Heterogeneous Unit Clustering for Efficient Operational Flexibility Modeling

Bryan S. Palmintier, *Member, IEEE*, and Mort D. Webster, *Member, IEEE*

***Abstract*— Designing future capacity mixes with adequate flexibility requires capturing operating constraints through an embedded unit commitment approximation. Despite significant recent improvements, such simulations still require significant computation times. Here we propose a method, based on clustering units, for approximate unit commitment with dramatic improvements in solution time. This method speeds computation by aggregating similar but non-identical units. This replaces large numbers of binary commitment variables with fewer integers while still capturing individual unit decisions and constraints. We demonstrate the trade-off between accuracy and run-time for different levels of aggregation. A numeric example using an ERCOT-based 205-unit system illustrates that careful aggregation introduces errors of 0.05-0.9% across several metrics while providing several orders of magnitude faster solution times (400x) compared to traditional binary formulations. Further aggregation increases errors slightly (~2x) with further speedup (2000x). We also compare other simplifications that can provide an additional order of magnitude speed-up for some problems.**

***Index Terms*—Integer programming, Power generation scheduling, Power system modeling, Unit commitment, Flexibility, Capacity Expansion.**

Nomenclature

A. Indices

g, G	Generating unit, set of units
$\hat{g}$, $\hat{G}$	Generation cluster, set of clusters
t, τ, T	Time period
ρ	Reserve category {1,2,3}
$\hat{}$	Indicates clustered variable/parameter

B. Variables

C^{total}	Total system cost [$]
$C^{var}_{g,t}$	Variable costs [$]
$C^{start}_{g,t}$	Startup costs [$]
$P_{g,t}$	Power output [MWh]
$U_{g,t}$, $\hat{U}_{\hat{g},t}$	unit commitment state: 0/1, clustered integer
$S_{g,t}$, $\hat{S}_{\hat{g},t}$	startup indicator: individual, clustered
$D_{g,t}$, $\hat{D}_{\hat{g},t}$	shutdown indicator: individual, clustered
$R^{1,up}_{g,t}$	Primary (regulation) up reserves
$R^{1,down}_{g,t}$	Primary (regulation) down reserves
$R^{2,up}_{g,t}$	Secondary (load follow) up reserves
$R^{2,down}_{g,t}$	Secondary (load follow) down reserves
$R^{3}_{g,t}$	Tertiary reserves (quick start)

C. Parameters

c^{fuel}_{g}	Fuel costs [$/mmbtu]
$c^{varO\&M}_{g}$	Variable O&M costs [$/mmbtu]
$F_g(P_g)$	Heatrate (Fuel use) function [mmbtu/MWh]
f^{start}_{g}	Fuel usage at startup [mmbtu/start]
l_t	Load [MWh]
p^{min}_{g}	Minimum power output [MWh]
p^{max}_{g}	Maximum power output [MWh]
Δp^{down}_{g}	Maximum down-ramp rate [MWh/hr]
Δp^{up}_{g}	Maximum up-ramp rate [MWh/hr]
$r^{1,up}$	Primary up reserve load fraction
$r^{1,down}$	Primary down reserve load fraction
$r^{2,up}$	Secondary up reserve load fraction
$r^{2,down}$	Secondary down reserve load fraction
r^{outage}	Contingency reserve load fraction
$r^{replace}$	Tertiary reserve load fraction
x^{nosync}	Fraction of secondary reserves from offline
$a^{\rho,dir}_{g}$	Reserve capability by direction [per unit]
$a^{quickstart}_{g}$	0/1 quick start ability
$m^{up/down}_{g}$	Minimum up- or down-time [hrs]

This work was supported in part by the U.S. National Science Foundation under Grants 1128147 and 0835414.

B. S. Palmintier (e-mail: bryanp@ieee.org) and M. D. Webster (e-mail: mort@mit.edu) are with the Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

I. Introduction

Growth in variable renewables and other advanced power system technologies has prompted a need for researchers to capture operational flexibility in a range of models. Operational flexibility requires a balance between 1) requirements due to uncertainty (e.g., forecast errors and outages) and fluctuations (e.g., demand and wind ramps) and 2) limitations, typically from thermal generator technical constraints (e.g., minimum output levels, startup/shutdown limits, maximum ramping, etc.). In flexibility studies, variability and constraints are typically captured using unit commitment (UC) models [1-22].

Since some of the early pioneering work in UC models [1-5], there have been significant contributions in reformulating unit commitment models to appropriately represent variable generation and its impacts on reserve requirements and operations within an existing capacity mix [6-17]. Much of this work has been to develop improved algorithms for stochastic unit commitment, inclusion of transmission and security constraints, and the use of these models to develop optimal reserve allocation rules and to economically value the additional reserve requirements from renewables. In addition, there have been initial attempts to consider long-term flexibility needs within capacity planning models [18-21].

However, the unit commitment (UC) problem is by itself computationally intensive to solve because of the combination of the large number of discrete (binary) on/off decisions – one for each generator for each time period – and the number and complexity of the technical constraints. To include the full UC problem within a long-term planning model would make the solution of these models infeasible because the UC subproblem has to be solved for many alternative capacity mixes considered, often over a long timeframe (e.g., one year). What is needed within capacity expansion models is a simplified approximation of the UC problem that captures its large-scale features. The importance of representing UC within generation planning has been well demonstrated by Shortt, et al. [22].

This paper explores a range of approximations to the UC problem. Because our goal is to study future potential systems that do not yet exist, we neglect many of the details required for near-term operations models of current systems. In particular, we assume a well-developed transmission system and only consider security constraints through generation reserves. Moreover, the nuances of generator-specific characteristics, may not be necessary. Indeed, for any units not yet built only generalized data may be available. Relaxing the need for unit-specific data enables the aggregation or clustering of similar generation units. This transforms the large number of binary commitment variables to far fewer integer variables, and thereby drastically reduces the problem size and corresponding run times [18], [23], [24].

The concept of aggregating *identical* units is not new. As early as 1966, pioneering studies in computer based unit commitment, grouped identical generators together to illustrate simple solution techniques with limited computer hardware [25]. More recently, examples of combining identical units has also appeared in the literature. For example, Gollmer, et al. [23] also use grouped integer commitment for identical thermal plants and Garcia-Gonzalez, et al. [26] use a grouped integer on/off state when modeling banks of identical hydro turbines for optimal combined bidding with wind. Likely other implementations using such homogeneous clustering remain unpublished, since the computational advantages of binary aggregation to integers is well recognized in the operations research community [27]. For example in his dissertation, Cerisola describes homogeneous aggregation into "generalized" units with integer commitment variables [28], yet this formulation is not described in related journal articles [29]. When clustering identical, co-located units, clustering can provide identical solutions in faster times.

The concept of *heterogeneous* clustering extends this aggregation such that similar, but not identical, units are clustered together and assigned an integer commitment state. Conceptually, this approach is similar to that of Sen and Kothari [30], who also group units. However, their treatment assumes a binary commitment state for the entire group: all on or all off. This is computationally helpful, but is much less flexible than an integer formulation that allows some of the generators within a group to run while others are off. The all or none approach also prevents properly computing startup costs, minimum output levels, and reserve capability.

Recent work on heterogeneous clustering has demonstrated efficient unit-commitment-based computations over long time horizons (e.g. full year as 8760 sequential hours) as part of price estimation [24] and planning studies [18]. However, in both efforts, heterogeneous clusters are simply used to make the study tractable, without considering different clustering approaches or comparing the results to a full binary formulation.

A key contribution of this paper is to explore the trade-offs among accuracy, run-time and level of aggregation used in heterogeneous clustering. To do so, we introduce a set of performance metrics applicable to a wide range of decision objectives. We also compare the performance of other simplifying long-term UC assumptions with and without clustering.

In addition, this paper presents a streamlined implementation for clustered minimum up and down time, which uses only one integer variable per cluster. As described in [24] and [28], these dynamic, inter-period constraints require careful consideration since within the same cluster, some units may startup or shutdown while others continue to run. In the past, these constraints have been converted back to binary [28], not described fully [23], or are not relevant because non-thermal units are aggregated [26]. Langrene et al. describe heterogeneous clustered dynamic constraints in detail, but their minimum up/down formulation requires multiple integer variables per cluster per time period [24]. These separate integers explicitly represent startup/shutdown generator states: running but stoppable, must keep running, stopped but able to start, and must stay stopped. As described below, our formulation uses sums of existing continuous startup and shutdown variables to maintain only a single integer unit commitment variable per cluster, thereby further reducing the problem size.

In Section II, we formulate a standard binary UC model that we use to compare against our clustered formulation. We present the clustered formulation of the UC model in Section III. Section IV describes alternative speedup strategies used in the literature to which we compare our approach. The experimental setup and error metrics are defined in Section V. Section VI describes the 205-generator unit test system and Section VII presents the results. Section VIII provides a concluding discussion.

II. Traditional Unit Commitment

A. Core model

The generic unit commitment problem finds the minimum cost combination of generator commitment and power output to meet demand over time. Here we linearize the standard basic formulation [31], [32], for a thermal-only system. The resulting optimization problem is a large mixed-integer linear program (MILP) that can then be solved by powerful commercial solvers as is done by a growing number of power system operators [33]. For clarity, we use uppercase for variables, bold upper case for sets, and lowercase for parameters and set elements.

1) The Objective Function minimizes total operations costs:

$$C^{total} = \min \sum_{g\in\boldsymbol{G}} \sum_{t\in\boldsymbol{T}} \left(C^{var}_{g,t} + C^{start}_{g,t}\right) \tag{1}$$

computed as the sum of variable costs, $C^{var}_{g,t}$, and startup costs, $C^{start}_{g,t}$, for all units, g, and time periods, t.

2) The Variable Costs, $C^{var}_{g,t}$ include fuel usage as a function, F_g, of the instantaneous power output, $P_{g,t}$, times fuel costs, c^{fuel}_g; and variable operations and maintenance (O&M) costs, $c^{varO\&M}_g$:

$$\begin{aligned} &C^{var}_{g,t} = F_g(P_{g,t})c^{fuel}_g + P_{g,t}c^{varO\&M}_g \\ &\text{with} \quad P_{g,t} \geq 0, \end{aligned} \tag{2}$$

3) The Startup Costs, $C^{start}_{g,t}$, assume a constant fuel use per startup, f^{start}_g:

$$C^{start}_{g,t} = S_{g,t} f^{start}_g c^{fuel}_g \tag{3}$$

Startup events, $S_{g,t}$, are computed using the state equation:

$$U_{g,t} = U_{g,t-1} + S_{g,t} - D_{g,t} \tag{4}$$

$$\text{with} \quad U_{g,t} \in \{0,1\} \tag{5a}$$

Here $U_{g,t}$ represents the commitment (on/off) state of each unit, $S_{g,t}$ represents startup events, and $D_{g,t}$ represents unit shut down[1].

We note that (3) deviates from startup formulations that distinguish warm and cold startup costs, e.g. [31]. This constant startup cost simplification is commonly used for this class of long-term unit commitment problem [17], [20], [35].

4) The System Balance Constraint ensures that the sum of instantaneous power, $P_{g,t}$, equals total load, l_t, at all times:

$$\sum_{g\in\boldsymbol{G}} P_{g,t} = \ l_t \qquad\qquad \forall t \in \boldsymbol{T} \tag{6}$$

[1] In this formulation, $S_{g,t}$ and $D_{g,t}$ are continuous variables that will be forced to take on integer variables by (4). Recent work by Ostrowski, et al. [34] has shown improved performance with modern MILP solvers by constraining these variables as integers. The results reported here were run before [34] was published. Limited testing, has confirmed that these performance improvements hold with clustering, but show the same relative performance as presented in detail here.

5) Unit Minimum and Maximum Output Constraints use the binary commitment variable to imply that each generating unit is either off and outputting zero power ($U_{g,t} = 0$), or on and running within its operating limits, p^{min}_g and p^{max}_g ($U_{g,t} = 1$):

$$U_{g,t}p^{min}_g \leq P_{g,t} \leq U_{g,t}p^{max}_g \tag{7a}$$

B. Additional Constraints

A more realistic model includes additional cost components and generator and system reliability imposed technical constraints [32]. We focus on the most common extensions:

1) Ramping Limits capture limitations on how fast thermal units can adjust their output power:

$$\begin{aligned} P_{g,t-1} - P_{g,t} \leq \mathrm{U_{g,t}} \Delta p^{downmax}_g \\ + \max\left(p^{min}_g, \Delta p^{downmax}_g\right) D_{g,t} \end{aligned} \tag{8a}$$

$$\begin{aligned} P_{g,t} - P_{g,t-1} \leq \mathrm{U_{g,t}} \Delta p^{upmax}_g \\ + \max\left(p^{min}_g, \Delta p^{upmax}_g\right) S_{g,t} \end{aligned} \tag{9a}$$

where the Δp's are the ramp limits up or down.

2) Minimum Up and Down Times are modeled using the most computationally efficient formulation from [36], [37], with m^{up}_g and m^{down}_g for minimum up and down times:

$$U_{g,t} \geq \sum_{\tau=t-m^{up}_g}^{t} S_{g,\tau} \tag{10}$$

$$1 - U_{g,t} \geq \sum_{\tau=t-m^{down}_g}^{t} D_{g,\tau} \tag{11a}$$

3) Operating Reserves: Because power generated on the grid must match demand instantaneously, a number of operating reserves are maintained by allowing room between generator output levels and corresponding limits to provide on-line capacity able to quickly increase (or decrease) and compensate for generation or transmission outages, forecast errors, etc.:

a) *Primary, or regulation, reserves,* operate at the few second timescale to compensate for rapid stochastic changes:

$$\sum_{g\in\boldsymbol{G}} R^{1,up}_{g,t} \geq \ r^{1,up} l_t \tag{12}$$

$$\sum_{g\in\boldsymbol{G}} R^{1,down}_{g,t} \geq \ r^{1,down} l_t \tag{13}$$

where $R^{1,up}_{g,t}$ and $R^{1,down}_{g,t}$ are the quantities of primary reserves supplied by unit *g* in time period *t*. The totals of which must exceed the exogenously determined system-level frequency reserve requirements, $r^{1,up}$ and $r^{1,down}$.

b) *Secondary reserves* operate on the few minute timescale for both contingencies (spinning reserves) and load following. We allow a fraction of the reserve up supply, x^{nosync}, to be supplied by non-synchronized resources such as offline quick starting units or demand response.:

$$\sum_{g\in\boldsymbol{G}} R^{2,up}_{g,t} \geq (r^{2,up} l_t + r^{outage})(1 - x^{nosync}) \tag{14}$$

$$\sum_{g\in\boldsymbol{G}} R^{2,down}_{g,t} \geq \ r^{2,down} l_t \tag{15}$$

The $R_{g,t}$'s are the quantity of on-line secondary reserves supplied by each unit. $r^{2,up}$ and $r^{2,down}$ are the system load following requirements, a function of load/wind forecast error. r^{outage} is the additional reserve required for contingencies, typically set to the largest unit or transmission tie capacity.

c) Tertiary or quick start reserves are off-line but ready to run units that can be brought on-line quickly when needed:

$$\sum_{g\in \boldsymbol{G}}\left(R_{g,t}^{3}+R_{g,t}^{2,up}\right) \geq r^{2,up}l_t + r^{outage} + r^{replace} \tag{16}$$

The left-hand side includes both tertiary and secondary up reserves both to capture the fraction of the secondary reserve allowed by (15) from off-line units, and to enable tertiary reserves to be met by on-line units when appropriate.

d) *Unit reserve capabilities* are dictated by the units ability to provide each type of reserve, a_g^{ρ}:

$$R_{g,t}^{\rho,dir} \leq a_g^{\rho,dir} p_g^{max} \quad \forall \rho \in \{1,2\}, dir \in \{up, down\} \tag{17a}$$

For tertiary reserves, quick start capable units can only be drawn from the pool of non-active units:

$$R_{g,t}^{3} \leq (1-U_{g,t}) a_g^{quickstart} p_g^{max} \tag{18a}$$

where $a_g^{quickstart}$ represents the fraction of the unit capacity, p_g^{max}, that can be deployed fast enough.

e) Updated unit output constraints capture the need for a unit to run below maximum for upward and above minimum for downward reserves. This replaces (7a) with the pair:

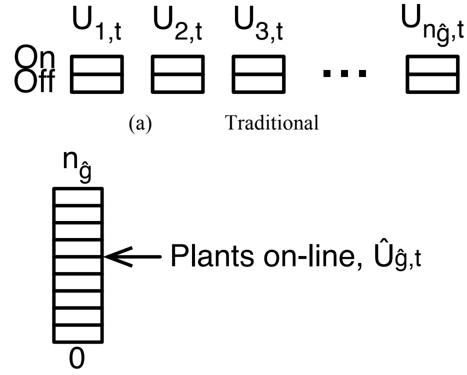
$$\begin{aligned} P_{g,t} &\geq U_{g,t}p_g^{min} + R_{g,t}^{1,down} + R_{g,t}^{2,down} \\ U_{g,t}p_g^{max} &\geq P_{g,t} + R_{g,t}^{1,up} + R_{g,t}^{2,up} \end{aligned} \tag{7b}$$

III. Clustered Unit Commitment

A. The Concept of Clustering

As described in the introduction, for problems with simplified or ignored transmission constraints, it is possible to combine similar generating units into clusters. As seen in Fig. 1, this replaces the large set of *binary* commitment decisions, one for each unit, with a smaller set of *integer* commitment states, one for each cluster. All of the other variables – such as power output level, reserves contribution, etc. – and constraints are then aggregated for the entire cluster. Within the cluster, however, the integer commitment variable still captures individual unit level relations.

Computationally, the integer variables provide structure that both reduces the dimensionality of and guides the search through the combinatorial commitment state space by eliminating identical or very similar permutations of binary commitment decisions. The number of possible discrete combinations of commitment variables with the traditional formulation scales exponentially as 2^N with the number of units *N*, while clustering scales as the product of the cluster sizes: $\prod n_{\hat{g}}$. For example, a system with 100 units clustered into three groups of sizes {10, 70, 20} would reduce the number of discrete combinations in each time period from $\sim 10^{30}$ to $\sim 10^{4}$.[2] In addition, clustering reduces the number of continuous equations and variables since all relations now apply over the smaller number of clusters rather than the full set of individual units.



(a) Traditional

(b) Clustered

Fig. 1. Conceptual comparison between traditional and clustered unit commitment for a single type of unit in a single time period. In the traditional formulation (a), each unit has a separate binary commitment variable, $U_{g,t}$. With clustering (b), the entire cluster of $n_{\hat{g}}$ units has only a single integer commitment variable, $\widehat{U}_{\hat{g},t}$.

B. Clustering Formulation

Mathematically, little of the traditional formulation changes with clustering. The key exceptions are replacing the individual unit index, g, with the cluster identifier, $\hat{g}$, and using a corresponding integer commitment variable, $\widehat{U}_{\hat{g}}$:

$$\widehat{U}_{\hat{g}} \in \{0,1,\dots,n_{\hat{g}}\} \tag{5b}$$

As before, the startup and shutdown variables, now $\hat{S}_{\hat{g}}$ and $\widehat{D}_{\hat{g}}$, continue to take continuous values, and are constrained by (4) and (5a) to take only positive integer values. The combination of minimum up/down time constraints and non-zero startup costs discourages simultaneous startup and shutdown within a cluster.

1) Relations With No Change Needed. Beyond this substitution no further changes are required for the objective (1), variable costs (2), startup costs (3), commitment state (4), system balance (6), unit output constraints (7b), minimum up time (10), and system reserve requirements (12) – (16).

2) Ramping Limits require the most extensive changes since hour-to-hour output for the entire cluster must account for units that start up, $\hat{S}_{\hat{g},t}$, and shut down, $\widehat{D}_{\hat{g},t}$. The ramp rates for on-line generators also scale by the number of plants actually on-line, $\hat{U}_{\hat{g},t}$. These modify (8a) & (9a) to:

$$\begin{aligned} P_{\hat{g},t-1} - P_{\hat{g},t} \leq \left(\widehat{U}_{\hat{g},t} - \hat{S}_{\hat{g},t}\right)\Delta p_{\hat{g}}^{downmax} - p_{\hat{g}}^{min}\hat{S}_{\hat{g},t} \\ + \max\left(p_{\hat{g}}^{min}, \Delta p_{\hat{g}}^{downmax}\right) D_{\hat{g},t} \end{aligned} \tag{8b}$$

$$\begin{aligned} P_{\hat{g},t} - P_{\hat{g},t-1} \leq \left(\widehat{U}_{\hat{g},t} - \hat{S}_{\hat{g},t}\right)\Delta p_{\hat{g}}^{upmax} - p_g^{min}\widehat{D}_{\hat{g},t} \\ + \max\left(p_{\hat{g}}^{min}, \Delta p_{\hat{g}}^{upmax}\right) S_{\hat{g},t} \end{aligned} \tag{9b}$$

In these relations, the first term on the right includes the core units that run in both time periods, the second corrects for startup/shutdowns to prevent artificial inflation of the ramping limits for the core units, and the third captures the allowable extra change in cluster production due to shutdown/startup.

3) Minimum Up and Down Time. Interestingly enough, the most efficient formulation for up and down time described in

[2] Modern MILP solvers use sophisticated branch-and-cut algorithms to explore only a tiny fraction of this combinatorial space. Still, the speedup with reduced dimensionality can be dramatic.

[36], [37] adapts to clustering with minimal changes. The formulation is based on the sum of starts or stops during the minimum up or down period which readily adapts to integer, rather than binary, commitment states. The only change required calculates the number of units currently off as the difference between $n_{\hat{g}}$ (as opposed to one) and the current commitment state, $\hat{U}_{\hat{g},t}$:

$$n_{\hat{g}} - \hat{U}_{\hat{g},t} \geq \sum_{\tau=t-m_{\hat{g}}^{mindown}}^{t} D_{\hat{g},\tau} \tag{11b}$$

4) Reserve capabilities change similarly to:

$$R_{\hat{g},t}^{\rho,dir} \leq \hat{U}_{\hat{g},t} a_g^{\rho,dir} p_g^{max} \quad \begin{array}{c}\forall dir \in \{up, down\},\\ \rho \in \{1,2\}\end{array} \tag{17b}$$

$$R_{\hat{g},t}^{tertiary} \leq (n_{\hat{g}} - \hat{U}_{\hat{g},t}) a_g^{quickstart} p_g^{max} \tag{18b}$$

C. Clustering Methodology

With the heterogeneity of generation units in real systems, the exact basis for clustering is a decision with important tradeoffs. Here we compare the results from four different approaches to aggregation:

1) Separate units – no clustering. This is the traditional formulation with binary commitment decisions for each unit.

2) Full clustering by unit type only – In this case all units with the same combination of fuel type and prime mover (e.g., coal steam, open cycle gas turbine, natural gas combined cycle) are combined into clusters.

3) Clustering by type and additional characteristics. This clustering approach sub-divides full clusters using an additional characteristic. For example, in this study we separately compare sub-dividing by size, age, or efficiency (heat rate). Cluster membership can be determined manually (as was done here) to provide roughly equal distributions of units between sub-clusters, or by using a formal clustering algorithm, such as *k*-means [38].

4) Clustering by plant. This approach clusters all units of the same type at the facility or plant level. Often, but not always, such units are identical.

For all clustering approaches, the representative unit for each cluster is assumed to have a size (nameplate capacity) equal to the average of cluster members. Technical characteristics such as heatrate, ramp rates, minimum output, etc., are taken as the size-weighted average. This representative plant is effectively duplicated such that the number of units in the cluster, $n_{\hat{g}}$, matches the original number of individual units.

D. Key Assumptions

In general, clustering assumes homogeneity of units within clusters. When clusters consist of identical units with constant incremental heat rates—i.e., only a single piecewise linear segment—the clustered solution exactly matches the traditional solution. For similar, but not identical, generators in the same cluster, they are assumed to have uniform technical characteristics such as minimum and maximum output levels, ramp rates, etc.

IV. Other Speedup Strategies

In addition to clustering, we explore other strategies for speeding up long-term unit commitment computations. We focus here on problem-specific simplifications to long-term unit commitment that can accelerate the optimization. These include:

- Constant incremental heat rate with offset: replace piecewise fuel-use with a single linear segment [23];
- Relaxed integer constraints for units with low min outputs: use relaxed commitment states for small units or units with small minimum output levels [22], [39];
- Combined reserves: aggregate reserve classes into three – off-line (tertiary), flexibility up, and flexibility down, similar to [40]; and
- Limited start-ups per time: replace minimum up/downtime with constraint on total startups per unit.

Other heuristics used in the literature include generic MILP solution tuning heuristics such as ε-optimal or "cheat", perturbing key parameters for identical units to introduce small artificial differences, or imposing a merit order. We have elsewhere [34] compared these techniques to clustering for a smaller system (IEEE Reliability Test System 1996), and found that clustering provides a two to three order of magnitude further reduction in solution time. The example system presented below does not have identical units, so we do not address the MILP heuristics further here.

V. Experimental Setup

A. Overview

We solve the unit commitment problem for an example power system to compare the computation time and results of clustering versus a traditional, binary formulation. Further comparisons are made for both formulations in conjunction with speedup strategies described in Section IV.

B. Metrics of comparison

To provide results relevant to a range of applications, we compute multiple comparison metrics. In all cases, we compare experimental runs to the full traditional binary unit commitment formulation, indicated with subscript "baseline":

1) Total Cost is the objective function value for the optimization and includes all operations costs. For comparison, we report the percent difference computed as:

$$\Delta C^{total} = (C^{total} - C_{baseline}^{total}) / C_{baseline}^{total} \tag{19}$$

2) CO_{2e} Emissions are computed system-wide based on fuel usage for both operations and startup. A scalar percent difference is computed in the same manner as total cost.

3) Energy Mix is based on total annual production by generator class divided in the same way as for clustering. The energy mix for each class is computed by summing the product of power output and duration for all time periods and dividing by the total system energy production:

$$E_{\hat{g}}^{fraction} = \left(\sum_{t \in \boldsymbol{T}} P_{\hat{g},t} \cdot 1hr \right) \Big/ \left(\sum_{t \in \boldsymbol{T}} D_t \cdot 1hr \right) \tag{20}$$

The mean absolute difference of this vector gives:

$$\Delta E^{mix} = \underset{\hat{g}\in \boldsymbol{G}}{\text{Mean}}\left|E_{\hat{g}}^{fraction} - E_{\hat{g},baseline}^{fraction}\right| \quad (21)$$

4) Commitment Plan differences are first computed as an array of differences with one element for each time period for each group of units aggregated to the cluster level. Two scalar comparisons are then made: a) The total count of differences between plans, computed as the number of non-zero elements in this array and b) the normalized mean absolute difference where commitment difference values for each time are normalized based on the total number of units committed for that time period in the baseline:

$$\Delta U = \underset{t\in \boldsymbol{T},\hat{g}\in \boldsymbol{G}}{\text{Mean}}\left|\frac{U_{\hat{g},t} - U_{\hat{g},t}\big|_{baseline}}{\sum_{\hat{g}\in \boldsymbol{G}}\left(U_{\hat{g},t}\big|_{baseline}\right)}\right| \quad (22)$$

5) Hourly Power Output differences are computed identically to commitment, except that for the count of differences, power levels are first rounded to the nearest 0.5MW.

6) Computation Time is reported as total solver (CPLEX) run time and excludes GAMS setup and output processing.

C. Implementation

All runs share a common model written in GAMS [41] that uses pre-compile flags for different data, model simplifications and solver configurations. The resulting problems were then solved using the state-of-the-art CPLEX 12.2 mixed-integer solver [42]. The solver was instructed to conserve memory when possible (memoryemphasis=1) to prevent out-of-memory errors for larger problems. The linear programming (LP) tolerance (epopt) was tightened to 1e-9 to ensure that the final LP solve matches the MILP branch-and-cut solution. The solver time limit (reslim) was set to 10 hours.

All model runs were conducted as a single thread running on a single 64-bit core (Intel Nehalem) at 2.67GHz clock speed. Up to 6 runs were run in parallel as sub-tasks of exclusive jobs on identical 8-core machines (2+ cores idle) with 24GB of shared RAM. Although run on a high performance cluster, the resulting resources allocated to each run are roughly equivalent to a modern personal computer.

VI. Test System: ERCOT

A. System Description

To test the impact of clustering on a system of realistic size, we modeled the entire Electric Reliability Council of Texas (ERCOT) balancing area using hourly historic demand and wind data from 2007. This system includes the entire Texas Interconnect, which covers the majority of the state of Texas and has negligible power exchange with other systems. ERCOT had a 2007 peak load of 62GW [43] supplied by a total of 92.5GW of generation capacity from 672 units [44].

To simplify the problem, we ignored the non-dispatchable combined heat and power facilities (15GW in 204 units), hydro (an additional 0.5GW in 41 units), units with uncommon fuel types (an additional 0.1GW in 72 units), and units with less than 50MW nameplate capacity (1GW in 56 units). In addition, we model combined cycle facilities as 36 groups instead of 115 individual combustion and steam turbines. This resulted in a total of 205 units in our model system. We also ignore wind expansion during the year and assume a fixed wind capacity equal to the final 2007 capacity of 3.7GW. Hourly wind production was taken as this capacity times the actual percent production based on the installed capacity in each time period. Historic hourly wind production and demand data from 2007 was obtained from ERCOT [43].

The week of Saturday Mar 17, 2007 was used for 1-week (168hr) analysis. This week contains both the peak wind and minimum demand. Thirteen week data include this peak wind week plus one week for each month.

Plant-level heat rate and unit nameplate (maximum) capacity data was taken from eGrid 2010 v1.1, which contains 2007 emission and plant data. Additional generator technical parameters were taken from the Sixth Northwest Power Plan appendix I [45] for corresponding plant types. Fuel costs were based on EIA 2007 data for south central west electric power sector use [46]. Reserve requirements were taken as 1% of load for regulation up and down, 1350MW for spinning reserves, and 2% of load for load following up and down. As a simple proxy for additional reserves required for wind uncertainty, load following requirements were increased as a function of both installed capacity and wind production using the factors in [40]. Up to 50% of the spinning reserve and load following up requirements can be met by quick start open cycle natural gas units.

Complete generator data tables are provided in [47]. Hourly demand and wind profile data is available by request from ERCOT. Based on the results in [47], we used no cheat with a 0.1% MIP gap for all runs.

B. Clustering Approach

We compared the four clustering approaches described in Section III-B. The resulting number of clusters and example problem sizes are included in Table I.

TABLE I:
Size and times for 1-week (168 hr) ERCOT case.

	Problem size (Before CPLEX pre-solve)					time
aggregation	**clusters equations**	**variables**	**discretes**	**non-zeros**		**(sec)**
None	205 446,394	349,960	34,272	2,068,949		4517.2

aggregation	clusters	equations	variables	discretes	non-zeros	time (sec)
None	205	446,394	349,960	34,272	2,068,949	4517.2
By Plant	90	197,922	151,048	14,952	943,685	435.3
Type & Age	17	37,650	27,400	2,688	186,173	8.0
Type & Size	17	37,650	27,400	2,688	186,173	10.2
Type & Eff.	17	37,650	27,400	2,688	186,173	11.9
Type only	7	14,802	10,264	1,008	74,957	2.2

VII. Results

A. Unit Commitment simplifications.

As seen in Table II, each of the unit commitment simplifications provides some performance improvement. With separate units, combined reserves and constraining the number of startups, rather than using the minimum up and down time, provided the most significant speed-up of around 10 times faster calculation. But, none of the simplifications were as effective as clustering alone, which was 200 times faster than any other simplification. In all cases, clustering further reduced computation time by a factor of between 350 to more than 2000.

TABLE II: RUNTIME COMPARISON FOR UNIT COMMITMENT SIMPLIFICATIONS (1-WEEK)

	Run Time (sec)	
	separate	full cluster
Full Problem	4517.2	2.2
No Integer <5MW	3693.8	1.9
No Integer <20MW	951.7	0.5
Combine Reserves	376.6	0.3
Limit startups	508.0	1.5
up/down + limit	1681.9	1.5

As seen in Fig 2, the errors for the various metrics were minimal, below 0.5% for separate units and near or below 1% for clusters.[3] The only exception was with separate units and combined reserves where normalized commitment error rose to 2.3%. The heterogeneity in generator characteristics results in relatively large (~1.25%) errors in CO_2 emissions with full clustering. The CO_2 errors are notably reduced with less aggregated clustering (next section) and longer model periods (not shown). In capacity expansion applications, future hypothetical units would be more homogenous by technology, and these errors would likely be smaller.

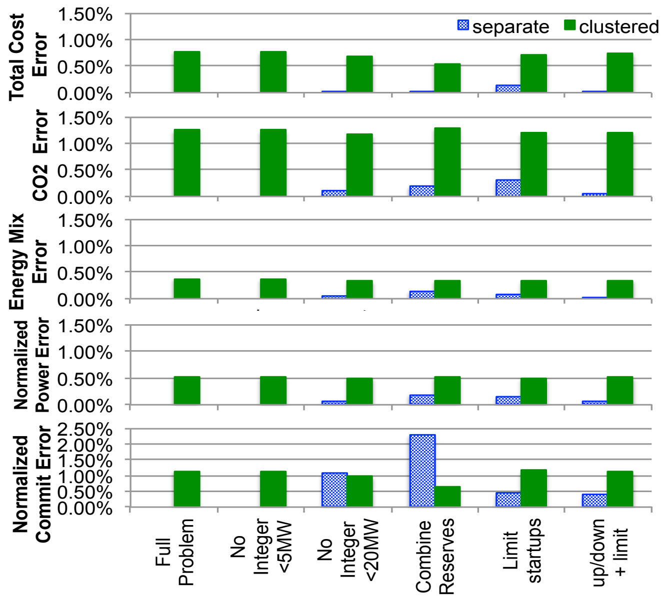


Fig. 2: Comparison of key error metrics for different Unit Commitment simplification approaches for the ERCOT 2007 test system. The full problem with separate units was used as a baseline.

B. Cluster Comparison

Fig. 3 shows how most sub-clustering schemes result in small errors (around or below 1%) with the exception of clustering by age, which has larger errors (2.3-4.5%) for all metrics except CO_2 emissions. The large errors in clustering by age result from large differences in coal output and the erroneous use of natural gas steam during peaking periods. Clustering by efficiency resulted in the lowest errors among the 17-cluster runs, for all other metrics, often close to or slightly better than the larger clustering by plant formulation (90 clusters).

[3] Unlike operational unit commitment where a 1% cost savings represents a major difference, here a few percent error is in line with other expected errors.

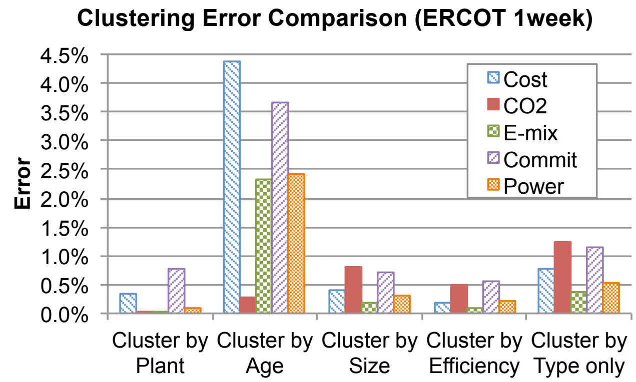


Fig. 3: Error comparison for different clustering approaches for ERCOT 2007. In all cases, the full problem with separate units was used as a baseline. All runs used a MIP gap of 0.1% and no cheat.

C. Detailed Performance Comparison

To provide more detail about the performance of clustering, we explore the distribution of normalized power errors using cumulative distribution functions (CDFs) for the sample week. We first look at CDFs across clustering method and then disaggregate the "clustering by efficiency" case by technology and time of day. Time series plots of operations over a sample week for individual units versus clustering by efficiency are visually indistinguishable, and therefore not included.

In Fig. 4, we present CDFs of the errors from each clustering strategy, aggregated over all generation technologies and all hours of the day. Similar to the average errors above, the full distributions of errors shows that the worst performance comes from clustering by age, whereas the other clustering strategies have no errors for most of the units during most hours. The quantile ranges with errors less than 1% in absolute value are (0.04, 0.95) for full clustering, (0.03, 0.95) for clustering by size, (0.02, 0.98) for clustering by efficiency, and (0.01, 0.99) for clustering by plant.

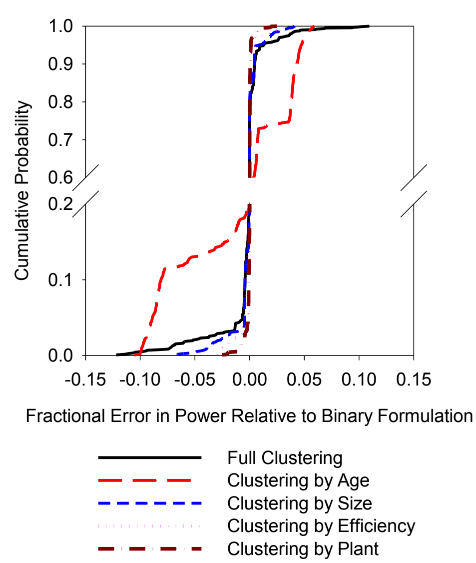


Fig. 4: Cumulative distributions of errors in power for each clustering type. Errors are calculated as the difference in power output of each generating unit in each hour from the binary formulation minus the power output from the same unit in the same hour from the clustering formulation, then normalized by the total hourly power from the binary formulation. All technologies and all hours are aggregated into a single CDF.

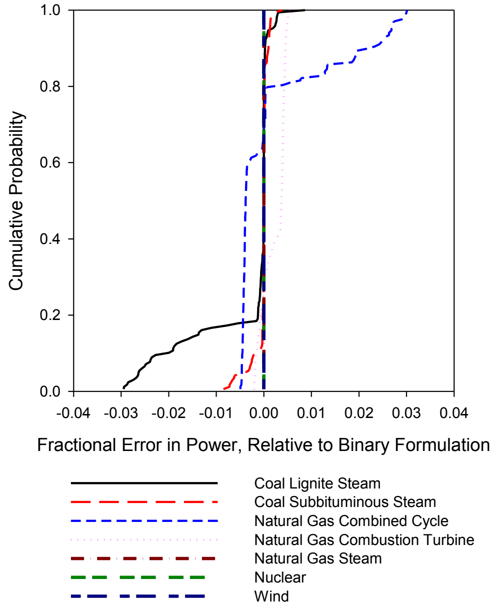


Fig. 5: Cumulative distributions of errors in power for clustering by efficiency, broken out for each generation technology. Errors are calculated as the difference in power output of each generating unit in each hour from the binary formulation minus the power output from the same unit in the same hour from the clustering formulation, then normalized by the total power from the binary formulation. Errors across all hours are aggregated into each CDF.

We further explore the error distributions by technologies and hour using the clustering by efficiency case. Fig. 5 shows the CDFs of errors from each technology aggregated over all hours. The largest biases are an overestimate of power output by 1-3% from coal lignite steam units in roughly 20% of the hours, and an underestimate of power from natural gas combined cycle units by 1-3% also for about 20% of the hours. The other technologies have errors that are always less than ±1%. Disaggregating instead by time of day (Fig. 6) shows that the largest errors occur in the overnight hours (hours 1-6). All other times of day have errors that are always less than ±1%.

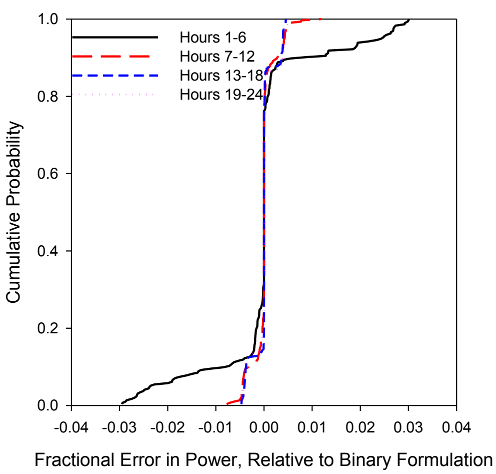


Fig. 6: Cumulative distributions of errors in power for clustering by efficiency, broken out for different times of day. Errors are calculated as the difference in power output of each generating unit in each hour from the binary formulation minus the power output from the same unit in the same hour from the clustering formulation, then normalized by the power from the binary formulation. Errors across all technologies for a range of hours are aggregated into each CDF.

For space considerations, we do not show the CDFs of errors in commitment states, but the results are qualitatively similar. With clustering by type and efficiency only natural gas combined cycle and natural gas combustion turbines had non-zero commitment errors. The general bias from clustering is to commit slightly too many NGCC units in the overnight hours, and slightly too few in the late afternoon and evening.

D. *Impact of Wind Penetration*

Future power systems will likely have significantly increased penetrations of variable renewables, such as wind, making it critical to verify that clustering continues to provide accurate results with more renewables. Fig. 7 compares our five aggregate error metrics for 0-40GW of wind, a range nearly eleven times larger than the baseline ERCOT 2007 capacity of 3.7GW wind. In all cases, the existing thermal capacity is unchanged. Wind shedding is allowed without penalty. The errors compare clustering by type and efficiency to a baseline with individual units.

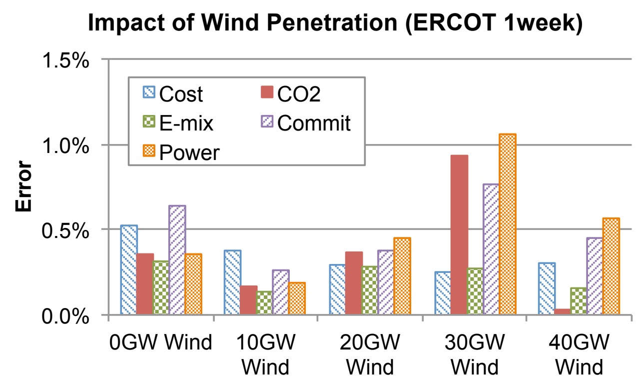


Fig. 7: Error comparison for clustering by efficiency for different installed wind capacity. In all cases, the corresponding simulation using separate units was used as a baseline. All runs used a MIP gap of 0.1% and no cheat.

As before, all of these errors remain well contained mostly near or below 0.5%, although climbing to near 1% for CO_2, commitment, and power in the 30GW wind case. These somewhat higher errors result from clustering by efficiency using somewhat more coal lignite with a corresponding reduction in subbituminous coal. In this case, small absolute CO_2 errors are further magnified by a nearly 60% reduction in total CO_2 emissions.

The results for 0-40GW are also noteworthy for the lack of an increasing error trend with larger quantities of wind. Instead the errors across all metrics vary up and down with increasing wind, suggesting that thermal unit discretization, natural variances within the MIP gap, or other factors are more important than the quantity of variable renewables.

At all wind capacities, clustering by type and efficiency continues to provide dramatic speedups ranging from 92 to over 800x faster than the separate unit baseline.

E. *Cluster Scaling.*

Fig. 8 shows how the total solver time is greatly reduced by clustering, enabling tractable computation of a full year, 8760 hour, optimal unit commitment for both 17 clusters (less than 3 hours) and 7 clusters (130 seconds). The primary driver for these speed-ups is a drastic reduction in the numbers of variables and equations which both scale roughly proportionally to the number of clusters.

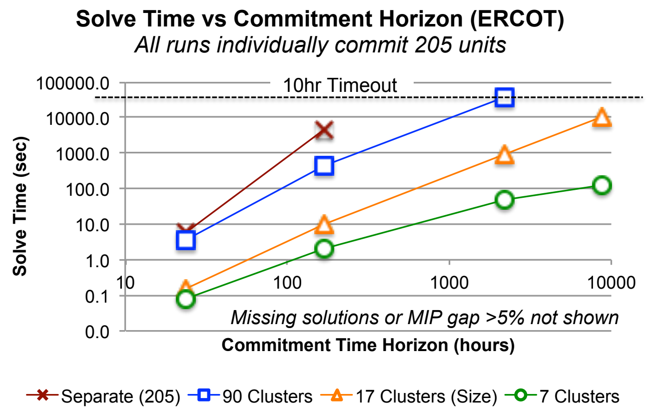


Fig. 8: Impact of clustering and model time horizon on solution time. Note both axes are logarithmic. All runs conducted with a 0.1% MIP gap and no cheat. Due to data limitations, constant heat rates are assumed. No other simplifications were used.

VIII. Conclusions

In this paper, we demonstrated the tradeoff between accuracy and runtime resulting from different levels of aggregation for heterogeneous clusters and other heuristic simplifications in unit commitment. In comparison to traditional binary formulations, clustering provides orders of magnitude faster computation – from 10 to over 1000 times faster depending on the configuration – by grouping similar units into clusters and assigning an integer, rather than binary, commitment decision to the group. This assumption builds on the existing concept of aggregating identical units. Clustering allows capturing full unit commitment constraints – including ramping, startup costs, minimum output levels, and minimum up and down times – at an individual unit level under the key assumption that all units with in a cluster have averaged technical characteristics. Despite this assumption, we show that errors are small for a wide range of metrics. Furthermore, a detailed look at the results shows that these errors are largely concentrated in (near) marginal units and certain hours. This suggests that the accuracy/performance trade-off could be further improved by sub-clustering the marginal technologies (e.g., NGCC and NGCT) more than other technologies and/or perhaps further sub-clustering these marginal units by size, operating costs, and/or startup costs.

A numeric example using an ERCOT-based 205-unit system shows that careful aggregation (17 clusters) introduces errors of 0.05-0.2% for total cost, CO_2 emissions, energy mix, and dispatch schedule while providing several orders of magnitude faster solution times (400x) compared to traditional binary formulations. The unit commitment metric exhibits errors of around 0.9%. More aggressive aggregation (seven clusters) increases errors somewhat (roughly double) but achieves further speedup (2000x). We also demonstrate a full year (8760 hour) unit commitment for a 205-unit system in less than three minutes with personal computer hardware.

We also compared other unit commitment simplifications – notably combining reserves and relaxing integer constraints for units with small minimum output levels – that can provide an additional order of magnitude speed-up for some problems. The clustering approach demonstrated here provides the ability to capture unit-level commitment decisions with intertemporal (hourly) constraints within a single optimization problem, which can be embedded within longer term operational and strategic analyses such as hydro-thermal coordination or capacity expansion under emissions or other policy constraints, especially when the long-term problem is stochastic. The application of clustering to capacity expansion and other long-term strategic decisions are left for future research.

IX. Acknowledgements

The authors wish to thank A.J. Conejo, J.I. Perez-Arriaga, C. Batlle, and our anonymous reviewers for valuable feedback on this manuscript. This work was supported by the National Science Foundation Grant Numbers 0835414 and 1128147.

X. References

[1] C. J. Baldwin, K. M. Dale, and R. F. Dittrich, "A Study of the Economic Shutdown of Generating Units in Daily Dispatch," *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, vol. 78, no. 4, pp. 1272–1282, Dec. 1959.

[2] L. L. Garver, "Power Generation Scheduling by Integer Programming-Development of Theory," *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, vol. 81, no. 3, pp. 730–734, Apr. 1962.

[3] J. A. Muckstadt and R. C. Wilson, "An Application of Mixed-Integer Programming Duality to Scheduling Thermal Generating Systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-87, no. 12, pp. 1968–1978, Dec. 1968.

[4] T. S. Dillon, K. W. Edwin, H.-D. Kochs, and R. J. Taud, "Integer Programming Approach to the Problem of Optimal Unit Commitment with Probabilistic Reserve Determination," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 6, pp. 2154–2166, Nov. 1978.

[5] G. B. Sheble and G. N. Fahd, "Unit commitment literature synopsis," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 128–135, Feb. 1994.

[6] J. M. Arroyo and A. J. Conejo, "Optimal response of a thermal unit to an electricity spot market," *Power Systems, IEEE Transactions on*, vol. 15, no. 3, pp. 1098 –1104, Aug. 2000.

[7] A. Tuohy, P. Meibom, E. Denny, and M. O'Malley, "Unit Commitment for Systems With Significant Wind Penetration," *Power Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 592–601, 2009.

[8] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve Requirements for Wind Power Integration: A Scenario-Based Stochastic Programming Framework," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197–2206, Nov. 2011.

[9] J. M. Morales, A. J. Conejo, and J. Perez-Ruiz, "Economic Valuation of Reserves in Power Systems With High Penetration of Wind Power," *Power Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 900 –910, May 2009.

[10] F. Bouffard and F. D. Galiana, "Stochastic Security for Operations Planning With Significant Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 306–316, 2008.

[11] V. S. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski, "A Stochastic Model for the Optimal Operation of a Wind-Thermal Power System," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 940–950, 2009.

[12] E. M. Constantinescu, V. M. Zavala, M. Rocklin, Sangmin Lee, and M. Anitescu, "A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment With Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 431–441, Feb. 2011.

[13] P. Meibom, R. Barth, B. Hasche, H. Brand, C. Weber, and M. O'Malley, "Stochastic Optimization Model to Study the Operational Impacts of High Wind Penetrations in Ireland," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1367–1379, Aug. 2011.

[14] J. J. Hargreaves and B. F. Hobbs, "Commitment and Dispatch With Uncertain Wind Generation by Dynamic Programming," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 724–734, 2012.

[15] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 52–63, 2013.

[16] 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, Feb. 2009.

[17] J. F. Restrepo and F. D. Galiana, "Assessing the Yearly Impact of Wind Power Through a New Hybrid Deterministic/Stochastic Unit Commitment," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 401–410, Feb. 2011.

[18] B. Palmintier and M. Webster, "Impact of Unit Commitment Constraints on Generation Expansion Planning with Renewables," in *Proceedings of 2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, 2011.

[19] S. Kamalinia and M. Shahidehpour, "Generation expansion planning in wind-thermal power systems," *IET Generation, Transmission & Distribution*, vol. 4, no. 8, p. 940, 2010.

[20] J. Ma, V. Silva, R. Belhomme, D. S. Kirschen, and L. F. Ochoa, "Evaluating and Planning Flexibility in Sustainable Power Systems," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 200–209, 2013.

[21] D. S. Kirschen, J. Ma, V. Silva, and R. Belhomme,, "Optimizing the Flexibility of a Portfolio of Generating Plants to Deal with Wind," in *Power and Energy Society General Meeting, 2011 IEEE*, Detroit, MI, 2011.

[22] A. Shortt, J. Kiviluoma, and M. O'Malley, "Accommodating Variability in Generation Planning," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 158–169, 2013.

[23] R. Gollmer, M. P. Nowak, W. Römisch, and R. Schultz, "Unit commitment in power generation–a basic model and some extensions," *Annals of Operations Research*, vol. 96, no. 1, pp. 167–189, 2000.

[24] N. Langrene, W. van Ackooij, and F. Breant, "Dynamic Constraints for Aggregated Units: Formulation and Application," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1349–1356, Aug. 2011.

[25] K. Hara, M. Kimura, and N. Honda, "A Method for Planning Economic Unit Commitment and Maintenance of Thermal Power Systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-85, no. 5, pp. 427–436, May 1966.

[26] J. Garcia-Gonzalez, R. M. R. de la Muela, L. M. Santos, and A. M. Gonzalez, "Stochastic Joint Optimization of Wind Generation and Pumped-Storage Units in an Electricity Market," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 460–468, May 2008.

[27] H. D. Sherali and J. C. Smith, "Improving Discrete Model Representations via Symmetry Considerations," *Management Science*, vol. 47, no. 10, pp. 1396 –1407, Oct. 2001.

[28] S. Cerisola, "Benders Decomposition for Mixed Integer Problems: Application to a Medium Term Hydrothermal Coordination Model," PhD Dissertation, Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Madrid, Spain, 2004.

[29] S. Cerisola, Á. Baíllo, J. M. Fernández-López, A. Ramos, and R. Gollmer, "Stochastic Power Generation Unit Commitment in Electricity Markets: A Novel Formulation and a Comparison of Solution Methods," *Oper. Res.*, vol. 57, pp. 32–46, Jan. 2009.

[30] S. Sen and D. P. Kothari, "An Equivalencing Technique For Solving The Large-Scale Thermal Unit Commitment Problem," in *The Next Generation of Electric Power Unit Commitment Models*, B. F. Hobbs, M. H. Rothkopf, R. P. O'Neill, and H. Chao, Eds. Boston: Kluwer Academic Publishers, 2002, pp. 211–225.

[31] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. Wiley-Interscience, 1996.

[32] R. Baldick, "The generalized unit commitment problem," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 465–475, Feb. 1995.

[33] R. P. O'Neil, "It's Getting Better All the Time (with Mixed Integer Programming)," presented at the HEPG Forty-Ninth Plenary Session, Los Angeles, CA, Dec-2007.

[34] J. Ostrowski, M. F. Anjos, and A. Vannelli, "Tight Mixed Integer Linear Programming Formulations for the Unit Commitment Problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 39–46, Feb. 2012.

[35] L. Wu, M. Shahidehpour, and T. Li, "Stochastic Security-Constrained Unit Commitment," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 800–811, May 2007.

[36] D. Rajan and S. Takriti, "Minimum Up/Down Polytopes of the Unit Commitment Problem with Start-Up Costs," IBM T.J. Watson Research Center, Yorktown Heights, NY, IBM Research Report, Jun. 2005.

[37] K. W. Hedman, R. P. O'Neill, and S. S. Oren, "Analyzing valid inequalities of the generation unit commitment problem," in *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, 2009, pp. 1–6.

[38] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[39] P. A. Ruiz, C. R. Philbrick, and P. W. Sauer, "Modeling Approaches for Computational Cost Reduction in Stochastic Unit Commitment Formulations," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 588–589, Feb. 2010.

[40] C. De Jonghe, E. Delarue, R. Belmans, and W. D'haeseleer, "Determining optimal electricity technology mix with high level of wind power penetration," *Applied Energy*, vol. 88, no. 6, pp. 2231–2238, Jun. 2011.

[41] GAMS Development Corporation, *General Algebraic Modeling System (GAMS)*. 2010.

[42] IBM ILOG, *CPLEX Optimizer*. 2010.

[43] ERCOT, "2007 Wind Output." 2010.

[44] US Environmental Protection Agency (EPA), *Emissions & Generation Resource Integrated Database (eGrid)*. 2010.

[45] Northwest Power and Conservation Council, "Sixth Northwest Conservation and Electric Power Plan," 2010-09, Feb. 2010.

[46] US Energy Information Administration [EIA], "Annual Energy Outlook 2010, Table 3," 2010. [Online]. Available: http://www.eia.gov/oiaf/aeo/tablebrowser/#release=AEO2010&table=3-AEO2010. [Accessed: 05-Nov-2011].

[47] B. Palmintier and M. Webster, "Heterogeneous Unit Clustering for Efficient Operational Flexibility Modeling," MIT Engineering Systems Division, Cambridge, MA, ESD Working Paper ESD-WP-2013-04, 2013.

XI. Biographies

Bryan Palmintier (M '07) holds a Ph.D. from the Engineering Systems Division (2013) at the Massachusetts Institute of Technology; a Degree of Engineer in mechanical engineering (2004) and an M.S. in aero/astro engineering (2000) from Stanford; and a B.S. in aerospace engineering (1996) from Georgia Tech. He is currently a Department of Energy EERE/SunShot Postdoctoral Fellow at the National Renewable Energy Lab, Golden, CO, USA. His research focuses on integration of renewable and efficient electric power systems.

Mort Webster (M '10) holds a Ph.D. from the Engineering Systems Division (2000) at the Massachusetts Institute of Technology. Professor Webster holds an M.S. from MIT's Technology and Policy Program (1996), and a B.S.E. in Computer Science and Engineering (1988) from the University of Pennsylvania. He is an Assistant Professor of Engineering Systems at MIT, Cambridge, MA, USA. His research and teaching focuses on environmental and energy systems analysis and decision-making under uncertainty.