Multi-Timescale Simulation of Non-Spinning Reserve in Wholesale Electricity Markets

Binghui Li, *Member*, *IEEE*, Cong Feng, *Student Member*, *IEEE*, and Jie Zhang, *Senior Member*, *IEEE*
The University of Texas at Dallas
Richardson, TX 75080, USA
Email: {binghui.li, cong.feng1, jiezhang}@utdallas.edu

***Abstract*—Non-spinning reserve (NSR) represents an ancillary service that is intended to help power systems recover from unexpected contingencies. Many innovative methods to size NSR have been proposed in the literature. To evaluate their performance in terms of system reliability and economics, NSR allocation and deployment are needed in power market simulation models. This paper presents a multi-timescale market simulation to examine the allocation of NSR in wholesale electricity markets. Specifically, we develop a set of unit commitment (UC) and economic dispatch (ED) models and place emphasis on the simulation of thermal units over a variety of timescales. The performance of the models are evaluated with a synthetic power system that is built on the footprint of the Texas power market. The multi-timescale forecasts of load, wind, and solar power are aligned with the real-world Texas market timeline. By applying NSR requirements generated from a baseline and a dynamic sizing method, the results indicate that the dynamic method can significantly reduce the penalty cost associated with reserve shortage, though the total cost is slightly increased due to greater production cost.**

Index Terms*—*Non-spinning reserve, power market simulation, multi-timescale simulation, dynamic reserve

NOMENCLATURE

Sets

$\mathbf{G}$	Generators.
$\mathbf{G}^T \subseteq \mathbf{G}$	Thermal generators.
$\mathbf{G}^N \subseteq \mathbf{G}$	Non-thermal generators, $\mathbf{G}^N = \mathbf{G} \setminus \mathbf{G}^T$.
$\mathbf{G}^{TC} \subseteq \mathbf{G}^T$	Thermal generators, committed.
$\mathbf{G}^{TU} \subseteq \mathbf{G}^T$	Thermal generators, uncommitted.
$\mathbf{G}^{TUI}$	Thermal, uncommitted instant ($\subseteq \mathbf{G}^{TU}$).
$\mathbf{G}^{TUE}$	Thermal, uncommitted extended ($\subseteq \mathbf{G}^{TU}$).
$\mathbf{T}$	Time intervals.
$\mathbf{L}$	Transmission lines.
$\mathbf{B}$	Buses.
$\mathbf{B}^L \subseteq \mathbf{B}$	Load buses.
$\mathbf{B}^G \subseteq \mathbf{B}$	Buses with generators.
$\mathbf{G}_b \subseteq \mathbf{G}$	Generators on bus $b \in \mathbf{B}^G$.

Parameters

$D_{b,t}$	Demand at bus $b \in \mathbf{B}^L$ during interval $t \in \mathbf{T}$.
P_g^{max}, P_g^{min}	Upper and lower bounds of power production from thermal generator $g \in \mathbf{G}^T$.
$P_{g,t}^F$	Forecasted power at time t from non-thermal generator $g \in \mathbf{G}^N$.
K_g	Number of segments in piecewisely linearized cost function of generator $g \in \mathbf{G}^T$.
C_g^0	Intercept of the cost function of generator g, in \$.
C_g	Slope of the cost function of generator g, in \$/MWh.
C_g^{SU}, C_g^{SD}	Start-up and shut-down cost of generator g.
$C^{\Delta L}$	System-wide load curtailment cost, in \$/MWh.
$C^{\Delta R}$	System-wide regulating up/down reserve curtailment cost, in \$/MWh.
R_g^U, R_g^D	Ramp up/down rates of generator g.
R_g^{SU}, R_g^{SD}	Start-up/shut-down ramp rates of generator g.
$p_{g,0}$	Initial power production of generator g.
$v_{g,0}$	Initial commitment status of generator g.
T_g^{SU}, T_g^{SD}	The start-up/shut-down time of generator g in hours.
$SF_{\ell,b}$	Power shift factor, power flow increment on branch ℓ due to power injected at bus b.
Λ_ℓ	Power limit of branch ℓ.
RQ_t^{RU}	System regulation up reserve requirement at time t.
RQ_t^{RD}	System regulation down reserve requirement at time t.
RQ_t^{NSR}	System NSR requirement at time t.

Variables

$p_{g,t} \in \mathbb{R}_+$	Power level from generator $g \in \mathbf{G}$ at the end of interval $t \in \mathbf{T}$.
$\bar{p}_{g,t} \in \mathbb{R}_+$	Maximum power available from generator $g \in \mathbf{G}$ during interval $t \in \mathbf{T}$.
$\underline{p}_{g,t} \in \mathbb{R}_+$	Minimum power available from generator $g \in \mathbf{G}$ during interval $t \in \mathbf{T}$.
$v_{g,t} \in \{0,1\}$	Commitment status of generator $g \in \mathbf{G}^T$ during interval $t \in \mathbf{T}$.
$\delta_{b,t} \in \mathbb{R}_+$	Load curtailment at bus b in interval t.
$\delta_t^{RU} \in \mathbb{R}_+$	System-wide regulating up reserve shortage during interval t.
$\delta_t^{RD} \in \mathbb{R}_+$	System-wide regulating down reserve shortage during interval t.
$\delta_t^{NSR} \in \mathbb{R}_+$	System-wide NSR shortage during interval t.
$r_{g,t}^{RU} \in \mathbb{R}_+$	Regulation up reserve provided by generator $g \in \mathbf{G}^T$ during interval t.
$r_{g,t}^{RD} \in \mathbb{R}_+$	Regulation down reserve provided by generator $g \in \mathbf{G}^T$ during interval t.
$r_{g,t}^{NSR} \in \mathbb{R}_+$	NSR provided by generator $g \in \mathbf{G}^T$ during interval t.

$z_{g,t}^{P} \in \mathbb{R}_+$ Production cost of generator $g \in \mathbf{G}^T$ during interval $t \in \mathbf{T}$.

$z_{g,t}^{SU} \in \mathbb{R}_+$ Start-up cost of generator $g \in \mathbf{G}^T$ during interval $t \in \mathbf{T}$.

$z_{g,t}^{SD} \in \mathbb{R}_+$ Shut-down cost of generator $g \in \mathbf{G}^T$ during interval $t \in \mathbf{T}$.

$z_{t}^{\Delta} \in \mathbb{R}_+$ Load and reserve curtailment cost during interval $t \in \mathbf{T}$.

I. INTRODUCTION

A critical challenge in today's power system is the increasing risk of imbalance of load and demand caused by the growing penetration of non-dispatchable resources. Traditionally, load is balanced by fossil fuel fired units, which are dispatchable and predictable, and the only uncertainties in power systems come from load uncertainty and unexpected contingencies of thermal generators. In deregulated power markets, system operators play an important role in maintaining power system reliability by running ancillary service markets, where operating reserves are procured to mitigate the impacts of load uncertainty and generator contingencies [1]. Although a common purpose of these operating reserves is to maintain system reliability levels, their services, functions, and characteristics can vary. For example, regulating reserves are usually used to balance fluctuations of load and supply at second level, spinning reserves are provided by online units and deployed when contingencies occur. NSR is also used to mitigate impacts of system contingencies, however, it is often provided by offline units and can take longer to deploy.

The recent decade has witnessed a rapid growth of wind and solar power, which poses great threats to power system reliability because of their variability and uncertainty. Specifically, variability implies the need for more flexible units to counteract the rapid change of output, and uncertainty means greater reserve needs because of deviations from forecasts. The reserve margin in many power markets have increased significantly to comply with reliability standards, usually at the cost of market efficiency. The determination of appropriate reserve requirements to meet a given level of system reliability remains an open question, and usually involves trade-offs between market efficiency and system reliability. Traditionally, the NSR requirement is often assumed to be the capacity of the largest generator [2]. This simple method fails to reflect the increased variability and uncertainty owing to the addition of non-dispatchable resources. Recently, probabilistic methods have increasingly been used by system operators to account for wind or solar uncertainties [3]–[5]. Although these methods can reduce the amount of reserves, their impacts on system economics and reliability must be examined in a real-world power market setting, which is usually conducted by using market simulation models [6]–[8].

Most wholesale electricity markets adopt a two-settlement system that consists of a day-ahead market and a real-time market [9]. The day-ahead market runs one day ahead to commit slow-starting units and the real-time market runs periodically to commit fast-starting units and update power output settings of all participating units. This design allows system operators to respond to deviations from previous schedules as more up-to-date forecasts become available. Consequently, the data used in real-world power market simulations should follow real-world timelines [10].

The contributions of this paper are threefold. First, we develop a multi-timescale market simulation model that can accurately simulate the operation of thermal generators over varying timescales, and demonstrate its applicability using a synthetic network that is built on the footprint of the Texas power grid. Second, we adopt a set of forecasts that are generated by strictly following the real-world Electric Reliability Council of Texas (ERCOT) timeline. Last, we examine the economic impacts of different NSR requirements and reveal the trade-off between flexibility of online and offline units.

The remaining of this paper is organized as follows: Section II formulates the fundamental UC and ED models used in our analysis and describes our enhancement to represent thermal generators over multi-timescales. Section III briefly summarizes the data used in our analysis and case studies. Section IV present our discoveries of the impacts of NSR requirement on system scheduling and operation. Section V concludes this paper.

II. METHODOLOGY

Power markets in real world are typically operated over multiple timescales ranging from day ahead to real time. Like many power markets in the U.S., ERCOT has a day-ahead market and a real-time market. The day-ahead reliability unit commitment (DRUC) runs day ahead by taking day-ahead forecasts of load and renewable power productions, and commits slow thermal units. In the operating day, hourly reliability unit commitment (HRUC) models are run 60 mins before the operating hour based on the latest system states and forecasts to fine-tune the commitment statuses of other fast thermal units. After all thermal units are committed, ERCOT runs real-time economic dispatch (RTED) every 5 minutes to determine the desirable generation resource output levels. Usually, energy and ancillary services are co-optimized in the UC and ED models.

A. The fundamental model

The UC model is formulated as a mixed integer program by following Ref. [11], and the ED model is derived by fixing all commitment variables in the UC model. The objective of the model is to minimize total costs, which include actual production costs, fixed costs, and penalizing terms. The detailed model formulation is given as follows.

System-level load balance: We assume during each interval the power level ramps up or down linearly. Therefore, the energy production during interval t is obtained by the mean of $p_{g,t}$ and $p_{g,t-1}$, and the system-level load balance is:

$$\sum_{g \in \mathbf{G}} \frac{p_{g,t} + p_{g,t-1}}{2} + \sum_{b \in \mathbf{B}^L} \delta_{b,t} = \sum_{b \in \mathbf{B}^L} D_{b,t}, \forall t \in \mathbf{T} \quad (1)$$

Maximum and minimum power generation: Note that for $g \in \mathbf{G}^N$, $P^F_{g,t}$ is given by forecasts:

$$\underline{p}_{g,t} \leq p_{g,t} \leq \bar{p}_{g,t}, \forall g \in \mathbf{G} \tag{2a}$$

$$\bar{p}_{g,t} \leq P^F_{g,t}, \forall g \in \mathbf{G}^N \tag{2b}$$

Active power flow: $\forall \ell \in \mathbf{L}$:

$$-\Lambda_\ell \leq \sum_{b \in \mathbf{B}} SF_{\ell,b} \left(\sum_{g \in \mathbf{G}_b} p_{g,t} - \delta_{b,t} - D_{b,t} \right) \leq \Lambda_\ell \tag{3}$$

Regulating up/down reserve requirements: $\forall t \in \mathbf{T}$:

$$\sum_{g \in \mathbf{G}^T} r^{RU}_{g,t} + \delta^{RU}_t = RQ^{RU}_t \tag{4a}$$

$$\sum_{g \in \mathbf{G}^T} r^{RD}_{g,t} + \delta^{RD}_t = RQ^{RD}_t \tag{4b}$$

NSR requirements: $\forall t \in \mathbf{T}$:

$$\sum_{g \in \mathbf{G}^T} r^{NSR}_{g,t} + \delta^{NSR}_t = RQ^{NSR}_t \tag{5}$$

Cost of thermal generators: $\forall g \in \mathbf{G}^T, t \in \mathbf{T}$:

$$z^P_{g,t} = C^0_g \cdot v_{g,t} + C_g \cdot p_{g,t} \tag{6a}$$

$$z^{SU}_{g,t} \geq C^{SU}_g (v_{g,t} - v_{g,t-1}) \tag{6b}$$

$$z^{SD}_{g,t} \geq C^{SD}_g (v_{g,t-1} - v_{g,t}) \tag{6c}$$

Penalty costs associated with load curtailment and reserve shortage: $\forall t \in \mathbf{T}$:

$$z^\Delta_t = C^{\Delta L} \sum_{b \in \mathbf{B}^L} \delta^\xi_{b,t} + C^{\Delta R} (\delta^{RU}_t + \delta^{RD}_t + \delta^{NSR}_t) \tag{7}$$

Objective function:

$$\min z = z^\Delta_t + \sum_{g \in \mathbf{G}^T, t \in \mathbf{T}} \left(z^P_{g,t} + z^{SU}_{g,t} + z^{SD}_{g,t} \right) \tag{8}$$

Constraint (1) ensures the load balance on each bus. Equations (2a) allows all units to operate within their output limits, while the outputs of renewable units are limited by power forecasts in (2b). The power flow on all transmission lines are limited by (3) using shift factors. Requirements of reserves are met in (4)–(5), where the regulating reserve requirements in the upward and downward directions are enforced in (4a) and (4b), respectively. The NSR requirements are enforced in (5). Costs of thermal generators consist of production costs and fixed costs. Typically, production costs of thermal generators are represented by quadratic functions, and here for simplicity, we use linear functions to approximate them, as shown in (6a). The fixed costs are given by (6b)–(6c), which include start-up and shut-down costs. Equation (7) calculates total curtailment costs by summing up load curtailment costs and reserve curtailment costs. Last, the objective function in the deterministic run is given in (8) by summing up all costs.

Note that the model in (1)–(8) do not constrain the maximum/minimum available power or maximum/minimum available reserve from thermal generators, which we will discuss in the coming section.

B. Model Enhancement

1) Representation of Non-Spinning Reserve: To be qualified as an ancillary service provider, a generating resource must meet market-specific time response requirements, which indicate the maximum time the generating resource can take before it ramps up/down to the scheduled amount of reserve [1], [12]. Based on ERCOT's nodal protocol, the time limit for regulating reserves, spinning reserve, and NSR to be deployed are 5, 10, and 30 minutes, respectively [13]. Typically, such time response requirements are enforced in market simulation models by ramp rate and maximum available power constraints. For example, the available upward regulating reserve from an online thermal unit $g \in \mathbf{G}^{TC}$ during time $t \in \mathbf{T}$ is subject to the following constraints:

$$r^{RU}_{g,t} \leq \frac{5}{60} \cdot R^U_g \tag{9a}$$

$$r^{RU}_{g,t} \leq \bar{p}_{g,t} - p_{g,t} \tag{9b}$$

Similarly, the available downward regulating reserve from a thermal generator $g \in \mathbf{G}^{TC}$ is subject to:

$$r^{RD}_{g,t} \leq \frac{5}{60} \cdot R^D_g \tag{10a}$$

$$r^{RD}_{g,t} \leq p_{g,t} - \underline{p}_{g,t} \tag{10b}$$

Unlike regulating reserve and spinning reserve, NSR is typically provided by off-line thermal units in ERCOT, i.e., $g \in \mathbf{G}^{TU}$. Therefore, the available NSR from an offline thermal unit g must be subject to the following constraint:

$$r^{NS}_{g,t} \leq (1 - v_{g,t}) \cdot P^{NS}_g \tag{11}$$

where P^{NS}_g denotes the available NSR from unit g, and is given by the following equation:

$$P^{NS}_g = \begin{cases} 0, & T^{SU} > 0.5 \text{ h} \\ P^{min}_g + (0.5 - T^{SU}) \cdot R^U_g, & T^{SU} \leq 0.5 \text{ h} \end{cases} \tag{12}$$

2) Multi-Timescale Representation of Thermal Generators: A critical challenge in multi-timescale market simulation is the commitment status of a thermal generator. Typically, the day-ahead market is used to determine the commitment statuses of slow-starting units, which are given by DRUC. Once committed, these units will go online in the operating day and follow their committed schedules. The remaining thermal generators, which are usually medium- to fast-starting units, will be committed in the real-time market, which runs HRUC to determine the schedules. The ED models, which run in real time, do not commit any units, but are used to produce a least-cost dispatch of online resources and calculate locational marginal prices. Once a commitment is made, the following models must start up or shut down the unit by following the committed schedules, therefore the HRUC model must distinguish the committed units from the uncommitted units. In our study, we use $\mathbf{G}^{TU}$ and $\mathbf{G}^{TC}$ to represent the set of uncommitted and committed units, respectively. Note that we assume $\mathbf{G}^{TC} = \emptyset$ in the DURC model because all units are uncommitted in the day-ahead market. Similarly, we assume $\mathbf{G}^{TU} = \emptyset$ in all RTED models because all units are committed.

The start-up/shut-down time of thermal units in real world varies with the types of prime driver and operating conditions owing to their different thermodynamic characteristics [14]. For example, the start-up of a coal-fired steam unit may take hours or even days, while the most flexible gas-fired units can start up within several minutes. Some models in the literature neglect the start-up/shut-down trajectories and assume the thermal units can switch between online and offline status instantly [15]. Although such assumptions might suffice for a model with a fixed timescale, failure in accurately modeling the start-up/shut-down dynamics of thermal units could lead to inconsistent results in multi-timescale simulations. If a unit takes half an hour to start up, it can be viewed as an instant unit in the hourly UC model, whereas its start-up trajectory could span multiple intervals in the sub-hourly UC or ED models. In addition, as NSR can only be provided by offline units, the ability of accurately modeling the start-up trajectories of thermal units is fundamental to the simulation of the allocation and deployment of NSR.

Many models [16], [17] have been proposed in the literature to formulate the start-up/shut-down trajectories of thermal units, and some of them can even distinguish between different start-up modes, i.e., cold, warm, and hot start [17]. Because of the varying time interval size in the multi-timescale market simulation, it is necessary to model the start-up/shut-down trajectories of a thermal unit when the time interval is both longer and shorter than its start-up/shut-down time. In our study, we follow Ref. [16] to model the start-up/shut-down trajectories of a thermal unit if its start-up/shut-down time spans more than one interval—i.e., we treat it as an "extended" unit ($g \in \mathbf{G}^{TUE}$). In addition, we treat a thermal unit as an "instant" unit ($g \in \mathbf{G}^{TUI}$) if its start-up/shut-down time is less than one interval, and we follow the formulations in Ref. [15] to describe its behaviors.

In summary, the set of thermal generators ($\mathbf{G}^{T}$) is further divided into three subsets—i.e., committed units ($\mathbf{G}^{TC}$), instant units ($\mathbf{G}^{TUI}$), and extended units ($\mathbf{G}^{TUE}$)—and each subset will be modeled differently. The maximum ($\bar{p}_{g,t}$) and minimum ($\underline{p}_{g,t}$) available power of a thermal generator are constrained according to the subset it belongs to.

Committed units ($\mathbf{G}^{TC}$): Once a thermal unit is committed, its power generation can range between P_g^{max} and P_g^{min}, i.e., its upper and lower bounds of power production, respectively. In addition, we calculate the dispatch limits of a thermal unit if this unit is scheduled to shut down such that the power generation schedules will not violate its commitment schedules. The $\bar{p}_{g,t}$ and $\underline{p}_{g,t}$ are jointly determined by its upper and lower bounds of power production and its dispatch limits.

Instant units ($\mathbf{G}^{TUI}$): The maximum and minimum available power of an instant unit is determined jointly by its ramp rates and upper/lower bounds of power production:

$$\bar{p}_{g,t} - p_{g,t-1} \leq R_g^U \cdot v_{g,t-1} + R_g^{SU}(v_{g,t} - v_{g,t-1}) + P_g^{max}(1 - v_{g,t}), \quad \forall g \in \mathbf{G}^{TUI}, t \in \mathbf{T} \tag{13a}$$

$$\bar{p}_{g,t} \leq P_g^{max} \cdot v_{g,t+1} + R_g^{SD}(v_{g,t} - v_{g,t+1}), \quad \forall g \in \mathbf{G}^{TUI}, t \in \mathbf{T} \tag{13b}$$

$$p_{g,t-1} - p_{g,t} \leq R_g^D \cdot v_{g,t} + R_g^{SD}(v_{g,t-1} - v_{g,t}) + P_g^{max}(1 - v_{g,t-1}), \quad \forall g \in \mathbf{G}^{TUI}, t \in \mathbf{T} \tag{13c}$$

where (13a) and (13b) limit the maximum available power of thermal generators by accounting for ramp rates, and (13c) limits the power generation during a ramp-down or shut-down process.

Extended units ($\mathbf{G}^{TUE}$): The start-up/shut-down trajectory of $g \in \mathbf{G}^{TUE}$ can span multiple intervals, and the $\bar{p}_{g,t}$ and $\underline{p}_{g,t}$ are jointly determined by its commitment status, ramp rates, and start-up/shut-down indicators. We follow [16] to formulate these constraints, and the model formulations are omitted due to the page limit.

C. Multi-Timescale UCED model

The timeline of our simulation is shown in Fig. 1. We simulate the ERCOT market process by following its temporal characteristics: The DRUC runs once per day and consists of twenty-four 1-hour intervals, the HRUC runs every 1 hour and consists of twelve 5-min intervals, and the RTED runs every 5 min and covers three 5-min intervals. Note that at the end of each hour, the RTED horizons are limited by the numbers of remaining 5-min intervals in that hour, and this is because all thermal generators must be committed in an RTED run, whereas the previous RTUC model can only provide commitment statuses up to the last 5-min interval of this hour. In summary, one day's simulation includes 1 DRUC run, 24 HRUC runs, and 288 RTED runs.

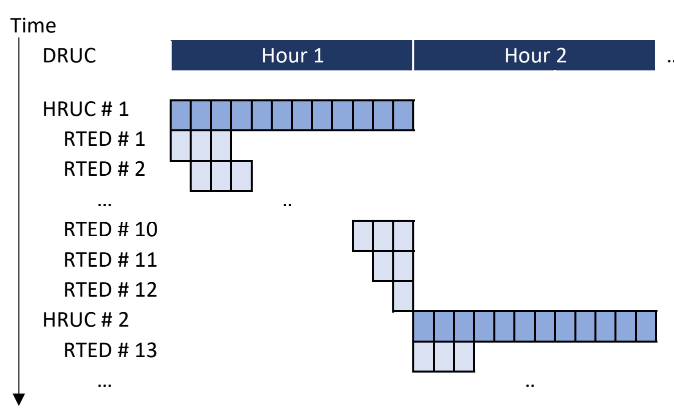


Fig. 1: Timeline of the multi-timescale simulation. Note that the DRUC covers 24 hours, and we only show the first 2 hours.

Initial power generation levels ($p_{g,0}$) and commitment statuses ($v_{g,0}$) must be given to the DRUC, HRUC # 1 and RTED #1. To initialize the model, we run a DC optimal power flow (DC-OPF) model by using the forecasts of load and power generation from non-thermal units in the first DRUC interval. In addition, we assume that only the first interval of each RTED model is financially binding, and the initial power generation levels and commitment statuses of all models after DRUC, HRUC # 1 and RTED #1 are obtained from the binding interval of the RTED run right before that model.

III. Data Summary

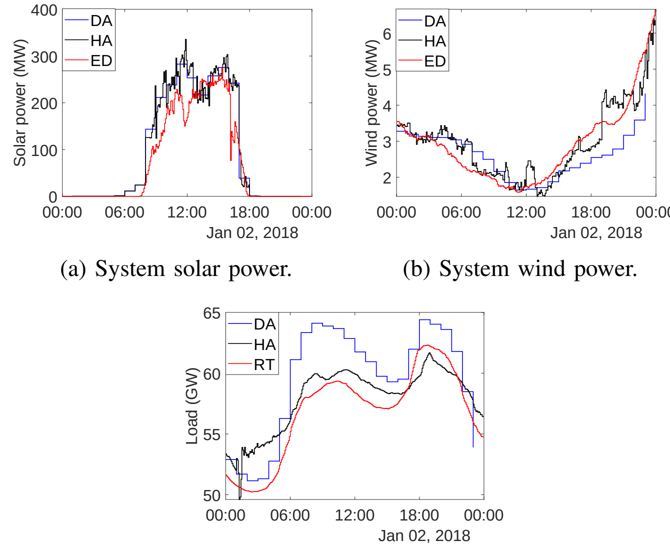
To simulate real-world power systems, we use a well-developed and publicly available synthetic network developed by [18], which is built on the footprint of ERCOT. The forecasts of load, solar, and wind power are made by following real-world schedules, i.e., the day-ahead forecasts are made at noon on the previous day, hour-ahead forecasts are made one hour ahead, and we use the realized data as the RTED inputs. Specifically, day-ahead forecasts from ERCOT are directly used. A machine learning-based multi-model (M3) forecasting methodology that has been proven to be effective in wind [19], solar [20], and load [21] forecasting, and we utilize it to generate hour-ahead forecasts. Detailed temporal characteristics of the forecasts used in this study are summarized in Table I. Fig. 2 show the data used in the three sets of models.

TABLE I: Temporal characteristics of model inputs.

Model	Frequency	Refresh	Horizon	Forecasts used
DRUC	Daily	Hourly	24 hour	Day-ahead
HRUC	Hourly	5 min	60 min	Hour-ahead
RTED	Every 5 min	5 min	15 min	Real-time

We assume that the load curtailment cost is \$9,000/MWh, according to the value of load loss (VOLL) used by ERCOT [22]. We use \$2,000/MWh, the low system-wide offer cap (LCAP) provided by the Texas Public Utility Commission [23], as the penalty price for both responsive and non-responsive reserve shortage. We use \$5,500/MWh, the average of LCAP and VOLL, as the penalty price for the regulating up/down reserve shortages. The penalty prices are selected such that responsive reserves are curtailed ahead of regulating reserves, and load will be the last to be curtailed [13].

To demonstrate our model's capability of simulating the allocation of NSR and examine the economic impacts of different NSR requirements, we use two sets of NSR requirements: baseline and dynamic. The baseline method is the method used by ERCOT in their daily operations. This method determines the NSR requirement using the 70th to 95th percentile of hourly net load uncertainty from the same month of the previous three years [24]. The dynamic method is a method developed in our previous work that incorporates probabilistic net load forecasts [25]. As shown in Fig. 3, the NSR requirements in the baseline are greater than those in the dynamic case during most time intervals. We also introduce another case where NSR constraints are not enforced, i.e., no NSR requirement, and we denote it as the "No NSR" case. We use 2% of demand as the regulating reserve requirement. Note that we assume a constant reserve margin for the regulating reserves to focus on the NSR requirements.



(a) System solar power. (b) System wind power.

(c) System load.

Fig. 2: System-level forecasts of (a) solar power, (b) wind power, and (c) load.

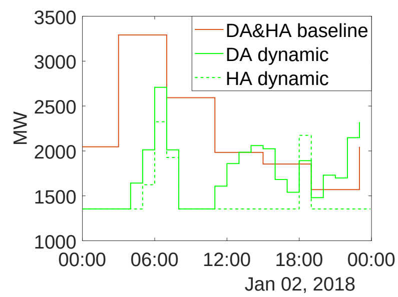


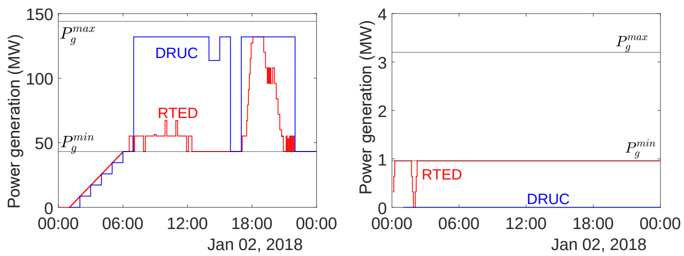
Fig. 3: Requirements of NSR requirements in both the baseline and the dynamic case.

IV. Results

A. Profiles of thermal generators

We first examine the power generation schedules of thermal units. Fig. 4 shows the scheduled power generation levels of two natural gas fired units in both the day-ahead and real-time markets. We demonstrate the start-up/shut-down trajectories by examining two units: The gas unit in Fig. 4a is a slow unit that takes 5 hours to start up completely, whereas the unit in Fig. 4b is a fast unit that takes 15 min to start up completely. The slow unit is scheduled to start at 1am in the DRUC model, and the RTED profile shows that it follows the DRUC schedule and ramps up to its minimum online capacity P_g^{min} at 6am exactly. By contrast, although the fast unit is scheduled to remain offline throughout the day in the DRUC model, the RTUC model updates its schedule by starting it up at 0am and ramping up to P_g^{min} after 3 RTED intervals. In addition, it shuts down at 1:40am, completes the shut-down process at 1:55am, and starts up again at around 2am.

Fig. 4 also shows the schedule updates owing to forecast errors. The slow unit is scheduled to ramp up to approximately 130 MW at 7am in the day-ahead schedule, however, the real-time load is significantly lower, as shown in Fig. 2c, and it only dispatches around 55 MW in the the RTED schedule.



(a) A slow gas turbine. (b) A fast gas turbine.

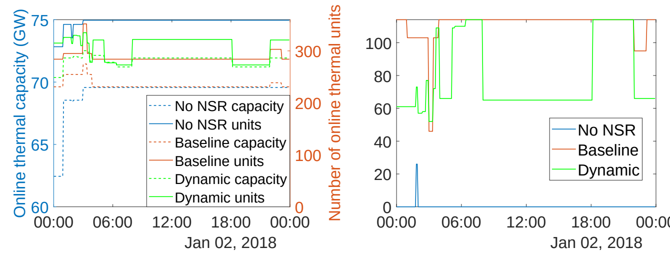
Fig. 4: Scheduled power generation of gas turbines in the DURC and RTED models. Note that $T^{SU} = 5$ h in (a) and $T^{SU} = 15$ min in (b).

B. The impact of NSR requirements on thermal units

In this study, we have requirements for both regulating and non-spinning reserves, which are provided by different units in compliance with ERCOT operating protocols: regulating reserves must be provided by online units ($v_{g,t} = 1$), whereas NSR can only be provided by offline units ($v_{g,t} = 0$). By examining the numbers and capacities of online and offline units, we can understand the effects of different NSR requirements on system flexibility and adequacy. Figs. 5a and 5b show the numbers and capacities of online and offline units, respectively. Note that Fig. 5b only includes the number of offline units that are qualified as NSR resources, i.e., $T^{SU} \leq$ 30 min.

As shown in Fig. 5a, the adoption of NSR significantly reduces the number of online units. In the "No NSR" case, there are over 350 units online throughout most of the day, whereas the baseline only has fewer than 300 units online. In addition, the size of NSR requirements has direct impacts on the numbers and capacities of online units: the comparison between the baseline and the dynamic case suggests that the dynamic case has both greater numbers and capacities of online units as a result of the reduced NSR requirements.

Although the online capacity is typically positively correlated with the number of online units, Fig. 5a shows otherwise. For example, despite that the "No NSR" case has the greatest numbers of online units throughout the day, it has the least online capacities and the dynamic case has 2.4 GW of additional online capacity compared to the other two cases. In addition, the online capacities of the "No NSR" case and the baseline are almost identical after 4am, although the latter has significantly fewer online units. This implies that the adoption of NSR constraints can affect the types of online thermal units. Specifically, the qualified NSR units are typically small but fast natural gas fired gas turbines, which are kept offline in the NSR enforced cases. Consequently, the two NSR enforced cases use large units to meet the load balance requirements. This can explain the smaller numbers of online units but greater online capacities in both NSR enforced cases.



(a) Online units. (b) Offline NSR units.

Fig. 5: (a) Capacities and numbers of online units. (b) Numbers of offline NSR units.

The numbers of offline NSR units are directly relevant to the NSR requirements and we further examine them in Fig. 5b. As shown in Fig. 5b, when there are no NSR requirements, all qualified NSR units are online in the "No NSR" case. The baseline keeps more NSR units offline as a result of its greater NSR requirements. In addition, the dynamic case starts up and shuts down approximately 60 NSR units following the changes of NSR requirements during early morning and late afternoon. As shown in Fig. 3, the peaks of the numbers of offline NSR units are overlapping with the peaks of NSR requirements in the dynamic case, which are around early morning (6am) and late afternoon (6pm). These NSR units can follow such schedules because they are very flexible units that can start up or shut down within half an hour, however, the frequent start-ups and shut-downs can incur additional costs, therefore negatively impact system economics.

C. Economic insights

The costs are tabulated by category in Table II. The "No NSR" case presents the least total cost because the NSR requirements are not enforced; and because in minimization problems, additional constraints can only lead to greater objective values. By contrast, both the baseline and the dynamic cases show greater total costs as a result of the NSR requirements.

We further demonstrate the economic impacts of the NSR requirements by examining each cost term. The production cost comes primarily from the fuel and maintenance costs of online units, and therefore are driven by the online capacities. Fig. 6 shows the production costs during all binding RTED intervals, and a comparison with the capacities of online units in Fig. 5a indicates that the dynamic case has the greatest production costs because of the greatest online capacities. The fixed cost, on the other hand, is mainly driven by the start-up and shut-down of thermal units. As we have discussed in the previous section, the more frequent start-up and shut-down of NSR units in the NSR enforced cases results in greater fixed costs: the fixed cost in the dynamic and baseline case are 40% and 20% greater than the "No NSR" case, respectively.

The reserve shortage costs in both NSR enforced cases come from a shortage of regulating reserves. This implies a reduction of flexibility of online units as a result of fewer flexible units online. As discussed in the previous section, the online units in both NSR enforced cases consist of large but less flexible units, whereas regulating reserves require 5 minutes response time; therefore, although both NSR enforced units have greater online capacities, their capabilities of responding to rapid net load changes are compromised. In most power markets, the regulating reserve is used to mitigate small fluctuations in net load in real time, whereas the NSR is intended to help the system recover from unanticipated contingencies [1]. The comparison between the "No NSR" case and the NSR enforced cases implies that the capabilities of dealing with contingencies are improved by increasing the risks of violating the requirements of regulating reserve. In addition, the comparison between the baseline and the dynamic case indicates that the dynamic NSR requirements can bring down the shortage of regulating reserves. This can be explained by the fact that the dynamic case requires fewer NSR units, which leaves more flexible units online towards the requirements of regulating reserve.

TABLE II: Cost breakdown of all cases (in $).

	No NSR	Baseline	Dynamic
Production	21,141,243	21,221,403	21,309,697
Fixed	63,888	75,967	91,441
Load curtailment	0	0	0
Reserve Shortage	0	94,714	33,205
Total cost	21,205,131	21,392,083	21,434,343

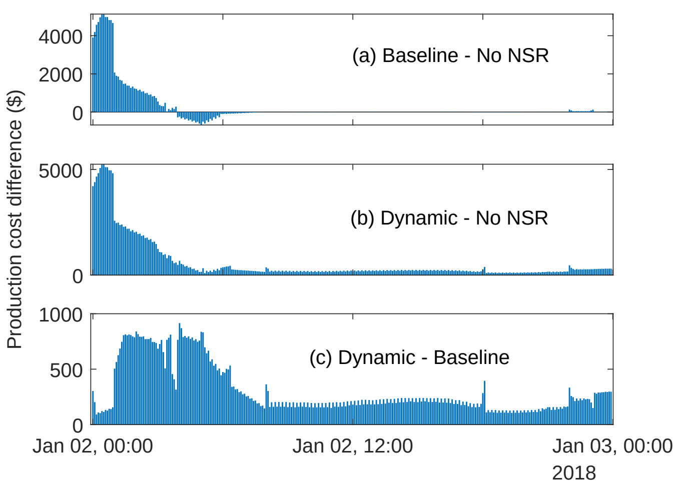


Fig. 6: Differences of production costs as a function of time between (a) The baseline and the "No NSR" case. (b) The dynamic case and the "No NSR" case. (c) The dynamic case and the baseline.

V. Conclusion

We have developed a multi-timescale market simulation model to simulate the allocation of NSR reserves in wholesale electricity markets. This model consists of a set of sequential UC and ED model runs to simulate the real-world market process. To address the challenge associated with the representation of start-up and shut-down trajectories of thermal units over varying timescales, we have designed three sets of constraints that can be applied to each thermal unit depending on its start-up/shut-down time and commitment status. In addition, we formulate the capability of thermal units to provide regulating and NSR services by strictly following their time response requirements.

We validate the model and evaluate its performance by using a synthetic power system that is built on the footprint of ERCOT. All forecasts used in our case studies are generated following the real-world timeline in ERCOT. The resulting thermal generator profiles suggest that our model accurately reflects the start-up and shut-down trajectories over all studied timescales, and all energy and reserve services are provided without violating the operating limits of thermal generators.

We also examine the impacts of NSR requirements on market economics by applying different NSR requirements. Our results suggest that although the NSR requirements can improve the system's capability of recovering from contingencies, this improvement is made available by keeping more flexible units offline. Consequently, the net load and regulating reserve requirements must be met by larger yet less flexible units, which could incur additional costs because of greater production costs and the compromise of flexibility of online units.

Finally, there are several open research questions that can be studied in the future. First, our study does not show the benefits of adopting dynamic NSR requirements in terms of reliability improvement. Many studies have shown that there is a trade-off between system reliability and economics [4], whereas our focus is mainly placed on system economics. To evaluate the impact of different NSR reserves on system reliability, contingency must be simulated because NSR is designed to mitigate the risk caused by contingencies. In addition, our results also suggest the trade-off between the flexibility of online units and offline units, which implies that market operators should consider the overall system reliability by jointly determine the requirements of operating reserves and contingency reserves. This could yield new insights towards the decision-making of market operators. In addition, the finest time interval in our model is 5 min, and it cannot reflect system dynamics within a 5-min time frame, which is often necessary to evaluate system reliability performance in compliance of North American Electric Reliability Corporation standards [26], therefore, our analysis can be supplemented by simulations over finer timescales.

Acknowledgment

This work was supported by the National Renewable Energy Laboratory under Subcontract No. ZDJ-8-82257-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308).

References

[1] Z. Zhou *et al.*, "Survey of us ancillary services markets," tech. rep., Argonne National Lab.(ANL), Argonne, IL (United States), 2016.

[2] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*. John Wiley & Sons, 2013.
[3] M. A. Matos and R. J. Bessa, "Setting the operating reserve using probabilistic wind power forecasts," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 594–603, 2011.
[4] M. A. Ortega-Vazquez and D. S. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 114–124, 2009.
[5] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. J. Bessa, D. Flynn, E. Gomez-Lazaro, and N. K. Detlefsen, "Methodologies to determine operating reserves due to increased wind power," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 713–723, 2012.
[6] M. Cui, J. Zhang, H. Wu, and B. Hodge, "Wind-friendly flexible ramping product design in multi-timescale power system operations," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1064–1075, 2017.
[7] M. Cui and J. Zhang, "Estimating ramping requirements with solar-friendly flexible ramping product in multi-timescale power system operations," *Applied energy*, vol. 225, pp. 27–41, 2018.
[8] H. Wu, I. Krad, A. Florita, B. Hodge, E. Ibanez, J. Zhang, and E. Ela, "Stochastic multi-timescale power system operations with variable wind generation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3325–3337, 2017.
[9] S. Stoft, "Power system economics," *Journal of Energy Literature*, vol. 8, pp. 94–99, 2002.
[10] D. Yang, E. Wu, and J. Kleissl, "Operational solar forecasting for the real-time market," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1499–1519, 2019.
[11] B. Li *et al.*, "A clustering-based scenario generation framework for power market simulation with wind integration," *J. Renew. Sustain. Ener.*, vol. 12, no. 3, p. 036301, 2020.
[12] J. L. Jorgenson and P. L. Denholm, "Modeling primary frequency response for grid studies," tech. rep., National Renewable National Lab.(NREL), Golden, CO (United States), Jan. 2019.
[13] ERCOT, "Ercot nodal protocols." http://www.ercot.com/mktrules/nprotocols/current, 2020.
[14] M. A. Gonzalez-Salazar *et al.*, "Review of the operational flexibility and emissions of gas- and coal-fired power plants in a future with growing renewables," *Renew. Sust. Energ. Rev.*, vol. 82, pp. 1497 – 1513, 2018.
[15] M. Carrión and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1371–1378, 2006.
[16] J. M. Arroyo and A. J. Conejo, "Modeling of start-up and shut-down power trajectories of thermal units," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1562–1568, 2004.
[17] G. Morales-España *et al.*, "Tight and compact milp formulation of start-up and shut-down ramping in unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1288–1296, 2013.
[18] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, 2017.
[19] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Applied Energy*, vol. 190, pp. 1245–1257, 2017.
[20] C. Feng, M. Cui, B.-M. Hodge, S. Lu, H. F. Hamann, and J. Zhang, "Unsupervised clustering-based short-term solar forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2174–2185, 2018.
[21] C. Feng, M. Sun, and J. Zhang, "Reinforced deterministic and probabilistic load forecasting via q-learning dynamic model selection," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1377–1386, 2019.
[22] R. Surendran, W. Hogan, H. Hui, and C.-N. Yu, "Scarcity Pricing in ERCOT," 2016.
[23] Public Utility Commission of Texas, "Electric Substantive Rules - Chapter 25 Rules," 2018.
[24] The Electric Reliability Council of Texas, "2019 Methodology for Determining Minimum Ancillary Service Requirements." (2019).
[25] C. Feng, M. Sun, J. Zhang, K. Doubleday, B.-M. Hodge, and P. Du, "A data-driven method for adaptive reserve requirement estimation via probabilistic net load forecasting," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2020.
[26] North American Electric Reliability Corporation, "Probabilistic Adequacy and Measures: Technical Reference Report Final," 2018.