Modeling flexible generator operating regions via chance-constrained stochastic unit commitment

Bismark Singh · Bernard Knueven · Jean-Paul Watson

Received: 19 October 2019 / Accepted: 10 April 2020 / Published online: 11 July 2020 © The Author(s) 2020

Abstract
We introduce a novel chance-constrained stochastic unit commitment model to address uncertainty in renewables’ production in operations of power systems. For most thermal generators, underlying technical constraints that are universally treated as “hard” by deterministic unit commitment models are in fact based on engineering judgments, such that system operators can periodically request operation outside these limits in non-nominal situations, e.g., to ensure reliability. We incorporate this practical consideration into a chance-constrained stochastic unit commitment model, specifically by infrequently allowing minor deviations from the minimum and maximum thermal generator power output levels. We demonstrate that an extensive form of our model is computationally tractable for medium-sized power systems given modest numbers of scenarios for renewables’ production. We show that the model is able to potentially save significant annual production costs by allowing infrequent and controlled violation of the traditionally hard bounds imposed on thermal generator production limits. Finally, we conduct a sensitivity analysis of optimal solutions to our model under two restricted regimes and observe similar qualitative results.

Keywords Stochastic optimization · Unit commitment · Power systems operations · Chance constraints · Emergency operations

Bismark Singh bismark.singh@fau.de
Bernard Knueven Bernard.Knueven@nrel.gov
Jean-Paul Watson jeanpaulwatson@llnl.gov

1 Discrete Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany
2 Computational Science Center, National Renewable Energy Laboratory, Golden, CO 80401, USA
3 Center for Applied Scientific Computing, Global Security Directorate Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
1 Introduction

The standard unit commitment (UC) problem for power systems operations involves determining which thermal generators should be scheduled to meet projected demand for power over a given time horizon, while ensuring physical and operational constraints are satisfied. The time horizon is typically one or two days at an hourly resolution. Feasible operating constraints for thermal generators include limits on ramping, minimum up and down times, startup and shutdown ramp rates, and minimum and maximum production levels. The UC problem is extensively studied in the optimization and power systems literature (Anjos and Conejo 2017; Damci-Kurt et al. 2016; Ostrowski et al. 2015; Padhy 2004; Queyranne and Wolsey 2017; Silbernagl 2016). It can be formulated as a mixed-integer linear program (MILP) and solved with commercial branch-and-cut software packages (Knueven et al. 2020, 2018b; Morales-Espana et al. 2013; O’Neill 2017). Prior to recent advances in MILP solver technology and UC formulations, even modest instances were computationally challenging to solve in time limits required by operations. Consequently, techniques such as Lagrangian decomposition (Borghetti et al. 2003), Benders decomposition (Wu and Shahidehpour 2010), and metaheuristics (Kazarlis et al. 1996) were previously employed. Presently, MILP solvers are regularly employed by Independent System Operators (ISOs) and Vertically Integrated Utilities to solve UC instances for real time operations; see, e.g. O’Neill (2017) and Ott (2010).

The standard UC model is implicitly a deterministic MILP, with inherently uncertain parameters such as demand being populated with their forecasted “point” quantities, with potential deviations from forecasted quantities being addressed by reserve margins. With the growing deployment of renewable energy sources, particularly wind and solar generators, there is increasing interest in explicitly treating uncertain aspects of the UC problem—yielding stochastic UC models; see, e.g., van Ackooij et al. (2018), Wang et al. (2012) and Zhao et al. (2013). Takriti et al. (2000) present a stochastic single-UC formulation with uncertain electricity prices, but from the perspective of an individual generator participating in a market. Uncertain demand is also a factor driving interest in stochastic UC formulations; see, e.g., Ozturk et al. (2004).

Stochastic UC is commonly formulated as an extensive stochastic program; see, e.g., Birge and Louveaux (2011) for an introduction to this decision formalism. Standard stochastic programming models assume that uncertainty is captured via a finite set of discrete scenarios, and that any solution to the model ensures feasibility of all constraints in all scenarios. A less stringent approach is to assume that some constraints can be violated with a small probability, resulting in chance constrained stochastic programming models. Due to their flexibility and explicit acknowledgment that certain constraints may be violated while ensuring reliability, chance constrained models are frequently found in the stochastic UC literature; see, e.g., Pozo and Contreras (2013), Singh et al. (2018), Wang et al. (2012) and Wu et al. (2014). Chance constrained models are seldom solved using a naive extensive formulation; see (Ozturk et al. 2004) for an approximation using iteratively updated union bounds, Pozo and Contreras (2013) for an approximation using conditional value-at-risk, Zhao et al. (2014) for a Monte Carlo sample average approximation, and Watson et al. (2010) for decomposition methods.
Operational limits on thermal generators are key inputs to both deterministic and stochastic UC models. Often, these limits are based on engineering judgment or economic considerations, rather than physical limitations. For example, minimum power outputs are commonly set to levels such that lower production levels are not profitable for operators, while maximum power outputs can be exceeded in practice for short periods of time to ensure reliable system operation. Consequently, system operators can, and do, run thermal generators beyond these limits (Kraemer 2013). We further note that such limits are also explicitly recognized by operators as “soft” constraints (Blumsack 2018), reinforcing the notion that they can be occasionally violated in practice. ISOs within the US, such as MISO (2018) and System Operations Division, PJM (2019), have mechanisms in place to allow for infrequent exceedance of normal production (i.e., dispatch) limits. Such mechanisms are different from spinning reserves and other ancillary service products in that they are typically held back for when net-load is significantly different from the forecasted quantity, whereas ancillary services are commonly dispatched to address contingencies (generator and/or transmission) and more typical deviations from net-load forecasts.

In principle, a stochastic programming model for UC can address anticipated variations in net-load (Wang and Hobbs 2014). For reserve products, market operators typically procure sufficient generation to cover a $1 - \varepsilon$ proportion of possible net-load realizations (Cornelius 2014). Compared to holding energy in reserve, stochastic programming can yield overly conservative schedules by scheduling sufficient generation to cover all—or a very large range of—possible load realizations. Our proposed stochastic UC model yields a method for incorporating the practical consideration of procuring sufficient generation to address a large proportion of possible net-load realizations into stochastic UC, while allowing for a currently-used recourse action (emergency capacity) if a low probability net-load realization is not covered under normal operations.

To the best of our knowledge, previous work has not analyzed the effect of periodically violating these soft operational constraints associated with thermal generators, in support of either cost reductions or system reliability. We acknowledge that a thermal generator should not be operated outside its prescribed engineering and/or economic limits for significant amounts of time, as this would likely incur increased maintenance costs and lifetime reductions (Dahal and Chakpitak 2007). Further, running thermal generators beyond their prescribed ratings is not fuel efficient (Knudsen et al. 2017). However, occasional violations may not significantly inflate maintenance costs or reduce lifetime, and may yield significant overall operational cost savings. Because of the scales involved, even a 1% savings in energy production costs is worth more than $10 billion per year globally, and more than $1 billion per year in the U.S. alone (O’Neill 2007). Thus, it is important to understand the trade-offs between the frequency and magnitude of these violations and any potential cost savings.

Here, we explicitly allow thermal generators to occasionally produce beyond their prescribed minimum and maximum ratings, while simultaneously limiting the extent of violations. We informally refer to the operation of a thermal generator in this mode as an “non-nominality”. We limit the number of these non-nominalities to a small quantity using a chance constraint, in the broader context of a stochastic UC formulation. Unlike traditional chance-constrained stochastic programming models
where violations are generally unbounded, thermal generators are still restricted by their absolute minimum and maximum ratings. As these absolute maximum ratings are typically proprietary information and consequently unknown to system operators, we take the absolute ratings to be a few percent greater than the prescribed ratings. Thus, unlike a traditional chance constraint where the magnitude of violations could be unrestricted, in our work the non-nominalities qualify as (tightly) bounded violations.

We summarize our main contributions in this article as follows:

- We present a mathematical formulation for a chance-constrained stochastic UC model where thermal generators are allowed to produce modestly beyond their technical minimum and maximum ratings with a small probability, in order to address discrepancies between forecasted and actual net-load.
- We show that an extensive form of the resulting chance-constrained stochastic UC model can be solved directly by commercial MILP branch-and-cut solvers for modestly sized problem instances.
- We demonstrate significant operational cost savings (on the order of $\approx 1\%$) can be obtained via infrequent and modest exceedance of nominal thermal generator production limits.
- We analyze the structure of optimal solutions to our chance-constrained UC model and assess the generalizability of our results to distinct test cases and operational restrictions.

The remainder of this paper is organized as follows. In Sect. 2, we present our chance-constrained stochastic UC model. In Sect. 3, we describe the data and sources associated with two different UC test cases that serve as the basis for our computational experiments. Our experimental results are detailed in Sect. 4, and include a sensitivity analysis to the parameters associated with the limit violation’s frequency and magnitude. We conclude in Sect. 5 with a summary of our contributions and plans for future work.

2 Chance-constrained stochastic unit commitment formulation

We now present our chance-constrained stochastic unit commitment formulation. We begin in Sect. 2.1 by introducing the notation, and then formally describe the mathematical programming model in Sect. 2.2.

2.1 Notation

Indices and Sets

\begin{align*}
g & \in \mathcal{G} & \text{Thermal generators}. \\
t & \in \mathcal{T} & \text{Hourly time steps: } 1, \ldots, T; \text{i.e., } [a, b) \in \mathcal{T} \times \mathcal{T} \text{ such that } b \geq a + UT^g. \\
l & \in \mathcal{L}^g & \text{Piecewise production cost intervals for generator } g: 1, \ldots, L_g. \\
s & \in \mathcal{S}^g & \text{Start-up categories for generator } g, \text{ from hottest (1) to coldest (} S_g). \\
\omega & \in \Omega & \text{Scenarios: } \omega_1, \ldots, \omega_N.
\end{align*}

Parameters: First Stage

\[\mathcal{G}, \mathcal{T}, \mathcal{L}^g, \mathcal{S}^g, \Omega\]
\( C^{l,g} \) Marginal cost for piecewise segment \( l \) for generator \( g \) (S/MWh).
\( \bar{C}^g \) Marginal cost for production above \( \bar{P}^g \) (S/MWh).
\( C^g \) Marginal cost for production below \( \underline{P}^g \) (S/MWh).
\( C^{R,g} \) Cost of generator \( g \) running and operating at minimum production \( \underline{P}_g \) (S/h).
\( C^{S,g} \) Start-up cost of category \( s \) for generator \( g \) (S).
\( DT^g \) Minimum down time for generator \( g \) (h).
\( \bar{P}^g \) Maximum power output for generator \( g \) under normal operations (MW).
\( \underline{P}^g \) Maximum power output for generator \( g \) under non-normal operations (MW).
\( P^g \) Minimum power output for generator \( g \) under normal operations (MW).
\( \underline{P}_g \) Minimum power output for generator \( g \) under non-normal operations (MW).
\( P^{l,g} \) Maximum power available for piecewise segment \( l \) for generator \( g \) (MW) (with \( \underline{P}^{0,g} = \underline{P}^g \)).
\( RD^g \) Ramp-down rate for generator \( g \) (MW/h).
\( RU^g \) Ramp-up rate for generator \( g \) (MW/h).
\( SD^g \) Shutdown ramp rate for generator \( g \) (MW/h).
\( SU^g \) Start-up ramp rate for generator \( g \) (MW/h).
\( TC^g \) Time down after which generator \( g \) goes cold (h).
\( \bar{T}^{S,g} \) Time offline after which the start-up category \( s \) is available (h) (with \( \bar{T}^{1,g} = DT^g, \bar{T}^{S,g} = TC^g \)).
\( UT^g \) Minimum up time for generator \( g \) (h).

Parameters: Second Stage
\( D^\omega_t \) Load (demand) at time \( t \) in scenario \( \omega \) (MW).
\( W^\omega_t \) Maximum power from renewables at time \( t \) in scenario \( \omega \) (MW).
\( \underline{W}^\omega_t \) Minimum power from renewables at time \( t \) in scenario \( \omega \) (MW).

Variables: First Stage
\( c^{SU,g}_t \) Start-up cost of generator \( g \) at time \( t \) (S).
\( u^g_t \) Commitment status of generator \( g \) at time \( t \), \( \in \{0, 1\} \).
\( v^g_t \) Start-up status of generator \( g \) at time \( t \), \( \in \{0, 1\} \).
\( w^g_t \) Shutdown status of generator \( g \) at time \( t \), \( \in \{0, 1\} \).
\( x^{[t,t']}_{g} \) Indicator arc for shutdown at time \( t \), start-up at time \( t' \), uncommitted for \( i \in [t, t') \), for generator \( g, \in \{0, 1\} \), \( [t, t') \) such that \( t + DT^g \leq t' \leq t + TC^g - 1 \).

Variables: Second Stage
\( p^{g,\omega}_t \) Power above minimum from generator \( g \) at time \( t \) in scenario \( \omega \) (MW).
\( \underline{P}^{g,\omega}_t \) Power above maximum from generator \( g \) at time \( t \) in scenario \( \omega \) (MW).
\( p^{g,\omega}_t \) Power below minimum from generator \( g \) at time \( t \) in scenario \( \omega \) (MW).
\( p^{l,g,\omega}_t \) Power from piecewise interval \( l \) for generator \( g \) at time \( t \) in scenario \( \omega \) (MW).
\( r^{n,\omega}_t \) Power from renewables at time \( t \) in scenario \( \omega \) (MW).
\( y^{g,\omega}_t \) Non-nominal operation status of generator \( g \) at time \( t \) in scenario (MW).
2.2 Mathematical programming model

We assume that the generator operating cost is increasing, piecewise linear and convex in \( p_{i}^{g,\omega} \) under normal operating conditions. Further, let \( G^{-1} = \{ g \in G \mid UT^{g} > 1 \} \) and \( G^{1} = \{ g \in G \mid UT^{g} = 1 \} \); i.e., \( G^{-1} \) and \( G^{1} \) denote the set of generators with an uptime greater than one and equal to one, respectively. We use the so-called “3-bin” formulation from (Knueven et al. 2020; Morales-España et al. 2013), with the ramping constraints from (Damcı-Kurt et al. 2016).

\[
\begin{align*}
\min & \sum_{g \in G} \sum_{t \in T} \left( \sum_{i \in L^{g}} \mathcal{E}[C^{I,g} p_{i}^{l,g,\omega} + \overline{C}^{g} p_{i}^{g,\omega} + C^{R,g} u_{t}^{g} + c_{t}^{SU,g}] \right) \\
\text{subject to:} & \\
& u_{t}^{g} - u_{t-1}^{g} = v_{t}^{g} - w_{t}^{g} \quad \forall t \in T, \forall g \in G \quad (2a) \\
& \sum_{t = t-UT^{g}+1}^{t} v_{t}^{g} \leq u_{t}^{g} \quad \forall t \in [UT^{g}, T], \forall g \in G \quad (2b) \\
& \sum_{t = t-DT^{g}+1}^{t-DT^{g}} u_{t}^{g} \leq 1 - u_{t}^{g} \quad \forall t \in [DT^{g}, T], \forall g \in G \quad (2c) \\
& \sum_{t' = t-TC^{g}+1}^{t+TC^{g}-1} x_{t'[t,t]}^{g} \leq v_{t}^{g} \quad \forall t \in T, \forall g \in G \quad (2d) \\
& \sum_{t' = t+DT^{g}}^{t+TC^{g}-1} x_{t'[t,t]}^{g} \leq w_{t}^{g} \quad \forall t \in T, \forall g \in G \quad (2e) \\
& c_{t}^{SU,g} = C^{S,g} v_{t}^{g} + \sum_{s=1}^{S-1} (C^{S,g} - C^{S,g}) \left( \sum_{t' = t-TC^{g}+1}^{t-T^{g}} x_{t'[t,t']}^{g} \right) \quad \forall t \in T, \forall g \in G \quad (2f) \\
& p_{i}^{l,g,\omega} \leq (\overline{F}^{g} - \underline{F}^{g}) u_{i}^{g} - (\overline{F}^{g} - SU^{g}) v_{i}^{g} - (\overline{F}^{g} - SD^{g}) w_{i}^{g} \quad \forall t \in T, \forall g \in G^{-1}, \forall \omega \in \Omega \quad (3a) \\
& p_{i}^{l,g,\omega} \leq (\overline{F}^{g} - \underline{F}^{g}) u_{i}^{g} - (\overline{F}^{g} - SU^{g}) v_{i}^{g} \quad \forall t \in T, \forall g \in G^{1}, \forall \omega \in \Omega \quad (3b) \\
& p_{i}^{g,\omega} \leq (\overline{F}^{g} - \underline{F}^{g}) u_{i}^{g} - (\overline{F}^{g} - SU^{g}) v_{i}^{g} \quad \forall t \in T, \forall g \in G^{1}, \forall \omega \in \Omega \quad (3c) \\
& p_{i}^{g,\omega} \leq (SU^{g} - RU^{g} - \underline{F}^{g}) v_{i}^{g} + RU^{g} u_{i}^{g} \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (3d) \\
& p_{i}^{g,\omega} \leq (SD^{g} - RD^{g} - \underline{F}^{g}) u_{i}^{g} + RD^{g} w_{i}^{g} \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (3e) \\
& p_{i}^{l,g,\omega} = \sum_{i \in L^{g}} p_{i}^{l,g,\omega} \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (3f) \\
& p_{i}^{l,g,\omega} \leq (\overline{F}^{g} - \overline{F}^{g-1}) u_{i}^{g} \quad \forall t \in T, \forall i \in L^{g}, \forall g \in G, \forall \omega \in \Omega \quad (3g) \\
& y_{i}^{g,\omega} = u_{i}^{g} - v_{i}^{g} - w_{i+1}^{g} \quad \forall t \in T, \forall g \in G^{1}, \forall \omega \in \Omega \quad (4a)
\end{align*}
\]
the stochastic bounds on the renewable power. Particular choice of this chance constraint in Section 2.3. The remaining constraints be reasonable as

\[
y_i^{g,\omega} \leq u_i^g - v_i^g \quad \forall t \in T, \forall g \in G^1, \forall \omega \in \Omega \quad (4b)
\]

\[
y_i^{g,\omega} \leq u_i^g - u_{i+1}^g \quad \forall t \in T, \forall g \in G^1, \forall \omega \in \Omega \quad (4c)
\]

\[
\bar{p}_i^{g,\omega} \leq (\bar{P} - \bar{P}' \bar{p}_i^{g,\omega}) \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (4d)
\]

\[
p_i^{g,\omega} \leq (P - P') \bar{p}_i^{g,\omega} \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (4e)
\]

\[
\sum_{g \in G} \left( p_i^{g,\omega} + \bar{p}_i^{g,\omega} - P_i^{g,\omega} + P_i^g u_i^g \right) + \nu_i^{\omega} = D_i^{\omega} \quad \forall t \in T, \forall \omega \in \Omega \quad (5)
\]

\[
\frac{1}{|G||T||\Omega|} \sum_{\omega \in \Omega} \sum_{g \in G} \sum_{t \in T} y_i^{g,\omega} \leq \varepsilon \quad (6)
\]

\[
p_i^{1^g,\omega} \in \mathbb{R}_+ \quad \forall t \in T, \forall l \in L^2, \forall g \in G, \forall \omega \in \Omega \quad (7a)
\]

\[
p_i^{g,\omega}, \bar{p}_i^{g,\omega}, \bar{P}_i^{g,\omega} \in \mathbb{R}_+ \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega \quad (7b)
\]

\[
r_i^{n,\omega} \in \left[ W_i^{n,\omega}, \bar{W}_i^{n,\omega} \right] \quad \forall t \in T, \forall n \in N, \forall \omega \in \Omega \quad (7c)
\]

\[
u_i^g, v_i^g, w_i^g \in \{0, 1\} \quad \forall t \in T, \forall g \in G \quad (7d)
\]

\[
x_{(t,t')}^g \in \{0, 1\} \quad \forall \{t, t'\} \in \mathcal{X}^g, \forall g \in G \quad (7e)
\]

\[
y_i^{g,\omega} \in \{0, 1\} \quad \forall t \in T, \forall g \in G, \forall \omega \in \Omega. \quad (7f)
\]

We introduce this model only briefly here, details are available in Morales-Espeña et al. (2013). Constraints (2) describe the use of the generator with its on and off variables and encode the start-up costs (which are directly substituted into the objective function via equation (2f)). Constraints (3) represent the generator’s start-up and shutdown requirements, ramping requirements, and piecewise power production. Constraint (5) ensures we meet the uncertain demand exactly with the thermal and renewable generators. Variable \(y_i^{g,\omega}\) is one when the generator \(g\) is operating in a non-nominal mode at hour \(t\) in scenario \(\omega\), and zero under normal operating conditions. Turning to constraints (4), first notice by constraints (4a)-(4c) that a generator can only be in non-nominal mode at time \(t\) if it is on \((u_t = 1)\), has been on \((u_t = 0)\), and will be on for at least one hour \((u_{t+1} = 0)\). Constraints (4d) and (4e) enforce that in a non-nominal mode the generator can produce power up to \(\bar{P}_i^g\) or down to \(\bar{P}_i^g\). Notice that the total output of the generator at \(t\) in scenario \(\omega\) is represented by the quantity \(\left( p_i^{g,\omega} + \bar{p}_i^{g,\omega} - P_i^{g,\omega} + P_i^g u_i^g \right) \). Finally, constraint (6) restricts the proportion of non-nominalities across all generators, times and scenarios to be no more than \(\varepsilon\). Here, \(\varepsilon\) is a small number less than one, such as 0.01 or 0.05. We explain the particular choice of this chance constraint in Section 2.3. The remaining constraints ensure the non-negativity and binary restrictions on the relevant decision variables and the stochastic bounds on the renewable power.

Because of the additional variables used to represent dispatch over \(\bar{P}_i^g\) and under \(\bar{P}_i^g\), this formulation implicitly relaxes the ramping requirements. We believe this to be reasonable as \(\bar{P}_i^g\) \((\bar{P}_i^g)\) will be not too different from \(\bar{P}_i^g\) \((\bar{P}_i^g)\), so the relaxation of the ramping requirements will be of the same scale as the relaxation of the operating limits.
In the two-stage stochastic program represented by model (1)–(7), we decide each generator’s operation status in the first stage. Then, we observe the uncertainty in the demand and wind power. After this, we decide the amount of power to employ from each generator that was declared to be “on” in the first stage. Under normal operating conditions, generator $g$ can use power up to $P^g$. This limit is increased to $\overline{P}^g$ under non-nominal operations. Similarly, if there is the potential for an over-generation event, the lower limit can be reduced from $P^g$ to $\underline{P}^g$. Also notice that power below minimum $p_{g,\omega}^g$ is penalized with a positive term in the objective function, and similarly for $\underline{p}_{g,\omega}^g$.

When the provided cost curve $C^{L,g}$ is convex, the overall production costs with non-nominalities will be convex when $C^g \geq C^{L,g}$ and $\underline{C}^g \geq -C^{1,g}$. The first condition says the marginal cost of providing non-nominal power by dispatching above $P^g$ is at least that of the generator operating at $P^g$, and the second says the marginal cost of dispatching below $P^g$ is at least that of being dispatched at $\overline{P}^g$. In practice we will consider $C^g > 0$ so as to compensate generators for dispatch below $P^g$.

During a non-nominality, we allow generator $g$ to produce power up to $\overline{P}^g > P^g$, at a cost greater than that during normal operations. Similarly, we allow generator $g$ to produce power down to $\underline{P}^g < P^g$, at cost $\underline{C}^g$. Given the lack of data, we selected parameters $\beta, \gamma > 0$ for use in our study:

$$\overline{P}^g = (1 + \beta)P^g$$
$$P^g = (1 - \beta)\underline{P}^g$$
$$\underline{C}^g = (1 + \gamma)C^{L,g}$$
$$\overline{C}^g = (1 + \gamma)\underline{C}^{L,g}.$$

### 2.3 Choice of the chance constraint

We begin with a few definitions, motivated by Prékopa (1988); Singh and Watson (2019). Define the event $E^g_t = \{ \omega : y_{g,\omega}^t = 0 \}$. Then, the complement of $E^g_t$ is $\overline{E}^g_t = \{ \omega : y_{g,\omega}^t = 1 \}$. Define the event $F = \{ \omega : \cap_{t \in T, g \in G} E^g_t \}$. Then, the complement of $F$ is $\overline{F} = \{ \omega : \cup_{t \in T, g \in G} \overline{E}^g_t \}$. Consider the following probabilistic constraints:

$$\mathbb{P}(F) \leq \varepsilon. \quad (8a)$$
$$\sum_{g \in G} \sum_{t \in T} \mathbb{P}(E^g_t) \leq \varepsilon. \quad (8b)$$

**Lemma 1** The constraint $(8b)$ implies the constraint $(8a)$.

**Proof** The proof follows from the definitions of $F$ and $\overline{E}^g_t$, and the classical probability inequality: $\mathbb{P}\left( \cup_{t \in T, g \in G} \overline{E}^g_t \right) \leq \sum_{g \in G} \sum_{t \in T} \mathbb{P}(E^g_t)$. \hfill \Box

Next, we construct a sample average approximation (SAA) of the above probabilistic constraints with $|\Omega|$ samples. To this end, we define a new binary variable: $z^\omega = 0 \forall \omega \in \Omega$. 
The constraints (8a) and (8b) are the SAA of the probabilistic constraints (8a) and (8b) are respectively:

\[
\frac{1}{|\Omega|} \sum_{\omega \in \Omega} z_\omega \leq \varepsilon. 
\]

\[
\frac{1}{|\Omega|} \sum_{g \in \mathcal{G}} \sum_{t \in T} \sum_{\omega \in \Omega} y_{g,t,\omega} \leq \varepsilon. 
\]

Equations (9b) and (10a) imply:

\[
\frac{1}{|\Omega| |T| |\mathcal{G}|} \sum_{g \in \mathcal{G}} \sum_{t \in T} \sum_{\omega \in \Omega} y_{g,t,\omega} \leq \varepsilon,
\]

which is weaker than constraint (10b). The above discussion can be summarized in the following lemma.

**Lemma 2** Under Eq. (9), the constraint (10b) implies the constraint (10a) which further implies the constraint (11).

In Sect. 4 we analyze the sensitivity of the stochastic solution to \( \beta, \gamma, \) and \( \varepsilon \). Larger values of \( \beta \) relax the nominal operating constraints further, and larger values for \( \gamma \) penalize deviation from the nominal operating constraints through the cost via the objective function.

### 3 Case study

To analyze the impact of considering bounded exceedances of generator ratings directly in a stochastic UC model, we consider two case studies. The first case study considers the WECC240++ system (Rachunok et al. 2018). WECC240++ UC instances have 85 thermal generating units. Demand profiles are based on real-world data from 2004. Wind generation scenarios are based on data from 2013, scaled to
achieve approximately 30% penetration levels. We selected one day in the 2013 simulation set, 11 May 2013, and consider 50 probabilistic scenarios and a time horizon of 48 hours at hourly resolution. Scenarios were constructed using the Markov-Chain Monte-Carlo procedure to model wind production; demand is treated as deterministic. Wind generation is fully curtailable. For additional details, the reader is referred to Rachunok et al. (2018).

The second case study is based on the recently introduced RTS-GMLC system (RTS-GMLC 2018; Barrows et al. 2020). The RTS-GMLC system has 73 thermal generators (oil, coal, gas, and nuclear) and 81 renewable generators (wind, hydro, utility-scale photo-voltaic, and rooftop photo-voltaic). Simulated load and renewable generation is provided on a 5-minute basis for the year 2020. We consider the date 10 July 2020 with a 48 hour planning horizon, at hourly resolution, and use 16 probabilistic wind scenarios obtained from Staid et al. (2017). All other data is considered deterministic for this study. We additionally assume hydro units are self-scheduling and the rooftop photo-voltaic is must-take. However, the wind and utility-scale photo-voltaic are fully curtailable. Hence, the aggregate renewables at each bus in the system is a mix of must-take and curtable resources.

4 Computational experiments

4.1 Computational setup

We encoded all models using the Pyomo 5.5 (Hart et al. 2017) algebraic modeling language. All models are solved using the commercial Gurobi 8.0.1 (Gurobi Optimization 2018) MILP solver, on a laptop comprising of a 2.8 GHz Intel Core i7 processor and 16 GB of RAM. We consider $\varepsilon = 0.01$ and 0.05, corresponding respectively to use of nominal generator operations at least 99% and 95%, across all scenarios, generators, and time periods. We attempt to solve all problems to a MILP optimality gap of 0.001 (i.e., 0.1%) within a time limit of 1800 seconds. Additionally, we set the Gurobi parameter `Method=1`, which dictates that dual simplex is to be used to solve the root LP relaxations. All other Gurobi solver parameter settings were preserved at their defaults. We report the MILP optimality gap at termination when a problem could not be solved within the 1800 second time limit. The expected cost at $\varepsilon = 0$ is the baseline for operations when no non-nominal states are permitted; i.e., standard stochastic UC. In this case, the values of $\beta$ and $\gamma$ are irrelevant. Setting $\varepsilon > 0$ allows for varying degrees of non-nominal operations. In addition to absolute cost, we report the percentage of savings relative to the base case ($\varepsilon = 0$). In general, this percentage represents the minimum percentage cost saved, as some problems could not be solved within the time limit, and in any case all models are only solved to a termination criteria of a 0.1% MILP gap.
4.2 Comparison of the two test systems

In Table 1 we present the results for the WECC-240++ case. All parameterizations of this case study were solved to 0.1% optimality gap within the time limit; this is reflected using “-” in the MILP gap column. In Table 2, we present analogous results for a single representative day for the RTS-GMLC case. Here, we do observe non-zero MILP gaps in some parameterizations of our model.

Both of our baseline case studies consider data from late spring and summer months in the US. Winter and fall months necessarily exhibit different load profiles. Next, we consider RTS-GMLC case results using data from the winter (for 10 January 2020). The results, shown in Table 3, indicate qualitatively identical behavior for our model.

Despite having a similar number of generators and fewer scenarios, the RTS-GMLC system is generally more computationally demanding than the WECC-240++ system. However, despite this difference, the qualitative performance of our model with respect to parameterizations of $\varepsilon$, $\beta$, and $\gamma$ across the two cases is identical. We observe that higher values of $\varepsilon$ result in lower operation costs, but at the expense of a larger compu-
Table 3  Computational results for the RTS-GMLC 16 scenario case for 10 January 2020. A blank in the “MILP gap (%).” column indicates that the instance was solved to the optimal tolerance level εβγ.

| ε   | β   | γ   | Cost (M$) | Savings (%) | Time (sec) | MILP gap (%) |
|-----|-----|-----|-----------|-------------|------------|--------------|
| 0   |     |     | 1.77      | 0.00%       | 18         | -            |
| 0.01| 0.05| 0.1 | 1.74      | 1.83%       | 41         | -            |
|     | 0.2 |     | 1.74      | 1.80%       | 41         | -            |
| 0.1 | 0.1 |     | 1.71      | 3.38%       | 583        | -            |
|     | 0.2 |     | 1.71      | 3.26%       | 317        | -            |
| 0.05| 0.05| 0.1 | 1.74      | 1.84%       | 46         | -            |
|     | 0.2 |     | 1.74      | 1.80%       | 49         | -            |
| 0.1 |     | 0.1 | 1.71      | 3.47%       | 286        | -            |
|     | 0.2 |     | 1.71      | 3.33%       | 308        | -            |

Sensitivity analysis of optimal solution

Next, we analyze the structure of an optimal solution (subject to the MILP optimality gap) to the RTS-GMLC system with ε = 0.01, β = 0.1, and γ = 0.1, in order to analyze differences relative to an optimal solution to the baseline stochastic UC model. Figure 1a presents the aggregate number of non-nominalities per scenario (i.e., ∑y_{g^,ω}^{t_1}). The total number of non-nominalities is 560, which is 1% of |G||Ω|. Clearly, the model incentivizes choosing a lower cost generator whenever possible. As a result, the sole nuclear generator in the RTS-GMLC system—which is the cheapest marginal unit—enters a non-nominal mode in all of the 16 scenarios for at least one hour in the operating horizon. We further observe (not shown) that this non-nominality is largely consistent in timing across scenarios; i.e., the nuclear
generator is in a non-nominal operations mode for almost all of the same hours across all the 16 scenarios. While nuclear generators tend to have low marginal costs, in practice, they are operated in narrow windows and are subject to rigorous oversight. Thus, we conducted another analysis to see the impacts if the nuclear generator were not allowed to run at all in a non-nominal mode. Figure 1b presents the aggregated non-nominalities per scenario for this case. The variability in the non-nominalities per scenario is notably larger in Fig. 1b than Fig. 1a. Specifically, the standard deviation in the number of non-nominalities across the 16 scenarios is 20.2 for Fig. 1b and 14.8 for Fig. 1a. While the cost savings relative to the stochastic solution for Fig. 1a is 1.51%, the cost savings for the solution represented in Fig. 1b is only 1.15%, indicating that shifting of non-nominal operations to more expensive units does decrease the overall cost benefit. This is, again, expected as the latter solution is suboptimal.

To examine the sensitivity of cost savings relative to any one generator, we now consider the instance analyzed in Table 2 under another restrictive operational regime; the results are shown in Table 4. The first considers only allowing units to activate non-nominal operations once per day; i.e., adds the following constraints

\[ \sum_{t=1}^{24} y_{g,\omega}^{t} \leq 1 \quad \forall g \in G, \forall \omega \in \Omega \]
Table 5  Computational results for the RTS-GMLC 16 scenario case for 11–16 July 2020. A blank in the “MILP gap (%)” column indicates that the instance was solved to the optimal tolerance level

| Date          | ε  | β  | γ   | Cost (M$) | Savings (%) | Time (sec) | MILP gap (%) |
|---------------|----|----|-----|-----------|-------------|------------|--------------|
| 11-July-2020  | 0  |    |     | 4.05      | 0.00%       | 24         |              |
|               | 0.01 | 0.05 | 0.1 | 4.04      | 0.25%       | 419        |              |
|               | 0.1  | 0.1  |     | 4.01      | 0.97%       | 553        |              |
| 12-July-2020  | 0  |    |     | 3.57      | 0.00%       | 29         |              |
|               | 0.01 | 0.05 | 0.1 | 3.54      | 0.87%       | 947        |              |
|               | 0.1  | 0.1  |     | 3.51      | 1.59%       | 1800       | 0.26%        |
| 13-July-2020  | 0  |    |     | 3.55      | 0.00%       | 68         |              |
|               | 0.01 | 0.05 | 0.1 | 3.50      | 1.43%       | 363        |              |
|               | 0.1  |     |     | 3.45      | 2.97%       | 941        |              |
| 14-July-2020  | 0  |    |     | 3.37      | 0.00%       | 29         |              |
|               | 0.01 | 0.05 | 0.1 | 3.30      | 1.81%       | 320        |              |
|               | 0.1  | 0.1  |     | 3.28      | 2.61%       | 882        |              |
| 15-July-2020  | 0  |    |     | 3.06      | 0.00%       | 21         |              |
|               | 0.01 | 0.05 | 0.1 | 3.01      | 1.86%       | 287        |              |
|               | 0.1  | 0.1  |     | 2.96      | 3.57%       | 527        |              |
| 16-July-2020  | 0  |    |     | 4.33      | 0.00%       | 41         |              |
|               | 0.01 | 0.05 | 0.1 | 4.27      | 1.53%       | 879        |              |
|               | 0.1  | 0.1  |     | 4.23      | 2.37%       | 1252       |              |

\[
\sum_{t=25}^{48} y_{t}^{g,\omega} \leq 1 \quad \forall g \in \mathcal{G}, \forall \omega \in \Omega
\]

to the model (1)–(7). Results for this regime are reported in the column labeled “Limited”. The second, mentioned above, disables non-nominal operation for the nuclear unit in this system, and is reported in the column labeled “No nuclear”. The “Optimal” column reports the savings achieved using the non-modified model (same as Table 2). We note that we can still observe cost savings on the order of 1% in both cases, which is particularly surprising for the much more conservative Limited case. Hence, despite placing these additional restrictions on when non-nominal generation can be dispatched, significant cost savings can still be realized.

To analyze the generalizability of our observed cost savings, we next solve our model for another six days of the RTS-GMLC system. Table 5 reports these results, for \( \varepsilon = 0.01, \beta = 0.05, 0.1, \gamma = 0.1 \). Of the 12 instances considered, only one could not be solved to within the specified MILP optimality gap in the 1800 second time limit. The percentage cost savings do not differ qualitatively from those in Table 2, which were generated under the same parameterization. The average and standard deviation across the six days in terms of percentage cost savings for the \( \beta = 0.05 \) (\( \beta = 0.1 \)) case are 1.29\% (2.34\%) and 0.56\% (0.86\%).
Table 6  Computational results for the WECC240++ 10 and 100 scenario test cases for 11 May 2013. A blank in the “MILP gap (%)” column indicates that the instance was solved to the optimal tolerance level. For instances that could not be solved within the time limit the MILP gaps are large, and thus the savings could not be accurately computed (denoted by “n/a”)

| ε     | β   | γ  | Cost (M$) | Savings (%) | Time (sec) | MILP gap (%) |
|-------|-----|----|-----------|-------------|------------|--------------|
| (a) 10 scenarios |
| 0     | 0.01| 0.05| 63.02     | 0.00%       | 69         | –            |
| 0.01  | 0.1 | 0.05| 62.85     | 0.28%       | 96         | –            |
| 0.2   | 0.1 | 0.05| 62.85     | 0.27%       | 95         | –            |
| 0.1   | 0.2 | 0.05| 62.68     | 0.55%       | 82         | –            |
| 0.2   | 0.2 | 0.05| 62.69     | 0.53%       | 84         | –            |
| 0.05  | 0.05| 0.1 | 62.53     | 0.78%       | 87         | –            |
| 0.2   | 0.2 | 0.05| 62.54     | 0.76%       | 88         | –            |
| 0.1   | 0.1 | 0.2 | 62.02     | 1.60%       | 95         | –            |
| 0.2   | 0.2 | 0.2 | 62.08     | 1.49%       | 89         | –            |
| (b) 100 scenarios |
| 0     | 0.01| 0.05| 64.31     | 0.00%       | 473        | –            |
| 0.01  | 0.1 | 0.05| 64.10     | 0.33%       | 992        | –            |
| 0.2   | 0.2 | 0.05| 64.11     | 0.32%       | 917        | –            |
| 0.1   | 0.1 | 0.05| 63.95     | n/a         | 1800       | 65.7%        |
| 0.2   | 0.2 | 0.05| 63.95     | 0.56%       | 1050       | –            |
| 0.05  | 0.05| 0.1 | 63.75     | 0.87%       | 1023       | –            |
| 0.2   | 0.2 | 0.05| 63.79     | 0.82%       | 1259       | –            |
| 0.1   | 0.1 | 0.2 | 63.24     | 1.67%       | 989        | –            |
| 0.2   | 0.2 | 0.2 | 63.31     | 1.56%       | 1371       | –            |

We next comment on the economic implications for our study. While saving operational costs is clearly desirable, our proposed non-nominality UC would also incentivize less expensive generators to relax their operating constraints, which would shift revenue from more expensive peaking units. However, this effect may not be so determinative if a more restrictive policy such as Limited is considered. Future extensions could consider a fixed cost for operating a generator in non-nominal mode or a maximum up-time for non-nominal mode, so as to fully capture the additional costs or restrictions for operating outside of non-nominal mode. In the context of stochastic UC, wide variation in renewables generation and/or load across scenarios may lead to solutions that are overly conservative. That is, there may be generators that are dispatched to provide power only for an event with low probability; see, e.g., Rachunok et al. (2018). Although we did not examine this for this study, the use of non-nominal modes could serve as an inexpensive way to add additional peaking capacity to a system so as to enable higher renewables penetration.

We observe that all of our computational experiments were conducted on a modest laptop. The problems are generally tractable even without any custom algorithms (e.g., decomposition) for the number of scenarios considered above. An ISO or a util-
ity would likely run the optimizations on larger machines and with a greater number of scenarios. In this context, we next test our stochastic UC model by using (i) a small batch of 10 scenarios and (ii) a large batch of 100 scenarios, both for the WECC-240++ case for 11 May 2013. Tables 6a and 6b report results analogous to Table 1, respectively, for the 10 and 100 scenario variants. These computational results reinforce two intuitive observations. First, as expected, the respective cost values do not differ significantly between the three instances, i.e., with 10, 50, and 100 scenarios. Second, the computational effort required increases with an increase in the number of scenarios. However, only one of the instances in Table 6b could not be solved within the 1800 second time limit. The 100 scenario instances have approximately half a million binary variables after Gurobi’s presolve. Future research could examine specialized algorithms that would further assist in tractability, as well as examining larger test cases with hundreds or thousands of generators.

5 Conclusion

We presented a chance-constrained unit commitment formulation to incorporate small violations of the technical ratings of a generator. Our motivation comes from the fact that system operators occasionally run generators in non-nominal conditions; this can create significant economical benefits, especially in the context of stochastic unit commitment with a wide diversity in scenarios. The model can also be useful when there is an outage, either unplanned or planned, since system operators often analyze contingencies on a case-by-case basis (Power Water 2017). In our study, we demonstrated a small percentage savings in the costs which could translate to a significant amount of dollars saved over the year; see, also (O’Neill 2007). We analyzed the sensitivity of the optimal solution under restricted regimes as well. The models we presented are generally tractable, however future work could examine tailored algorithms to achieve even faster solutions.

Acknowledgements

Open Access funding provided by Projekt DEAL. This work was completed while the authors were affiliated with Sandia National Laboratories, and supported by Sandia’s Laboratory Directed Research and Development (LDRD) program. The authors thank Sandia for their time.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Anjos MF, Conejo AJ (2017) Unit commitment in electric energy systems, vol. 1, pp. 220–310. Now Publishers, Inc
Queyranne M, Wolsey LA (2017) Tight MIP formulations for bounded up/down times and interval-dependent start-ups. Math Program 164(1–2):129–155
Rachunk B, Staid A, Watson JP, Woodruff DL, Yang D (2018) Stochastic unit commitment performance considering Monte Carlo wind power scenarios. In: 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–6. IEEE
RTS-GMLC (2018) Reliability test system - grid modernization lab consortium. https://github.com/GridMod/RTS-GMLC. [Online; accessed 19-July-2008]
Silbernagl M (2016) A polyhedral analysis of start-up process models in unit commitment problems. Ph.D. thesis, Technische Universität München
Singh B, Watson J (2019) Approximating two-stage chance-constrained programs with classical probability bounds. Optim Lett 13:1403–1416. https://doi.org/10.1007/s11590-019-01387-z
Singh B, Morton DP, Santoso S (2018) An adaptive model with joint chance constraints for a hybrid wind-conventional generator system. Comput Manag Sci 15:563–582
Staid A, Watson JP, Wets RJB, Woodruff DL (2017) Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators. Wind Energy 20(12):1911–1925
System Operations Division, PJM (2019) PJM manual 13: Emergency operations revision: 68 effective date: January 1, 2019. https://www.pjm.com/~/media/documents/manuals/m13.ashx. [Online; accessed January 15, 2019]
Takriti S, Krasenbrink B, Wu LSY (2000) Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. Oper Res 48(2):268–280
van Ackooij W, Lopez ID, Frangioni A, Lacalandra F, Tahanan M (2018) Large-scale unit commitment under uncertainty: an updated literature survey. Ann Oper Res 271(1):11–85
Wang B, Hobbs BF (2014) A flexible ramping product: Can it help real-time dispatch markets approach the stochastic dispatch ideal? Electr Power Syst Res 109:128–140
Wang Q, Guan Y, Wang J (2012) A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. IEEE Trans Power Syst 27(1):206–215
Watson JP, Wets RJ, Woodruff DL (2010) Scalable heuristics for a class of chance-constrained stochastic programs. INFORMS J Comput 22(4):543–554
Wu H, Shahidehpour M, Li Z, Tian W (2014) Chance-constrained day-ahead scheduling in stochastic power system operation. IEEE Trans Power Syst 29(4):1583–1591
Wu L, Shahidehpour M (2010) Accelerating the benders decomposition for network-constrained unit commitment problems. Energy Syst 1(3):339–376
Zhao C, Wang J, Watson JP, Guan Y (2013) Multi-stage robust unit commitment considering wind and demand response uncertainties. IEEE Trans Power Syst 28(3):2708–2717
Zhao C, Wang Q, Wang J, Guan Y (2014) Expected value and chance constrained stochastic unit commitment ensuring wind power utilization. IEEE Trans Power Syst 29(6):2696–2705

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.