s}_k + \mathbf{F}(\mathbf{G}\mathbf{y}_k^* + \mathbf{m}) = \mathbf{u}, \forall k \in \mathcal{K} \tag{31}$$

$$\beta \ge 0, \ \mathbf{x} \text{ binary}, \ \mathbf{p}, \mathbf{k}, \mathbf{w} \ge 0. \tag{32}$$

in which, $\mathcal{K} = \{1, 2, ..., K-1\}$ is the set composed of indices of worst scenarios $\mathbf{y}_1^*, \mathbf{y}_2^*, ..., \mathbf{y}_{K-1}^*$ that have been identified by the subproblem in the previous $K-1$ iterations. The variable $\mathbf{s}_k$ is the compensation variable corresponding to $\mathbf{y}_k^*$. Variable β is an auxiliary variable, represents the cost of wind curtailment and load shedding at the worst scenario. $\mathbf{m}$ is a constant vector.

2) Subproblem: It is to minimize the cost of wind curtailment and load shedding at the worst scenario given the solution $(\mathbf{x}^*, \mathbf{p}^*)$ from the master problem, and is a bilevel optimization problem with a max-min structure. The $\min_{\mathbf{s}\in\Omega(\mathbf{x},\mathbf{p},\mathbf{y})} \mathbf{d}^T\mathbf{s}$ determines the cost of wind curtailment and load shedding for a fixed commitment $\mathbf{x}^*$, first stage dispatch $\mathbf{p}^*$ and wind scenario $\mathbf{y}$, which is then maximized over the uncertainty set U. The form of the subproblem is formulated as:

$$Q(\mathbf{x}^*, \mathbf{p}^*) = \max_{\mathbf{y}\in U} \min_{\mathbf{s}\in\Omega(\mathbf{x},\mathbf{p},\mathbf{y})} \mathbf{d}^T\mathbf{s} \tag{33}$$

$$\mathbf{C}\mathbf{x}^* + \mathbf{D}\mathbf{p}^* + \mathbf{E}\mathbf{s} \le \mathbf{g} \tag{34}$$

$$\mathbf{O}\mathbf{p}^* + \mathbf{R}\mathbf{s} + \mathbf{F}\left(\mathbf{G}\mathbf{y} + \mathbf{m}\right) = \mathbf{u} \tag{35}$$

$$\mathbf{y} \text{ binary}, \ \mathbf{s} \ge 0 \tag{36}$$

According to the dual theory of the linear programming problem, the inner layer optimization problem can be dualized, and (33)-(36) can be transformed into a single layer optimization problem:

$$\begin{aligned} Q(\mathbf{x}^*, \mathbf{p}^*) = \max_{\mathbf{y}\in U,\boldsymbol{\lambda}} \big(&\boldsymbol{\lambda}_1^T \left(\mathbf{C}\mathbf{x}^* + \mathbf{D}\mathbf{p}^* - \mathbf{g}\right) \\ &+ \boldsymbol{\lambda}_2^T \left(\mathbf{u} - \mathbf{O}\mathbf{p}^* - \mathbf{F}\mathbf{m} - \mathbf{F}\mathbf{G}\mathbf{y}\right)\big) \end{aligned} \tag{37}$$

$$\mathbf{E}^T\boldsymbol{\lambda}_1 + \mathbf{R}^T\boldsymbol{\lambda}_2 \le \mathbf{d} \tag{38}$$

$$\boldsymbol{\lambda}_1 \ge 0, \ \boldsymbol{\lambda}_2 \text{ free}, \ \mathbf{y} \text{ binary} \tag{39}$$

among them, $\boldsymbol{\lambda}_1$ is the dual variable of constraint (34), $\boldsymbol{\lambda}_2$ is the dual variable of constraint (35). The dual objective function (37) contains the bilinear term $-\boldsymbol{\lambda}_2^T\mathbf{F}\mathbf{G}\mathbf{y}$, which can be linearized using the big-M method, and (37)-(39) could be turned into a MILP as below.

$$\begin{aligned} Q(\mathbf{x}^*, \mathbf{p}^*) = \max_{\mathbf{y}\in U,\boldsymbol{\lambda}} \big(&\boldsymbol{\lambda}_1^T \left(\mathbf{C}\mathbf{x}^* + \mathbf{D}\mathbf{p}^* - \mathbf{g}\right) \\ &+ \boldsymbol{\lambda}_2^T \left(\mathbf{u} - \mathbf{O}\mathbf{p}^* - \mathbf{F}\mathbf{m}\right)\big) - \mathbf{v} \end{aligned} \tag{40}$$

$$\mathbf{E}^T\boldsymbol{\lambda}_1 + \mathbf{R}^T\boldsymbol{\lambda}_2 \le \mathbf{d} \tag{41}$$

$$\boldsymbol{\lambda}_1 \ge 0, \ \boldsymbol{\lambda}_2 \text{ free}, \ \mathbf{y} \text{ binary} \tag{42}$$

$$-M(\mathbf{1} - \mathbf{y}) \le \boldsymbol{\xi} - {\boldsymbol{\lambda}_2}^T \le M(\mathbf{1} - \mathbf{y}) \tag{43}$$

$$-M\mathbf{y} \le \boldsymbol{\xi} \le M\mathbf{y} \tag{44}$$

$$\mathbf{v} = \boldsymbol{\xi}\mathbf{F}\mathbf{G} \tag{45}$$

in which, $\mathbf{v}, \boldsymbol{\xi}$ are auxiliary variables, M is a large enough positive number, and $\mathbf{1}$ is an all-ones vector of the same type as $\mathbf{y}$.

3) The Proof of the three Propositions and discussion: The general form of PUS or its variants is formulated as:

$$\begin{aligned} U_1 = \Big\{ &w_{r,t} \in R^{NW\times NT} : \\ &w_{r,t} \in \left[\overline{w}_{r,t} - \delta_{r,t}, \overline{w}_{r,t} + \delta_{r,t}\right], \\ &\sum_t \frac{|w_{r,t} - \overline{w}_{r,t}|}{\delta_{r,t}} \le \Gamma_{total} \Big\}. \end{aligned} \tag{46}$$

and the form of MUS is:

$$U_2 = \left\{ w_{r,t} \in R^{NW\times NT} : w_{r,t} = \sum_i \omega_{r,t}^{s_i} y_{r,t}^i \right. \tag{47}$$

$$\sum_i y_{r,t}^i = 1, y_{r,t}^i \in \{0, 1\}, \tag{48}$$

$$\sum_t \left| \frac{\sum_i y_{r,t}^i \left(i - (N+1)/2\right)}{(N-1)/2} \right| \le \Gamma_{total}, \forall r \tag{49}$$

$$y_{r,t}^i + y_{r,t+1}^j \le 1, if \ z_{i,j}^r = 0, i = 1, ..., N, j = 1, .., N \tag{50}$$

$$P\left(X_{r,t}^i\right) \le \mathrm{Pr}_R \tag{51}$$

$$\left. P\left(\omega_{r,t}^{s_1} \le \omega_{r,t} \le \omega_{r,t}^{s_N}\right) \le \mathrm{Pr}_P \right\}. \tag{52}$$

Proposition 1: PUS is a special case of MUS with $N = 3$ (i.e., WP only taking the lower/upper bound and the expected value).

Proof: Due to the solution characteristic of the linear optimization problem, WP in PUS-based robust unit commitment model (PRUCM) could only take one of the three values (i.e., the lower bound, the expected value and the upper bound), and could not be any other one in U_1. Therefore, U_1 could be transferred as:

$$U_1' = \Big\{ w_{r,t} \in R^{NW\times NT} : \tag{53}$$

$$w_{r,t} = \left(\overline{w}_{r,t} - \delta_{r,t}\right) y_{r,t}^1 + \overline{w}_{r,t} y_{r,t}^2 + \left(\overline{w}_{r,t} + \delta_{r,t}\right) y_{r,t}^3, \tag{54}$$

$$\left. y_{r,t}^1, y_{r,t}^2, y_{r,t}^3 \in \{0, 1\}, y_{r,t}^1 + y_{r,t}^2 + y_{r,t}^3 = 1, \right. \tag{55}$$

$$\sum_t \left| \frac{(\overline{w}_{r,t}-\delta_{r,t})\, y^1_{r,t} + \overline{w}_{r,t} y^2_{r,t} + (\overline{w}_{r,t}+\delta_{r,t})\, y^3_{r,t}}{\delta_{r,t}} \right| \le \Gamma_{total} \Big\}. \tag{56}$$

By (55), obtain $y^2_{r,t}$=1 $- y^1_{r,t} - y^3_{r,t}$. U_1' could be written as below by some simple operation:

$$U_1'' = \Big\{ w_{r,t} \in R^{NW\times NT} : \tag{57}$$

$$w_{r,t} = \overline{w}_{r,t} - \delta_{r,t} y^1_{r,t} + \delta_{r,t} y^3_{r,t}, \tag{58}$$

$$y^1_{r,t}, y^3_{r,t} \in \{0,1\}, y^1_{r,t} + y^3_{r,t} \le 1; \tag{59}$$

$$\sum_t \left| y^3_{r,t} - y^1_{r,t} \right| \le \Gamma_{total}, \forall r \Big\}. \tag{60}$$

In MUS, if all state transitions are allowable, the state transition constraints (50)-(51) can be removed. If $\mathrm{Pr}_P = 100\%$ or $P\left(\omega^{s_1}_{r,t} \le \omega_{r,t} \le \omega^{s_N}_{r,t}\right) \le \mathrm{Pr}_P$ is always satisfied, (52) can be removed, too. U_2 can thus be written as:

$$U_2' = \Big\{ w_{r,t} \in R^{NW\times NT} : w_{r,t} = \sum_i \omega^{s_i}_{r,t} y^i_{r,t}, \tag{61}$$

$$\sum_i y^i_{r,t} = 1, y^i_{r,t} \in \{0,1\}, \tag{62}$$

$$\sum_t \left| \frac{\sum_i y^i_{r,t}\,(i-(N+1)/2)}{(N-1)/2} \right| \le \Gamma_{total}, \forall r \Big\}. \tag{63}$$

According to the condition $N = 3$ in the proposition, the lower/upper bound, predicted value of WP and (63) can be written as below, respectively :

$$w^{s_1}_{r,t} = \overline{w}_{r,t} - \delta_{r,t}, w^{s_2}_{r,t} = \overline{w}_{r,t}, w^{s_3}_{r,t} = \overline{w}_{r,t} + \delta_{r,t} \tag{64}$$

$$\sum_t \left| y^3_{r,t} - y^1_{r,t} \right| \le \Gamma_{total}. \tag{65}$$

After replacing (61) by (64) , applying $y^2_{r,t} = 1 - y^1_{r,t} - y^3_{r,t}$ to U_2' and doing some manipulation, U_2' can be turned to:

$$\begin{aligned} U_2'' = \Big\{ & w_{r,t} \in R^{NW\times NT} : \\ & w_{r,t} = \overline{w}_{r,t} - \delta_{r,t} y^1_{r,t} + \delta_{r,t} y^3_{r,t}, \\ & y^1_{r,t}, y^3_{r,t} \in \{0,1\}, y^1_{r,t} + y^3_{r,t} \le 1; \\ & \sum_t \left| y^3_{r,t} - y^1_{r,t} \right| \le \Gamma_{total} \Big\}. \end{aligned} \tag{66}$$

Equation (66) and (57)-(60) are completely same, so Proposition 1 is proved.

Proposition 2: The conservativeness of RUCM based on PUS is larger than that based on MUS if any one of conditions as below holds:

1) $N = 3$, or $\Gamma_{total} \ge 24$, $\Gamma_{total} \in \mathbb{Z}^+$;

2) $\Gamma_{total} < 24$, $\Gamma_{total} \in \mathbb{Z}^+$; wind curtailment or load shedding will not occur if WP takes the predicted value in U_1 or intermediate state value in U_2', in which, U_2' is the U_2 without constraints (50)-(52); $\left| \frac{CR^i_{t,2}}{i-(N+1)/2} \right| \ge \left| \frac{CR^{i+1}_{t,2}}{i+1-(N+1)/2} \right|$ for $1 \le i < (N+1)/2$, in which, $CR^i_{t,2}$ is the larger one between the load shedding cost and the wind curtailment cost for WP at state i in MUS at time period t.

Prove: We first prove the proposition 2 under the first condition. By the dual theory, the two-stage robust optimization model (20)-(25) can be transformed as below:

$$\min_{\mathbf{x},\mathbf{p}} \left(\mathbf{c}^T\mathbf{x} + \mathbf{b}^T\mathbf{p} + f(\mathbf{x},\mathbf{p}) \right) \tag{67}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{p} \le \mathbf{h} \tag{68}$$

$$\mathbf{J}\mathbf{x} + \mathbf{N}\mathbf{p} = \mathbf{n} \tag{69}$$

$$\mathbf{x} \text{ binary, } \mathbf{p} \ge 0. \tag{70}$$

In which,

$$\begin{aligned} f(\mathbf{x},\mathbf{p}) = \max_{\mathbf{y}\in U, \boldsymbol{\lambda}} & \left(\boldsymbol{\lambda}_1^T \left(\mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{p} - \mathbf{g} \right) \right. \\ & \left. + \lambda_2^T \left(\mathbf{u} - \mathbf{O}\mathbf{p} - \mathbf{F}\mathbf{m} \right) \right) - \mathbf{v} \end{aligned} \tag{71}$$

(41)-(45).

For $w \in U_1$, $w \in U_2$ and $w \in U_2'$, the corresponding $f(\mathbf{x},\mathbf{p})$ is written as $f_1(\mathbf{x},\mathbf{p})$, $f_2(\mathbf{x},\mathbf{p})$ and $f_2'(\mathbf{x},\mathbf{p})$ respectively. For the subproblem composed of (71), (41)-(45), according to Proposition 1, U_1 and U_2' with $N = 3$ are completely same and we could deduce that:

$$f_2'(\mathbf{x},\mathbf{p}) = f_1(\mathbf{x},\mathbf{p}). \tag{72}$$

Compared U_2 with U_2', except for the additional constraints (50)-(51) in U_2, both the rest formulas of them are completely same, so U_2 is a subregion of U_2', that is, $U_2 \subseteq U_2'$, thus:

$$f_2(\mathbf{x},\mathbf{p}) \le f_2'(\mathbf{x},\mathbf{p}) = f_1(\mathbf{x},\mathbf{p}). \tag{73}$$

Since $\mathbf{x}$ and $\mathbf{p}$ are arbitrarily taken in their feasible regions, it can deduce that:

$$\min_{\mathbf{x},\mathbf{p}} \left\{ \mathbf{c}^\mathbf{T}\mathbf{x} + \mathbf{b}^\mathbf{T}\mathbf{p} + f_2(\mathbf{x},\mathbf{p}) \right\} \le \min_{\mathbf{x},\mathbf{p}} \left\{ \mathbf{c}^\mathbf{T}\mathbf{x} + \mathbf{b}^\mathbf{T}\mathbf{p} + f_1(\mathbf{x},\mathbf{p}) \right\}. \tag{74}$$

Therefore, for the two-stage robust optimization problem, the solution based on U_2 with $N = 3$ is less conservative than that based on U_1.

At the case of $\Gamma_{total} \ge 24$. The budget constraint (63) in U_2' and the counterpart in U_1 all make optimal WP value possibly take the upper/lower bounds of the uncertainty set in each of 24 periods of a day. The relaxed subproblem composed of (71), (41)-(45) is the linear programming problem, its optimal solution of WP must be obtained on the boundary of the uncertainty set due to the solution characteristic. Therefore, the optimal solution of both $f_1(\mathbf{x},\mathbf{p})$ and $f_2'(\mathbf{x},\mathbf{p})$ can only take the lower/upper bound, but can't be any other one within U_1 and U_2', so the optimal solution of $f_1(\mathbf{x},\mathbf{p})$ and $f_2'(\mathbf{x},\mathbf{p})$ is exactly the same. That is:

$$f_2'(\mathbf{x},\mathbf{p}) = f_1(\mathbf{x},\mathbf{p}). \tag{75}$$

Compared U_2 with U_2', except for the additional constraints (50)-(51), both the rest formulas of them are completely same, so U_2 is a subregion of U_2', that is, $U_2 \subseteq U_2'$, thus:

$$f_2(\mathbf{x},\mathbf{p}) \le f_2'(\mathbf{x},\mathbf{p}) = f_1(\mathbf{x},\mathbf{p}). \tag{76}$$

Since $\mathbf{x}$ and $\mathbf{p}$ are arbitrarily taken in their feasible regions, it can deduce that:

$$\min_{\mathbf{x},\mathbf{p}} \left\{ \mathbf{c}^\mathbf{T}\mathbf{x} + \mathbf{b}^\mathbf{T}\mathbf{p} + f_2(\mathbf{x},\mathbf{p}) \right\} \le \min_{\mathbf{x},\mathbf{p}} \left\{ \mathbf{c}^\mathbf{T}\mathbf{x} + \mathbf{b}^\mathbf{T}\mathbf{p} + f_1(\mathbf{x},\mathbf{p}) \right\}. \tag{77}$$

Therefore, for the two-stage robust optimization problem, the solution based on U_2 is less conservative than that based on U_1 when $\Gamma_{total} \geq 24$.

Combining the conclusion of $N = 3$ with that of $\Gamma_{total} \geq 24$ above, Proposition 2 is proved under the first condition.

We further prove the proposition under the second condition. some necessary definitions and notations are presented firstly for the convenience of description.

1. For the two-stage robust optimization problem, the optimal solution of (71) based on U_1 is denoted as $(\mathbf{x}^*, \mathbf{p}^*)$, which is a feasible solution of U_2 and U_2'.

2. $w_{r,t}^1, w_{r,t}^2, w_{r,t}^3$ represents the lower, expectation and the upper value of WP in U_1, respectively. For the multi-state values $\left\{\omega_{r,t}^1, \omega_{r,t}^2, ..., \omega_{r,t}^N\right\}$ in U_2', $\omega_{r,t}^1$ and $\omega_{r,t}^N$ also represent the lower bound and upper bound, respectively, and thus $\omega_{r,t}^1 = w_{r,t}^1$, and $\omega_{r,t}^N = w_{r,t}^3$.

3. At time period t, the system upward and downward reserve shortage based on U_1(denoted as a_1, a_2) and U_2' (denoted as a_3, a_4) at $(\mathbf{x}^*, \mathbf{p}^*)$ are respectively attained by

$$\begin{cases} a_1 = D_t - \left(\sum\nolimits_g p_{g,t}^* + R_t^{res-up*}\right) - w_{r,t}^1 \\ a_2 = \left(\sum\nolimits_g p_{g,t}^* - R_t^{res-down*}\right) - w_{r,t}^3 - D_t \end{cases} \tag{78}$$

$$\begin{cases} a_3 = D_t - \left(\sum\nolimits_g p_{g,t}^* + R_t^{res-up*}\right) - \omega_{r,t}^i \\ a_4 = \left(\sum\nolimits_g p_{g,t}^* - R_t^{res-down*}\right) - \omega_{r,t}^{N-i+1} - D_t \\ 1 \leq i < (N+1)/2 \end{cases} \tag{79}$$

in which, D_t, $R_t^{res-up*}$ and $R_t^{res-down*}$ are the total load requirement, the total upward ramp reserve, and the total downward ramp reserve at time period t, respectively. Accordingly, the cost of them, that is, denoted as a_{1_c} and a_{2_c} for the two cost based on U_1, a_{3_c} and a_{4_c} for the two cost based on U_2', are respectively attained by

$$\begin{cases} a_{1\_c} = C^{LS}\left(D_t - \left(\sum\nolimits_g p_{g,t}^* + R_t^{res-up*}\right) - w_{r,t}^1\right) \\ a_{2\_c} = C^{WS}\left(\left(\sum\nolimits_g p_{g,t}^* - R_t^{res-down*}\right) - w_{r,t}^3 - D_t\right) \end{cases} \tag{80}$$

$$\begin{cases} a_{3\_c} = C^{LS}\left(D_t - \left(\sum\nolimits_g p_{g,t}^* + R_t^{res-up*}\right) - \omega_{r,t}^i\right) \\ a_{4\_c} = C^{WS}\left(\left(\sum\nolimits_g p_{g,t}^* - R_t^{res-down*}\right) - \omega_{r,t}^{N-i+1} - D_t\right) \\ 1 \leq i < (N+1)/2 \end{cases} \tag{81}$$

in which, C^{LS}, C^{WS} represents the cost of unit load shedding, unit wind curtailment, respectively. Note that for the system based on U_1 or U_2' while $1 \leq i < (N+1)/2$, a_1 and a_2, a_3 and a_4, are respectively counterpart cost at time period t.

4. The larger of a_{1_c} and a_{2_c} at t is denoted as $CR_{t,1}$, which forms the series $\left\{CR_{1,1}, CR_{2,1}, ..., CR_{24,1}\right\}$, for 24 time periods of a day.

5. The larger of a_{3_c} and a_{4_c} at t while $0 < i < (N+1)/2$ is denoted as $CR_{t,2}^i$, which forms the series $\left\{CR_{t,2}^1, CR_{t,2}^2, ..., CR_{t,2}^i\right\}$, and their corresponding WP state deviation relative to the intermediate state is $\left\{1, \left|\frac{2-(N+1)/2}{(N-1)/2}\right|, ..., \left|\frac{i-(N+1)/2}{(N-1)/2}\right|\right\}$.

Since $\Gamma_{total} < 24$, $\Gamma_{total} \in \mathbb{Z}^+$, we sort the elements in $\left\{CR_{1,1}, CR_{2,1}, ..., CR_{24,1}\right\}$ in descending order and select the top Γ_{total} to be a series, which is rewritten as $\left\{CR_{k_1,1}, CR_{k_2,1}, ..., CR_{k_{\Gamma_{total}},1}\right\}$. Then the WP at the worst scenario of the second stage optimization problem of RUCM based on U_1 only takes the either upper bound or the lower bound at time $k_1, ..., k_{\Gamma_{total}}$, and the predicted values at time $k_{\Gamma_{total}+1}, .., k_{24}$, due to the budget constraint and the condition 2) of the proposition. Therefore, the maximum value of the second stage optimization problem of RUCM based on U_1 is calculated as:

$$f_1(\mathbf{x}^*, \mathbf{p}^*) = \sum_{j=1}^{\Gamma_{total}} CR_{k_j,1}. \tag{82}$$

For the second stage maximum problem of RUCM based on U_2', the largest of wind curtailment and load shedding costs among $1 \leq i \leq N$ of the system at each time period t is denoted as $CR_{t,2}^1$ and forms series $\left\{CR_{1,2}^1, CR_{2,2}^1, ..., CR_{24,2}^1\right\}$ for 24 time periods of a day. $\left\{CR_{1,2}^1, CR_{2,2}^1, ..., CR_{24,2}^1\right\}$ and $\left\{CR_{1,1}, CR_{2,1}, ..., CR_{24,1}\right\}$ is identical due to the definitions and notation 2 and the condition 2) of the proposition. Sort the elements in $\left\{CR_{1,2}^1, CR_{2,2}^1, ..., CR_{24,2}^1\right\}$ in descending order and select the top Γ_{total} to be a series, which is rewritten as $\left\{CR_{k_1',2}^1, CR_{k_2',2}^1, ..., CR_{k_{\Gamma_{total}}',2}^1\right\}$, and each element in $\left\{CR_{k_{\Gamma_{total}+1}',2}^1 CR_{k_{\Gamma_{total}+2}',2}^1 ..., CR_{k_{24}',2}^1\right\}$ is 0 due to budget constraint. Obviously,

$$\sum_{j=1}^{\Gamma_{total}} CR_{k_j,1} = \sum_{z=1}^{\Gamma_{total}} CR_{k_z',2}^1. \tag{83}$$

Moreover, the WP value corresponding to each $CR_{k_j,1}$ and $CR_{k_z',2}^1$ gained by second stage maximum problem of RUCM based on U_1, U_2', respectively, is exactly same due to the definition and notation 2. Thus, the maximum value and the solution of the second stage maximum problem of RUCM based on U_1 is a candidate maximum value and a feasible solution of that based on U_2', respectively. Adopt the following steps to explore whether exist or not WP worse scenarios in U_2' at which the subproblem attains larger optimal value than that whose corresponding series composed of 24 elements $\left\{CR_{k_1',2}^1, CR_{k_2',2}^1, ..., CR_{k_{\Gamma_{total}}',2}^1, 0, ..., 0\right\}$:

1) Select an element $CR_{k_z',2}^1, z = 1, ..., \Gamma_{total}$ from the series $\left\{CR_{k_1',2}^1, CR_{k_2',2}^1, ..., CR_{k_{\Gamma_{total}}',2}^1, 0, ..., 0\right\}$, and exchange it with an element $CR_{k_z',2}^{j_0}$ in series $\left\{CR_{k_z',2}^{j_0}\right\}(1 < j_0 < (N+1)/2\ )$ to obtain a new series $\left\{CR_{k_1',2}^1, CR_{k_2',2}^1, ..., CR_{k_z',2}^{j_0}, ..., CR_{k_{\Gamma_{total}}',2}^1, 0, ..., 0\right\}$. The sum of the deviation of each WP state corresponding to $CR_{k_z',2}^1$ in the new series from the intermediate (abbreviated as total budget hereafter for the convenience of description) is $\Gamma_{total} - (1 - |2\,(j_0 - (N+1)/2\ )/(N-1)\ |)$.

2) In the new series, there are total $24 - \Gamma_{total}$ of 0 elements. Choose M of them, each index of which is denoted as $i_m m = 1, ..., M, 1 \leq M \leq 24 - \Gamma_{total}$, $\Gamma_{total} + 1 \leq i_m \leq 24$

and $i_m \in \mathbb{Z}^+$. Replace each of the M of 0 elements in the series $\left\{CR^1_{k'_1,2}, CR^1_{k'_2,2}, ..., CR^{j_0}_{k'_z,2}, ..., CR^1_{k'_{\Gamma_{total}},2}, 0, ..., 0\right\}$ with the corresponding element $CR^{j_m}_{k'_{i_m},2}$ in the series $\left\{CR^{j_m}_{k'_{i_m},2}\right\}(1 < j_m < (N+1)/2)$, to obtain the series as: $\{CR^1_{k'_1,2}CR^1_{k'_2,2}, ..., CR^{j_0}_{k'_z,2}, ..., CR^1_{k'_{\Gamma_{total}},2}, 0, ..., CR^{j_1}_{k'_{i_1},2}$ $, CR^{j_2}_{k'_{i_2},2}, ..., CR^{j_M}_{k'_{i_M},2}, ..., 0\}$, whose total budget is equal to:

$$\Gamma_{total} - \left(1 - \sum_{i=0}^{M} \left| \frac{j_i - (N+1)/2}{(N-1)/2} \right| \right). \tag{84}$$

Repeat step2) to make the second part in (84) as close as possible to 0. The obtained closest series is rewritten as:

$$\{CR^1_{k'_1,2}CR^1_{k'_2,2}, ..., CR^{j_0}_{k'_z,2}, ..., CR^1_{k'_{\Gamma_{total}},2}, 0, ..., CR^{j_1}_{k'_{i_1},2}, CR^{j_2}_{k'_{i_2},2}, ..., CR^{j_M}_{k'_{i_M},2}, ..., 0\}. \tag{85}$$

3) Since the total cost of the wind curtailment and load shedding among the 24 time periods of a day, which is the object of second stage optimization problem of RUCM, increases gradually as the total budget of the series corresponding to WP scenario approaches to Γ_{total}, it thus achieves maximum when the series is (85), and the corresponding WP scenario is the worst. At the meantime, the formula as below is obtained.

$$\sum_{i=0}^{M} \left| \frac{j_i - (N+1)/2}{(N-1)/2} \right| \leq 1. \tag{86}$$

According to condition 2), we can derive:

$$\left| \frac{CR^1_{t,2}}{(1-(N+1)/2\,)\,CR^2_{t,2}} \right| \geq \left| \frac{1}{2-(N+1)/2} \right| \tag{87}$$

$$\left| \frac{CR^2_{t,2}}{(2-(N+1)/2\,)\,CR^3_{t,2}} \right| \geq \left| \frac{1}{3-(N+1)/2} \right|, t = 1, ..., 24 \tag{88}$$

From (88) we can obtain:

$$|3-(N+1)/2\,| \geq \left| \frac{(2-(N+1)/2\,)\,CR^3_{t,2}}{CR^2_{t,2}} \right| \tag{89}$$

After the left and right side of (87) multiplies the left and right side of (89), respectively, we can get:

$$\left| \frac{CR^1_{t,2}\,(3-(N+1)/2\,)}{(2-(N+1)/2\,)\,CR^2_{t,2}} \right| \geq \left| \frac{(1-(N+1)/2\,)\,CR^3_{t,2}}{CR^2_{t,2}\,(2-(N+1)/2\,)} \right| \tag{90}$$

Equation (90) is further simplified as:

$$\left| \frac{CR^1_{t,2}}{CR^3_{t,2}\,(1-(N+1)/2\,)} \right| \geq \left| \frac{1}{(3-(N+1)/2\,)} \right| \tag{91}$$

By the same process as (89) and (90), we can obtain the relationship between $CR^1_{t,2}$ and $CR^4_{t,2}$ is as:

$$\left| \frac{CR^1_{t,2}}{CR^4_{t,2}\,(1-(N+1)/2\,)} \right| \geq \left| \frac{1}{(4-(N+1)/2\,)} \right| \tag{92}$$

According to (87), (88) and (91), a recursive formula could be deduced as:

$$\left| \frac{CR^1_{t,2}}{CR^{j_m}_{t,2}\,(1-(N+1)/2\,)} \right| \geq \left| \frac{1}{(j_m-(N+1)/2\,)} \right| \tag{93}$$

$$\left| \frac{CR^{j_m}_{t,2}\,(1-(N+1)/2\,)}{CR^1_{t,2}(N-1)/2} \right| \leq \left| \frac{(j_m-(N+1)/2\,)}{(N-1)/2} \right| \tag{94}$$

in which, $0 \leq m \leq M$. After substituting (94) into (86), and replace t by k'_z and $k_{i'_m}$, it can be derived as:

$$\left| \frac{(1-(N+1)/2\,)\,CR^{j_0}_{k'_z,2}}{CR^1_{k'_z,2}\,(N-1)/2} \right| + \sum_{m=1}^{M} \left| \frac{(1-(N+1)/2\,)\,CR^{j_m}_{k'_{i_m},2}}{CR^1_{k'_{i_m},2}\,(N-1)/2} \right| \leq 1 \tag{95}$$

$$\left| \frac{CR^{j_0}_{k'_z,2}}{CR^1_{k'_z,2}} \right| + \sum_{m=1}^{M} \left| \frac{CR^{j_m}_{k'_{i_m},2}}{CR^1_{k'_{i_m},2}} \right| \leq 1 \tag{96}$$

$$CR^{j_0}_{k'_z,2} + CR^{j_1}_{k'_{i_1},2}\frac{CR^1_{k'_z,2}}{CR^1_{k'_{i_1},2}} + ... + CR^{j_M}_{k'_{i_M},2}\frac{CR^1_{k'_z,2}}{CR^1_{k'_{i_M},2}} \leq CR^1_{k'_z,2}. \tag{97}$$

Since the series $\left\{CR^1_{k'_1,2}, CR^1_{k'_2,2}, ..., CR^{j_0}_{k'_z,2}, ..., CR^1_{k'_{\Gamma_{total}},2}\right\}$ is composed of top Γ_{total} of $\left\{CR^1_{1,2}, CR^1_{2,2}, ..., CR^1_{24,2}\right\}$ and listed in descending order, in addition, there are conditions below:

$$\begin{cases} k'_z \in k'_1, k'_2, ..., k'_{\Gamma_{total}} \\ k'_{i_m} \in k'_{\Gamma_{total+1}}, k'_{\Gamma_{total+2}}, ..., k'_{\Gamma_{total+24}} (1 \leq m \leq M) \end{cases} \tag{98}$$

It thus can derive:

$$CR^1_{k'_z,2}/CR^1_{k'_{i_m},2} > 1. \tag{99}$$

Therefore, (97) could be further turned to:

$$CR^{j_0}_{k'_z,2} + CR^{j_1}_{k'_{i_1},2} + ... + CR^{j_M}_{k'_{i_M},2} \leq CR^1_{k'_z,2}. \tag{100}$$

Due to that k'_z, $k'_{i_m}(1 \leq m \leq M)$ is arbitrarily chosen from series $\{k'_1, k'_2, ..., k'_{\Gamma_{total}}\}$, $\{k'_{\Gamma_{total}+1}, k'_{\Gamma_{total}+2}, ..., k'_{\Gamma_{total}+24}\}$, respectively, (100) means we cannot find a worse scenario than $\left\{CR^1_{k'_1,2}, CR^1_{k'_2,2}, ..., CR^1_{k'_{\Gamma_{total}},2}, 0, ..., 0\right\}$ in U'_2. And thus, the optimal value of second stage optimization problem under U'_2 is $f'_2(\mathbf{x}^*, \mathbf{p}^*) = f_1(\mathbf{x}^*, \mathbf{p}^*) = \sum_{j=1}^{\Gamma_{total}} CR_{k_j,1}$ $= \sum_{i=1}^{\Gamma_{total}} CR^1_{k'_i,2}$. Therefore,

$$\mathbf{c}^T\mathbf{x}^* + \mathbf{b}^T\mathbf{p}^* + f_1(\mathbf{x}^*, \mathbf{p}^*) = \mathbf{c}^T\mathbf{x}^* + \mathbf{b}^T\mathbf{p}^* + f'_2(\mathbf{x}^*, \mathbf{p}^*) \tag{101}$$

Compared U_2 with U'_2, except for the additional constraints (50)-(51) in U_2, both the rest part of them are completely same, so U_2 is a subregion of U'_2, that is, $U_2 \subseteq U'_2$, thus:

$$f_2(\mathbf{x}^*, \mathbf{p}^*) \leq f'_2(\mathbf{x}^*, \mathbf{p}^*) = f_1(\mathbf{x}^*, \mathbf{p}^*). \tag{102}$$

Therefore, it can deduce that:

$$\min_{\mathbf{x},\mathbf{p}} \left\{\mathbf{c}^T\mathbf{x} + \mathbf{b}^T\mathbf{p} + f'_2(\mathbf{x}, \mathbf{p})\right\} \leq \min_{\mathbf{x},\mathbf{p}} \left\{\mathbf{c}^T\mathbf{x} + \mathbf{b}^T\mathbf{p} + f_1(\mathbf{x}, \mathbf{p})\right\} \tag{103}$$

And the proposition is thus proved under the second condition. Combining the proof under the first and second condition as above, proposition 2 is proved.

Discussion:explanation about the reasonableness of the two conditions in proposition 2:
1)Among the first condition, $N = 3$ and $\Gamma_{total} \geq 24$ are the pre-condition, $\Gamma_{total} \in \mathbb{Z}^+$ is the setting of the paper and is widely accepted condition. Therefore, the first condition is reasonable.
2) Since the second condition is composed of three sub-conditions, we explain its reasonability one by one as below:
a) The first sub-condition: $\Gamma_{total} < 24$ is the pre-condition, and $\Gamma_{total} \in \mathbb{Z}^+$ has been explained at 1), so it is reasonable.
b) The second sub-condition: Since the first stage of the two-stage RUCM is to schedule the states of start-up or shut-down of the units and the generation output with WP at predicted value (or base scenario), and the second stage is to find minimum cost of the wind curtailment and load shedding at the worst scenarios, the base scenario basically does not result in wind curtailment or load shedding, so it is valid.
c) The third sub-condition: When the difference of WP values between adjacent states equals or gradually increases with the state deviating farther from intermediate state, this sub-condition is automatically satisfied. The distribution of WP is a conditional probability distribution parameterized by predicted values and is approximately right-skewed or left-skewed for the small or large predicted value, respectively. This characteristic leads to that under the condition of all WP states being obtained by uniform division of probability range $[0, 1]$, and its values being approximated with the quantile of the CPs of the states, both of which are the strategy adopted at Section III in the paper, the difference of WP values between adjacent states gradually increase while the state deviating away from the intermediate state. Therefore, this sub-condition will be also satisfied.

Based on the all above explanation, the two conditions of proposition 2 are reasonable and valid.

D. Shifted Legendre Polynomial

The N order shifted Legendre polynomial is defined as below(https://mathworld.wolfram.com/LegendrePolynomial.html):

$$\widetilde{P_n}(x) = (-1)^n \sum_{k=0}^{n} \binom{n}{k} \binom{n+k}{k} (-x)^k \tag{104}$$

The specific expressions of 0-4 order shifted Legendre polynomial are listed at Table I.

TABLE I
5TH ORDER SHIFT LEGENDRE POLYNOMIAL

n	$\widetilde{P_n}(x)$
0	1
1	$2x - 1$
2	$6x^2 - 6x + 1$
3	$20x^3 - 30x^2 + 12x - 1$
4	$70x^4 - 140x^3 + 90x^2 - 20x + 1$

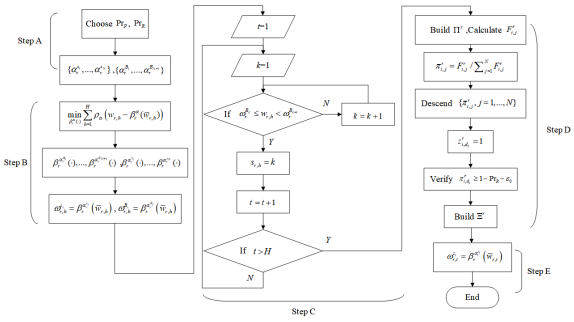

Fig. 1. Flow chart of MUS generation

E. Flow Chart of MUS Generation

F. The Expressions of Convex Hull method Based on Disjunction with SOS1 and Its Unification with the Concise Expression of big-M

The objective function of the subproblem (37) could be rewritten as:

$$min\left(R + \sum_t \sum_r w_{r,t}\lambda_{r,t}\right) = \min\left(R + \sum_t \sum_r \sum_i \omega^i_{r,t} y^i_{r,t} \lambda_{r,t}\right). \tag{105}$$

in which, R represents the rest part; the second term is the so-called bilinear term, $\lambda_{r,t}$ is the dual variable, and $y^i_{r,t}$ is a 0-1 variable and subject to the SOS1(constraint (48), note that is constraint (3) in the paper). After introducing the auxiliary variable $\phi_{r,t} = \sum_i \omega^i_{r,t} y^i_{r,t} \lambda_{11,r,t}$, (105) with constraint (48) can be expressed as a disjunctive form according to disjunctive programming theory as below.

$$min\left(R + \sum_t \sum_r \phi_{r,t}\right) \tag{106}$$

$$\begin{bmatrix} y^1_{r,t} = 1 \\ \phi_{r,t} = \omega^1_{r,t}\lambda_{r,t} \end{bmatrix} \underset{-}{\vee} \ldots \underset{-}{\vee} \begin{bmatrix} y^B_{r,t} = 1 \\ \phi_{r,t} = \omega^B_{r,t}\lambda_{r,t} \end{bmatrix}, \ \forall(r,t). \tag{107}$$

In which, $\underset{-}{\vee}$ is the symbol of XOR, N represents the total number of WP states, and $y^i_{r,t}$ is a 0-1 variable. Constraint (107) means that at any time, the WP of the wind farm r can only be at one state, that is, only one of the disjunctions in (107) holds. After introducing auxiliary variable $v^i_{r,t}, u^i_{r,t}, i = 1, ..., N$, the (107) can be transformed into a linear convex hull representation, that is,

$$\phi_{r,t} = \sum_i v^i_{r,t} \tag{108}$$

$$\lambda_{11,r,t} = \sum_i u^i_{r,t} \tag{109}$$

$$v^i_{r,t} = \omega^i_{r,t} u^i_{r,t} \tag{110}$$

$$-y^i_{r,t} M \leq u^i_{r,t} \leq y^i_{r,t} M \tag{111}$$

$$\sum_i y^i_{r,t} = 1. \tag{112}$$

After substituting (110) into (108), it can be turned to:

$$\phi_{r,t} = \sum_i \omega_{r,t}^i u_{r,t}^i \tag{113}$$

$$\lambda_{11,r,t} = \sum_i u_{r,t}^i \tag{114}$$

$$-y_{r,t}^i M \le u_{r,t}^i \le y_{r,t}^i M \tag{115}$$

$$\sum_i y_{r,t}^i = 1. \tag{116}$$

Equation (113)-(116) is completely the same as those of big-M method with SOS1 presented at the paper, indicating that both method is equivalent and unified at this situation.

G. Generation Method of Intra-day Scenarios

Since the probability distribution of WP is generally asymmetric, exists no analytical form and is a conditional probability distribution (CPD) with the predicted value as a parameter, the intra-day scenarios in the paper are produced according to the historical data without making any assumptions about the form of distribution, while considering asymmetry and CPD. The specific processes are as follows.

For the WP predicted value of wind farm r at time t–$\overline{w}_{r,t}$, select the data from the two-year historical actual WPs, whose corresponding predicted values fall within $\left[\overline{w}_{r,t} - \tau W_r^{\max}, \overline{w}_{r,t} + \tau W_r^{\max}\right]$(where τ is a small positive number and $W_r^{\max}$ is the installed capacity of the wind farm) ,to form sub-set, which is called D1 and its the total number is N_0. Randomly draw one data from D1 as the actual WP of the wind farm r at time t. The above sampling method reflects conditional sampling using the predicted values as parameters. For $t = 1, ..., NT, r = 1, ..., NW$, repeat the process as above to generate a scenario $\{\widetilde{w}_{r,t}\}_{t=1,...,NT,r=1,...,NW}$ of the WP with the given $\overline{w}_{r,t}, t = 1, ..., NT, r = 1, ..., NW$. After repeating the above process N_S(the number of scenarios to be generated) times, the N_S scenarios of WP of all wind farms in all time periods are produced.

Fig.2 presents the scatter plot of two years historical data of WP predicted and actual values of No.6389 wind farm. Fig.3 shows the part of the scatter plot at $\overline{w}_{r,t} = 0.7$p.u. and $\tau = 0.02$. The actual WP drawn from this part reflects that actual WP follows CPD parameterized by $\overline{w}_{r,t} = 0.7$p.u..

H. Verification of State Transition Matrices without Considering the Difference of Time Periods

In order to verify the state transition matrix Π^r gained by treating all the transition events happening at different period to be same only if initial states and the target states all being identical, respectively, we further calculate the state transition matrix for each time period, $\Pi^{r,i}(i = 1, 2, ..., 24)$. After obtaining the historical WP state series at step.C in Sec.II in the paper, $\Pi^{r,i}(i \in \{1, 2, ..., 24\})$ is attained by only counting the state transition events whose initial states are at time period i, and is shown at. Fig.4. The x-axis is the state at time t, the y-axis is the state at time $t+1$, and the z-axis is the time period. The orange and blue point (x, y, z) represent

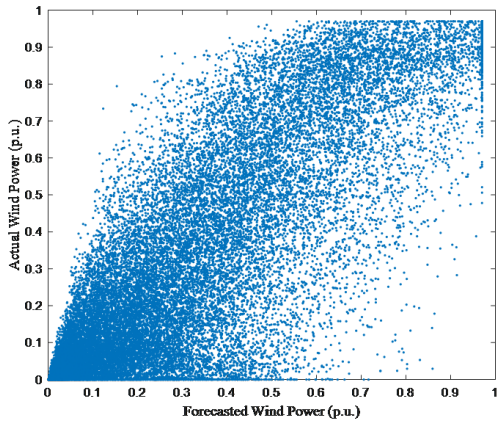

Fig. 2. Scatter plot of two-year historical data of WP predicted values and actual values

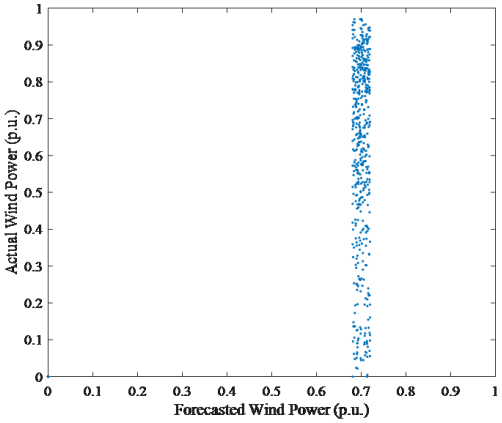

Fig. 3. Part of the scatter plot of historical data with WP predicted values between [0.68p.u., 0.72p.u.]

that the transition from state x to state y is allowable and not allowable at time period z, respectively. Fig.4 shows the color of the points in the z-axis direction only very mildly changes, indicating that the general statuses of state transitions of different time periods are very similar. Only the transitions between states 2 and 4, state 7 and state 5 at different time periods have relative obvious differences, which just accounts for 6% among the total transition events and is resulted from the small statistic data after dividing the 2-year total historical data into 24 sub-sets, the rest of the state transitions are almost identical.

In order to further illustrate the above observations, we expand each of the 24 transition matrices by rows into a series, that is, for $\Pi^{r,i}(1 \le i \le 24)$ being a 7×7 matrix, let $\Pi^{r,i,k}(1 \le k \le 7)$ be the vector of the kth row, the ith series is formed as $\left\{\Pi^{r,i,1}\Pi^{r,i,2}, ..., \Pi^{r,i,7}\right\}$, and then total 24 series are attained, accordingly. Fig.5 displays the Pearson correlation coefficients between any two of the 24 series. The horizontal and vertical axes represent time periods, and the square (i, j) represents the Pearson correlation coefficient between the ith series and jth series. The mean value, variance and the minimum of the Pearson correlation coefficient of the

24 series is 0.89, 0.004, 0.71, respectively. The mean value is close to 0.9 and the variance 0.004 is very small, indicating that there is a strong linear positive correlation between any two of the series. This further verifies that there is very slight difference between the different xOy cut plane matrices(that is, the state transition matrices at different periods) in Fig.4, indicating that the general fluctuation statuses of WP in a wind farm has little correlation with the time period. Therefore, it is valid and effective to build the state transition matrix based on the statistics of the whole time periods without distinguishing them at the paper.

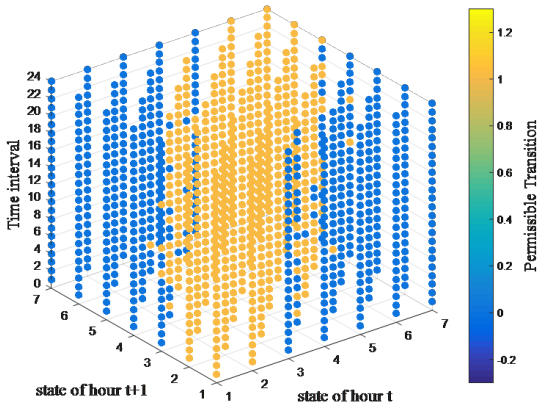

Fig. 4. permissible state transitions in different time periods

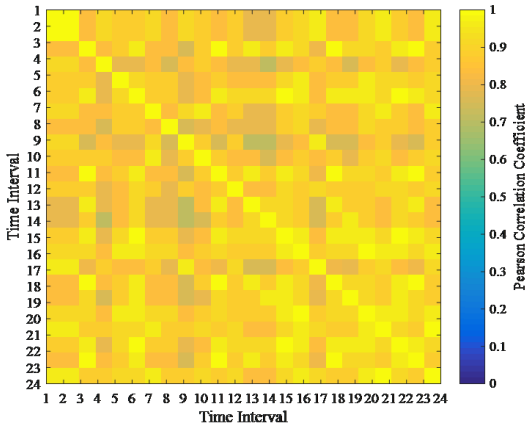

Fig. 5. Pearson correlation coefficient of permissible state transitions in different time periods

I. MUS of No.3902 and No.4562 Wind Farms

Let Pr_P and Pr_R all be 0.9, and N be 7. All the parameters in MUSs of No.3902 and No.4562 wind farms, which is denoted as wind farm 2, 3, respectively, are calculated and shown as follows:

1) Quantile functions calculation: Normalize the predicted and actual values of the historical onshore wind farm data No.3902 and No.4562 in Eastern Wind Dataset to range [0,1], the scatter plots of which are shown as the blue points in Fig.6 and Fig.7, respectively. The two figures show that the distribution relationship between the actual value and the predicted values of the two wind farms is similar as that of No.6389 wind farm that is explained in detail at the paper.

Employing the method in section III of the paper, the series of the quantile functions of the CPs of all power states and its boundaries for the two wind farms are obtained and listed as $\left\{\beta_2^{\alpha_2^{S_1}}(\cdot),\beta_2^{\alpha_2^{B_1}}(\cdot)...,\beta_2^{\alpha_2^{S_7}}(\cdot),\beta_2^{\alpha_2^{B_7}}(\cdot)\right\}$, $\left\{\beta_3^{\alpha_3^{S_1}}(\cdot),\beta_3^{\alpha_3^{B_1}}(\cdot)...,\beta_3^{\alpha_3^{S_7}}(\cdot),\beta_3^{\alpha_3^{B_7}}(\cdot)\right\}$, respectively, and the related curves under different predicted values are also presented in Fig.6 and Fig.7, respectively. In each of the both figures, the red curves from top to bottom are the quantile value curves of CPs of power state from 7 to 1 in order, respectively, and the corresponding α are 0.95, 0.80, 0.65, 0.5, 0.35, 0.20 and 0.05. The blue curves from top to bottom (some overlapping with red lines) are the quantile value curves of CPs of power state interval boundaries of power state from 7 to 1 in order, and the corresponding α are 0.95, 0.80, 0.65, 0.55, 0.45, 0.35, 0.20, 0.05.

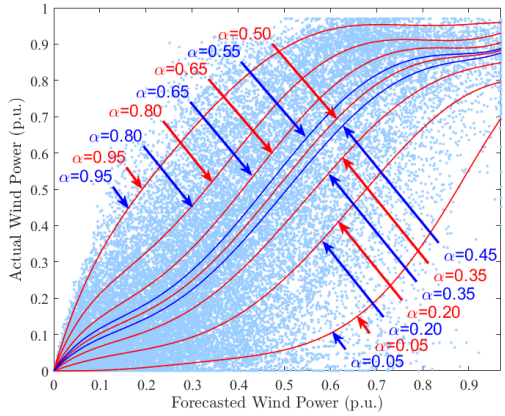

Fig. 6. Scatter plot of WP predicted and actual values of wind farm 2 (No.3902) and the curves of quantile function values

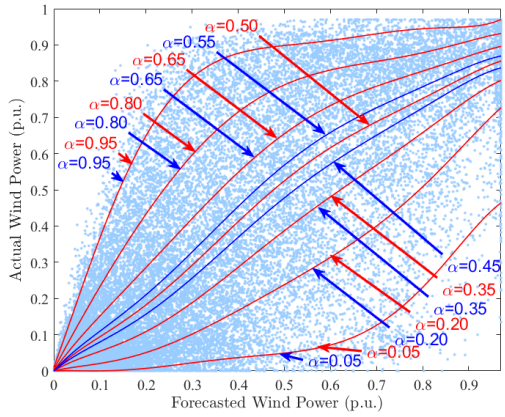

Fig. 7. Scatter plot of WP predicted and actual values of wind farm 3 (No.4562) and the curves of quantile function values

2) State transition probability matrix: The state transition probability matrices of No.3902 wind farm and No.4562 wind farm are shown at Fig.8 and Fig.9, respectively.

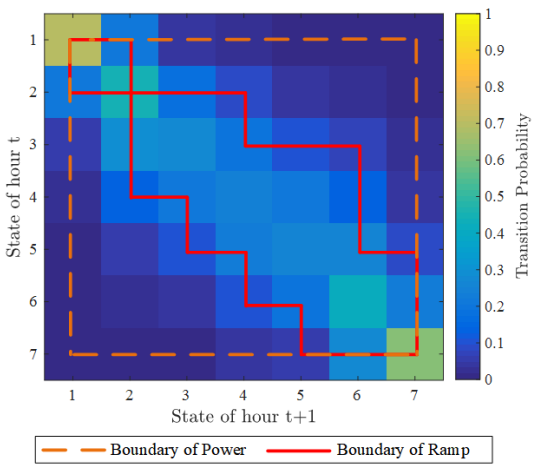

Fig. 8. State transition probability matrix of No.3902 wind farm Π^2

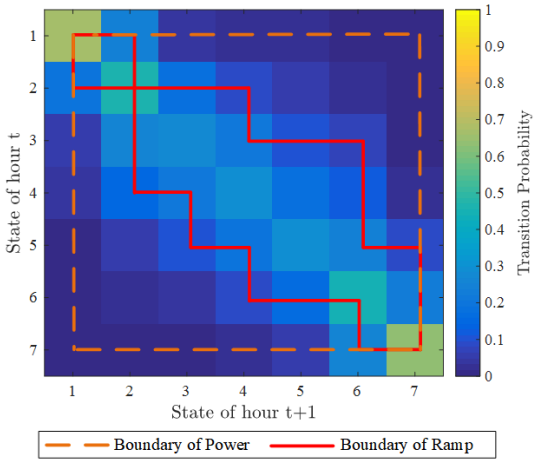

Fig. 9. State transition probability matrix of No.4562 wind farm Π^3

In Fig.8 and Fig.9, the yellow dashed box includes the state transitions without constraint (50)-(51), and the red solid box includes the permissible state transitions determined according to Pr_R. The figures show that from any state at time t, the transition probability to the same state or the adjacent at time $t+1$ is much larger than those to the rest. For wind farm No.3902, the permissible transitions by the transition matrix are state 1→1-2, state2→1-4, state 3→2-6, state 4→2-6, and state 5→3-7, state 6→4-7, state 7→5-7. For wind farm No.4562, the permissible transitions by the transition matrix are state 1→1-2, state2→1-4, state 3→2-6, state 4→2-6, and state 5→3-7, state 6→4-7, state 7→6-7. The largest state difference between the initial state and the target state among all the allowable state transitions in each of $\Pi^r(r=2,3)$ is all 3.

Let the state transition indicator $z^r_{i,j}(i,j=1,2,...,7,;r=1,2,3)$ equal 1 or 0, if its (i,j) is within or outside the red line box in Π^r, respectively. Then the permissible state transition matrix $\Xi^r(r=1,2,3)$ is attained.

3) Building of Next-day MUS: Substitute the hourly predicted values of WP at January 1, 2006 into $\beta_r^{\alpha_r^{s_1}}(\cdot),...,\beta_r^{\alpha_r^{s_7}}(\cdot)\,(r=2,3)$ to obtain the WP state values $\omega^{s_i}_{r,t}(i{=}1,2,...,7;t=1,2,...,24;r=2,3)$, the related curves are presented at Fig.10 and Fig.11. The red curves of WP state values from top to bottom in the both figures represent $\omega^{s_7}_{r,t},\omega^{s_6}_{r,t},...,\omega^{s_1}_{r,t}(t=1,2,...,24)$, respectively. The black curves and the cyan curves in the both figures represent the predicted value and the actual value of WP, respectively. The cyan curves in the two figures shows that the actual changes of the WP among the adjacent time periods is mostly among the neighborhood of current states and its largest range is no more than 2 states, which is smaller than that of allowable largest state difference between the initial and the target in Π^r and $\Xi^r(r=2,3)$. This demonstrates the reasonableness of MUSs. Up to now, the allowable state transitions and the WP state values are all attained, and thus the MUSs of wind farm 2 and 3 are finally built.

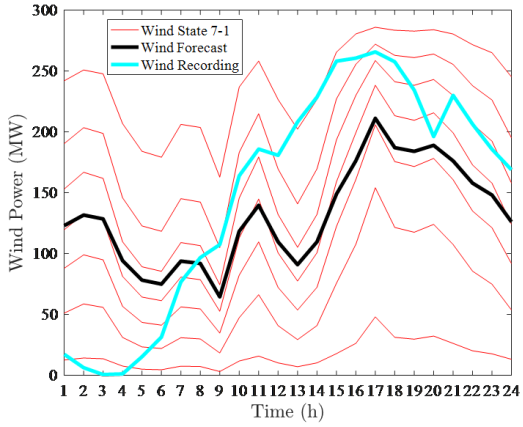

Fig. 10. The state values of WP of wind farm 2 on January 1, 2006

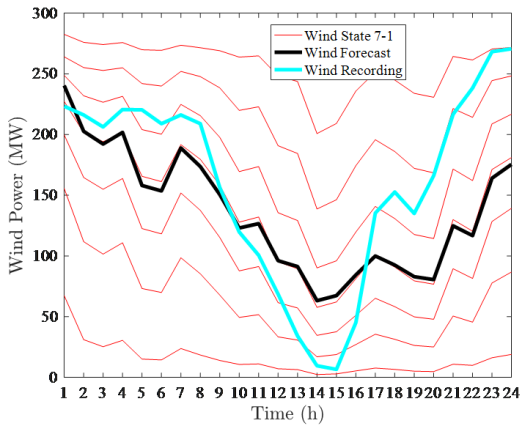

Fig. 11. The state values of WP of wind farm 3 on January 1, 2006

J. Verification of the Speedup of Big-M-CC-Compact

Table II presents the results of applying Big-M-CC-Original and Big-M-CC-Compact to MRUCM with $N = 3$ for the modified IEEE118-bus power system by turning on the presolve option.

TABLE II
RESULTS OF THE TWO METHODS USING GUROBI WITH PRESOLVE OPTION ON($N = 3$)

	Relaxation solution	Subproblem solution time(sec)	Total time (sec)
Big-M-CC -Compact	-1.57×10^7	0.11	70.83
	-2.16×10^6	20.40	
	-2.19×10^6	50.32	
Big-M-CC -Original	-1.78×10^7	0.18	98.53
	-5.66×10^6	35.89	
	-5.72×10^6	62.37	

(Note: the relaxation solution is the lower bound of SP)

Table II shows that after we turn on the Gurobi presolve options, the relaxation solution of Big-M-CC-Compact is still larger than that of Big-M-CC-Original at each iteration, though the extent is greatly reduced. Each iteration time and the total time of Big-M-CC-Compact of three iterations are still less than those of Big-M-CC-Original, only the extent is not so obvious as that when presolve options is turned off. This is because that the Gurobi might automatically reconstruct the original problem and greatly reduces the redundant constraints. However, the reconstructed model using Big-M-CC-Compact still has smaller model scale compared with that using Big-M-CC-Original, leading to less calculation time. More details about Gurobi reconstruction of the model can refer to the forum(http://www.gurobi.cn/picexhview.asp?id=90). Table IV shows that the lower bound of Big-M-CC-Compact of 3 iterations are -1.75×10^7, -2.16×10^6 and -2.19×10^6, which are larger than those of Big-M-CC-Original, respectively. This is because the feasible region of Big-M-CC-Original is larger than that of Big-M-CC-Compact and leads to the larger vertex number needed to be explored in the model using Big-M-CC-Original than that using Big-M-CC-Compact to increase the lower bound, as the result the calculation speed of Big-M-CC-Compact is mildly faster.

To further verify the effectiveness of Big-M-CC-Compact, Solver Cplex 12.8.0 (presolve options on) is applied to the MRUCM using Big-M-CC-Compact and Big-M-CC-Original for an IEEE 6-bus system, respectively. The system includes 3 thermal power units, 1 wind farm, and 3 loads. The total capacity of thermal power units is 420 MW, the total capacity of wind farm is 300 MW, and the peak load is 328 MW. The wind farm is connected at 4 nodes. The configuration of the system is presented at https://github.com/Tongjillj/Supplementary-Notes. The data of wind farm is also extracted from the National Renewable Energy Laboratory's Eastern Wind Dataset (https://www.nrel.gov/) for onshore wind farms No.6389. The parameters of MRUCM are $\mathrm{Pr}_P = 0.85$, $\mathrm{Pr}_R = 0.8$, $\Gamma_{total} = 16$. Table III shows the calculation results of MRUCM under different N.

TABLE III
COMPUTATIONAL RESULTS OF TWO METHODS

	State	Total solution time(sec)
Big-M-CC-Compact	3	11.7
	4	13.0
	5	30.0
	6	27.1
	7	26.7
Big-M-CC-Original	3	11.7
	4	15.2
	5	450.4
	6	219.5
	7	626.8

Table III shows that the computation time of the two methods all increases with the increase of N. When $N = 3$, the calculation time of the two methods is the same, which is because the scale of the model is too small to show the difference of calculation efficiency. As N increases, the computational time of the two models all gradually increase, but the time of the model using Big-M-CC-Compact is always much less than that using Big-M-CC-Original, especially only nearly $1/20$ of Big-M-CC-Original when $N = 7$. Therefore, the effectiveness and superiority of Big-M-CC-Compact are demonstrated strongly once again, compared with Big-M-CC-Original.