MIT CEEPR

MIT Center for Energy and Environmental Policy Research

Working Paper Series

Modeling Unit Commitment in Political Context: Case of China's Partially Restructured Electricity Sector

MICHAEL R. DAVIDSON AND IGNACIO J. PÉREZ-ARRIAGA

APRIL 2017

CEEPR WP 2017-010

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Modeling Unit Commitment in Political Context: Case of China's Partially Restructured Electricity Sector

Michael R. Davidson, J. Ignacio Pérez-Arriaga

Abstract

Restructuring an electricity sector entails a complex realignment of political and economic institutions, which may both delay and distort the achievement of satisfactorily competitive conditions. In research and planning for policy interventions in power systems under these varied regulatory environments, typical operational models may neglect important interactions between techno-economic criteria and political constraints, leading to poor understanding of underlying causes of inefficiency and to inappropriate recommendations. We develop tractable formulations of the unit commitment problem based on integer clustering of similar units that endogenize important political factors in the Northeast grid region of China. We demonstrate the importance of these interactions on operations and provide a set of options for researchers to explore further pathways for China's ongoing power system reforms. For example, wind integration, a key policy priority, is inhibited by the interaction of institutions limiting short- and long-term sources of flexibilities in inter-provincial trade.

Index Terms

Electricity deregulation, unit commitment, power systems operation, renewable energy policy, combined heat and power (CHP).

I. Introduction

A wide range of countries have initiated some form of electricity sector restructuring since the 1980s, choosing to introduce competition into one or several segments of the traditional vertically-integrated utility (VIU) model of electricity supply. Motivations for these transitions are varied, ranging from expected efficiency benefits and relaxing of demands on public finance associated with the entry of private actors and capital, to regulatory goals of tackling state-owned and entrenched interests [1]–[3].

M. R. Davidson is a Ph.D. candidate in the Massachusetts Institute of Technology (MIT) Institute for Data, Systems, and Society, Cambridge, MA, and Research Assistant with the MIT Joint Program on the Science and Policy of Global Change (contact: michd@mit.edu, www.mdavidson.org).

J. I. Pérez-Arriaga is Professor & Director of the BP Chair on Energy & Sustainability, Instituto de Investigacion Tecnologica (IIT), Universidad Pontificia Comillas, Madrid, Spain, Visiting Professor at MIT Sloan School of Management, and MIT Center for Energy and Environmental Policy Research Faculty Affiliate.

Nevertheless, due to differences in institutional make-up, resource endowments, regulatory philosophies and macro-economic conditions, among other factors, these transitions have been often protracted and incomplete [2]–[4]. In addition, increasing coordination even among neighboring well-established restructured systems can run into various institutional complications, such as the long road to establish a common European market [5] and addressing various interests in intertie markets in the western US interconnect [6]. These can lead to outcomes that deviate from efficient outcomes assuming ideal economic prescriptions are followed. Under these settings, it is important for policy-makers, regulators, and researchers to appropriately model and understand how transition electricity systems operate, and to set realistic baselines for policy analysis.

China, currently undergoing a long transition from a state-run VIU to competitive wholesale and retail markets, is an important area of power systems research due to its size and effects on global environmental challenges, as well as implications for other transitioning systems. Similar to other countries, unique institutional structures and entrenched relationships between government and industry have led to complications in ownership and regulatory reforms, delaying the introduction of competition [7]. In particular, a quota-based system whereby generation hours are guaranteed for generators at fixed prices was maintained during restructuring. This system, intended as interim until competitive conditions were achieved, has become one of the most difficult roadblocks to establishing price competition mechanisms [8].

Over the next several years, China is engaging in additional reforms to create competitive markets and address air pollution and climate change impacts of electricity generation [9]. Pilots primarily at the provincial level will test compatibility of incentives with existing institutions with the goal of moving all commercial and industrial electricity transactions to medium- to long-term contracts by 2020 [10], ensuring that a diverse set of rules in the sector will persist.

An essential recurring function of system operators in all varieties of regulated structures is the scheduling of startup/shutdown and dispatch of generators to meet expected load on a daily basis. This is typically solved using a unit commitment (UC) optimization which minimizes production cost subject to various technical constraints [11]–[19], though in China, due to its partial liberalization of operations, a complex mix of dispatch priorities exist that are not fully optimized [20]. In addition to ensuring economic and reliable operation of existing assets, the proper functioning of this dispatch optimization is also deemed essential in restructured markets to provide efficient long-term investment signals [21], [22]. Establishing the central position of UC in grid operations and reducing administrative constraints will thus be similarly important in China's restructuring efforts.

Much research into the UC model has been aimed at improving computational performance of the solution algorithm [12], [16], [17], incorporating uncertainty [15], [19], and widening the scope of decisions such as to include investments [13], [18]. In terms of analyzing institutional factors and degrees of restructuring, the difference between zonal and nodal market designs is an important area of research, especially as it relates to integrating renewable energy through market coupling mechanisms [23]. No UC work to the best of the authors' knowledge has focused on modeling operations under political constraints established during transition such as China's generation quota.

Meanwhile, equilibrium models have been important tools for investigating the impact of restructuring on electricity sector outcomes, such as to examine the presence of market power [24]. The goals of this research are different, and more research into tractable equilibrium UC models is needed as these mixed complementarity formulations do not scale well with integer variables [25].

This paper formulates a new UC model with details of key political institutions influencing system operations in China and applies it to China's Northeast Grid. New operational challenges in the Northeast China Grid are primarily the result of the increasing penetration of wind energy conflicting with inflexible coal operation and combined heat and power cogeneration in winter. The quota introduces a long-term coupling constraint causing computational time and convergence difficulties in typical UC formulations. To facilitate consideration of the quota and run sensitivities over uncertain political parameters, we advance a clustering technique traditionally designed to speed computation in planning models that makes similar units identical and generalizes binary commitment variables to integer variables. The paper's main contributions are to:

1) Implement elements of China's partially-restructured electricity sector, in particular its hierarchical dispatch structure, and a method for quantifying short-term and long-term inflexibilities associated with it;
2) Highlight interactive effects among political and technical constraints relevant for future consideration by modelers and policy-makers;
3) Demonstrate tractable approach with acceptable errors using similar unit clustering to optimize generator scheduling under annual coupling regulatory constraints.

II. Partial Liberalization of Operation

Since the 1980s, when China's generation sector was opened up to investment other than the primary network owner, China has struggled to define and implement consistent rules for generator access to the transmission network. For two decades, multiple generation owners competed with assets of the central state-run VIU in the absence of prices or other unambiguous criteria for deciding dispatch, leading to claims of discrimination [26]. Local government ownership of generators also created conditions for "reverse" discrimination against VIU-owned plants through local government authority over dispatch rule-making [27], [28].

In 1998, the State Council, China's highest-level policy-making body, initiated reforms to create an "open, competitive and orderly electricity market", proceeding to break up the former state-run electricity ministry into a state-owned electric power company and new regulatory and policy bodies [29]. In 2002, the State Council unbundled the electric power company into the current arrangement of two major network companies and five large generation SOEs ("Big Five"), and called for the creation of wholesale generation markets and an independent regulatory body [30].

Following unbundling, a "benchmark electricity tariff" for each province was established for thermal generators, reflecting unpublished cost and return expectations as well as affordability based on the economic development of the province [31]. This structure, sometimes referred to as a yardstick tariff, was only intended as a temporary measure before competitive wholesale regional markets were fully implemented, though it still remains in force [20], [26], [31].

In the absence of differentiated cost signals, either through internal cost accounting or market-clearing bids, other criteria are used to determine the dispatch order. First, an overriding priority across operation of China's electricity sector is the concept of *equitable dispatch*, specifying that benefits (and costs) of supplying electricity should be shared equitably among generators. Actual costs may differ from the benchmark tariff and there is no consistently applied method to adjust the tariff to respond to these changes. As a result, minimum generation quotas are allocated to generators on an annual basis to guarantee sufficient revenues [32]. In other countries, stranded assets and insufficient compensation as a result of a transition to wholesale power markets are sometimes handled by side-payments from the regulator [33]. However, no comprehensive system of transition payments was created in China, the absence of which creates political pressure to maintain production from inefficient generators.

Second, generator-specific operational restrictions including minimum generation outputs, minimum up/down times and ramp rates are either determined administratively for generic unit types or self-reported by generators themselves. These tend to be conservative: for example, minimum generation points of coal units are typically set at 50-60% [34]. Energy-only remuneration (i.e., no compensation for commitment costs) combined with above equity concerns lead the grid operator to set minimum up times at 5-7 days and typically much longer. Therefore, over time intervals of less than a week, commitment decisions are relatively rigid and the sum of minimum generation outputs, referred to as *minimum mode*, can be quite high [34]. The over-conservativeness of these self-reported arrangements leading to efficiency losses has been observed in other partially-restructured systems, such as Spain in the 1990s [35].

Third, since 2007, areas of the country have piloted an additional grid management priority called *energy-efficient dispatch*, which prioritizes first renewables and nuclear, and continuing with coal units in decreasing order of efficiency [36]. This was instituted in response to some provinces dispatching lower-efficiency units more than new high-efficiency units, thereby degrading the benefits of central energy conservation policies [37]. Implementation has been uneven, attributed to preferential treatment by provincial governments for government-invested plants, conflicts with government-mediated bilateral contracts between suppliers and large consumers, and lack of incentive for grid companies to comply [20], [38].

Defining relevant balancing areas for power system operation can be ambiguous in China as dispatch centers at the provincial and regional levels (6 in total, consisting of 3-6 provinces each) all have degrees of autonomy. Generation quotas and other dispatch priorities are determined and evaluated at the provincial level, hence a large fraction of generators are still dispatched at the provincial level. Adjustments for planned over-/under-supply can be negotiated through allocating inter-provincial transmission capacity in annual plans, and are coordinated by the regional dispatch operator [20].

Long-term transmission contracts and autonomous dispatch centers operating based on distinct annual plans and minimum modes thus highly restrict short-term (i.e., intra-day) balancing operations between bordering regions. Imbalances must be met mostly using intra-provincial resources, and can run into conflicts with long-term plans under changing conditions [39]. These delineations of authority were largely the result of a compromise during unbundling between provincial governments, who enjoyed significant influence over planning and operation under the previous vertically-integrated structure, and central grid agencies, who sought a well-connected grid network with centralized operation [40].

III. Model

A. Standard Unit Commitment

The standard UC problem seeks to minimize operational costs of meeting a given electricity demand, whose objective consists of variable generation costs and the startup (commitment) costs of thermal generators. We start with classic formulations [11] and linearize the objective function as in [12], [13]:

$$\min \quad \sum_{g \in G} \sum_{t \in T} \left(p_g^{su} \boldsymbol{v}_{g,t}^{up} + p_g^{var} \boldsymbol{y}_{g,t} \right) \tag{1}$$

where $\boldsymbol{y}_{g,t}$ is the dispatch (continuous) level and $\boldsymbol{v}_{g,t}^{up}$ the startup decision of generator g at time t, p_g^{var} and p_g^{su} are variable and startup costs (respectively) of generator g, G is the set of generators, T the set of time periods. Throughout, **bold typeface** refer to decision variables. This is subject to electricity demand and transmission constraints (Kirchhoff's first law):

$$\sum_{g \in G_p} \boldsymbol{y}_{g,t} - \sum_{p' \neq p} \left[\boldsymbol{f}_{p,p',t} + \boldsymbol{l}_{p,p',t}/2 \right] = d_{p,t} \ , \ \forall p \in P \tag{2}$$

where $d_{p,t}$ is the electricity demand at provincial node p at time t, $\boldsymbol{f}_{p,p',t}$ is transmission flow from p to p' at time t, and $\boldsymbol{l}_{p,p',t}$ is the non-negative transmission loss associated with that flow. Note that we are not considering intra-provincial transmission constraints or losses.

Network losses are functions of sinusoids of the angle differences between nodes, frequently approximated piecewise linearly such as in combination with the DC approximation of the optimal power flow problem (DC-OPF) [41]. In longer-time horizon models such as unit commitment or expansion planning, losses are typically ignored altogether (e.g., [12]). However, given the long transmission distances between provincial nodes, we felt that losses should not be neglected. We adopt a piece-wise linear loss formulation in terms of flow decision variables [41]:

$$\begin{aligned}
\boldsymbol{f}_{p,p',t} &= -\boldsymbol{f}_{p',p,t} && (3)\\
\boldsymbol{f}_{p,p',t} &= \boldsymbol{f}^{+}_{p,p',t} - \boldsymbol{f}^{-}_{p,p',t} && (4)\\
\sum_{s} \boldsymbol{j}_{p,p',t,s} &= \boldsymbol{f}^{+}_{p,p',t} + \boldsymbol{f}^{-}_{p,p',t} && (5)\\
\boldsymbol{f}_{p,p',t} + \boldsymbol{l}_{p,p',t}/2 &\leq \overline{F}_{p,p'} && (6)\\
\boldsymbol{l}_{p,p',t} &= \mu_{p,p'} \sum_{s} \alpha_{p,p',s} \boldsymbol{j}_{p,p',t,s} && (7)\\
\alpha_{p,p',s} &= (2s-1)\, \Delta f_{p,p'} \ , && (8)\\
& \qquad \forall s = 1..S \\
\Delta f_{p,p'} &= \overline{F}_{p,p'}/S && (9)\\
\boldsymbol{l}_{p,p',t}, \boldsymbol{f}^{+}_{p,p',t}, \boldsymbol{f}^{-}_{p,p',t}, \boldsymbol{j}_{p,p',t,s} &\geq 0 && (10)\\
& \forall t \in T,\, p, p' \in P
\end{aligned}$$

where $\boldsymbol{f}^{+}_{p,p',t}$, $\boldsymbol{f}^{-}_{p,p',t}$ are the positive and negative components of the flow from p to p' at time t ($\left|\boldsymbol{f}_{p,p',t}\right| = \boldsymbol{f}^{+}_{p,p',t} + \boldsymbol{f}^{-}_{p,p',t}$), and $\overline{F}_{p,p'}$ is the available transmission capacity from p to p'. Available capacity is divided into $S = 20$ pieces indexed by s, with flow in each segment given by $\boldsymbol{j}_{p,p',t,s}$, and maximum flow in each segment by $\Delta f_{p,p'}$. Resistive loss coefficients are $\mu_{p,p'}$, and the linear slope of the quadratic linearization by $\alpha_{p,p',s}$.

Generator constraints on production and commitment:

$$\underline{P}_g \boldsymbol{u}_{g,t} \leq \boldsymbol{y}_{g,t} \leq \overline{P}_g \boldsymbol{u}_{g,t}\ ,\ \forall g \in G_{thermal} \tag{11}$$

$$0 \leq \boldsymbol{y}_{g,t} \leq W_{g,t}\ ,\ \forall g \in G_{wind} \tag{12}$$

$$\boldsymbol{w}_{g,t} = \boldsymbol{y}_{g,t} - \underline{P}_g \boldsymbol{u}_{g,t} \tag{13}$$

$$\boldsymbol{w}_{g,t} - \boldsymbol{w}_{g,t-1} \quad \leq \quad RU_g \tag{14}$$

$$\boldsymbol{w}_{g,t-1} - \boldsymbol{w}_{g,t} \quad \leq \quad RD_g \tag{15}$$

$$\boldsymbol{u}_{g,t} \quad \geq \quad \sum_{t'=t-MU_g}^{t} \boldsymbol{v}^{up}_{g,t'} \tag{16}$$

$$1 - \boldsymbol{u}_{g,t} \quad \geq \quad \sum_{t'=t-MD_g}^{t} \boldsymbol{v}^{dn}_{g,t'} \tag{17}$$

$$\boldsymbol{u}_{g,t} - \boldsymbol{u}_{g,t-1} \quad = \quad \boldsymbol{v}^{up}_{g,t} - \boldsymbol{v}^{dn}_{g,t} \tag{18}$$

$$\forall g \in G_{thermal}, t \in T$$

where ($\boldsymbol{v}^{up}_{g,t}$, $\boldsymbol{v}^{dn}_{g,t}$) are startup and shutdown decisions, $\boldsymbol{w}_{g,t}$ is an auxiliary ramping variable, ($\underline{P}_g$, $\overline{P}_g$) are minimum and maximum outputs, (RU_g, RD_g) are maximum upward and downward ramp rates, (MU_g, MD_g) are minimum up and down times, G_{wind} is the set of wind generators, and $W_{g,t}$ is the available wind power by wind generator g at time t. To ensure feasibility of ramping and commitment decisions at the beginning and end of the time period, periodic boundary conditions are assumed (i.e., for negative time indices $-t' \equiv T - t'$).

Spinning reserves to ensure sufficient flexibility to respond to unpredicted changes in demand and supply are calculated from all committed units based on technical limits:

$$\boldsymbol{r}_{g,t} \quad \leq \quad \boldsymbol{u}_{g,t}\overline{P}_g - \boldsymbol{y}_{g,t} \tag{19}$$

$$\boldsymbol{s}_{g,t} \quad \leq \quad \boldsymbol{y}_{g,t} - \boldsymbol{u}_{g,t}\underline{P}_g \tag{20}$$

$$\boldsymbol{r}_{g,t} \quad \leq \quad RU_g \tag{21}$$

$$\boldsymbol{s}_{g,t} \quad \leq \quad RD_g \tag{22}$$

$$\forall g \in G, t \in T$$

$$\sum_{g \in G_{res}} \boldsymbol{r}_{g,t} \geq \overline{RES}_t \tag{23}$$

$$\sum_{g \in G_{res}} \boldsymbol{s}_{g,t} \geq \underline{RES}_t \tag{24}$$

$$\forall t \in T$$

where $(\boldsymbol{r}_{g,t}, \boldsymbol{s}_{g,t})$ are upward and downward reserve variables, $(\overline{RES}_t, \underline{RES}_t)$ are upward and downward system-wide reserve requirements, and G_{res} is the set of generators providing reserves.

B. Combined heat and power

Combined heat and power (CHP) for district heating is widespread in northern China, where much residential heating in urban areas as well as process steam for industrial applications are provided by centralized cogeneration facilities [42]. These primarily coal-fired cogeneration units have distinct operational constraints co-dependent on heat and electricity output, related to minimum and maximum stable boiler outputs, and minimum condenser threshold [43]. The latter leads to increasing $\underline{P}_g$ with heat output while maximum boiler outputs lead to decreasing $\overline{P}_g$ with heat output, which can be used to co-optimize heat and electricity production [44]. These two constrain the feasible range of electricity outputs under high heat extraction compared to electricity-only units.

A simplifying assumption used by Chinese grid operators for dispatch purposes is to specify for each CHP unit a fixed minimum electricity output that does not vary with diurnal heating load. Rather, this may take on different constant values throughout the heating season, typically divided into three stages: early, middle and late, with higher heating loads in the middle. This corresponds to adjusting $(\underline{P}_g, \overline{P}_g)$ for $g \in G_{CHP}$, the set of CHP generators. For this paper, these new constraints are assumed constant over the day.

C. Hydropower

Hydrothermal coordination models may consider multiple time horizons to predict, plan and adjust dispatch based on expected hydropower availability, for example optimized over a full year, then a month or week, and finally daily. Production functions converting water flow into electricity generation may be approximated as linear [14]. These models must consider a range of additional constraints related to other water uses such as irrigation, tourism and fisheries.

As our main thrust in this model is not to endogenize these constraints–which would also require significant additional data collection–we consider hydropower as a flexible resource over the model horizon, with inflows given by historic generation and fixed initial and final states, and minimum and maximum reservoir levels:

$$\boldsymbol{h}_{g,t} - \boldsymbol{h}_{g,t-1} = H_g - \boldsymbol{y}_{g,t} \tag{25}$$

$$\boldsymbol{h}_{g,t} = HL_{g,t} \,,\, t \in \{1, |T|\} \tag{26}$$

$$\boldsymbol{h}_{g,t} \geq \underline{HL}_g \tag{27}$$

$$\boldsymbol{h}_{g,t} \leq \overline{HL}_g \tag{28}$$

$$\boldsymbol{h}_{g,t} \geq 0 \quad \forall g \in G_{hydro}, t \in T \tag{29}$$

where $\boldsymbol{h}_{g,t}$ is the reservoir level in units of generation, H_g is mean inflow of generator g over a timestep, G_{hydro} the set of hydro generators, $HL_{g,t}$ for $t = \{1, |T|\}$ are the fixed initial and final levels, and $\underline{HL}_g, \overline{HL}_g$ are lower and upper reservoir levels, respectively.

D. Partially Restructured Operation

Due to the partial restructuring of China's electricity sector outlined in Section II, the formulation (1)-(29) does not represent the decision-making situation faced by grid operators. We introduce here a new formulation that captures the essential features of China's partially restructured operation, in particular, the strong role and autonomy of provincial balancing authorities and minimum generation quotas.

1) Provincial Dispatch: To completely model unit commitment and dispatch decisions across multiple provinces that are not centrally optimized would require either a complex simulation method based on heuristics of each province's decision-making process or, if possible, a multi-objective optimization. The former would require substantial granular detail into individual plants and grid company operations which are typically confidential and parameterizing of which could become quite complicated when dealing with various technical constraints. The latter would still require constraints on inter-provincial trade and, most importantly, if it were to be solved concurrently for all provinces would likely lead to an equilibrium formulation [24], which does not scale well with binary variables [25].

Therefore, in order to evaluate gains from lifting various operational restrictions, we propose a formulation that captures this decentralized decision-making in terms of a single-objective optimization. The unconstrained model represents the ideal reference, and we subsequently add restrictions to better reflect reality and evaluate efficiency losses. We identify at least two important changes that occur when dispatch is no longer centralized across provinces: transmission line capacities are constrained below their limits, and reserve requirements must be calculated separately for each province. These reflect, respectively, *long-term inflexibilities* associated with inter-provincial transmission contracts and *short-term inflexibilities* due to coordination challenges between distinct operators in charge of balancing operations (< 1 hour). The latter is similar to a pilot reserve-sharing mechanisms begun in 2014 in East China Grid, which was noted by regulators to lower costs and enhance security in the case of contingencies in long-distance transmission lines [45].

Inter-provincial transmission constraints are essentially derived from annual energy production and consumption planning, and then converted to power transfers on sub-monthly scales [32]. Ideally, granular data on fine temporal scales (e.g., weekly or daily) would be used to fix a narrow range of allowable transmission quantities; however, only annual aggregates of these data are available. Instead, transmission limits are modeled as uniform limits on transmission interconnection capacities $\overline{F}_{p,p'}$ based on historic aggregate transfers. The aggregate data are typically summarized in terms of flows over major lines, which indicate in some cases that particular export-import relationships exist in the transmission capacity allocation process.

In the standard UC model, aggregate reserve constraints (23)-(24) are imposed for the entire region. Under provincial dispatch, we also require each province to meet its reserve requirements internally, replacing (23)-(24) with:

$$\sum_{g \in G_{res} \cap G_p} \boldsymbol{r}_{g,t} \quad \geq \quad \overline{RES}_{p,t} \tag{30}$$

$$\sum_{g \in G_{res} \cap G_p} \boldsymbol{s}_{g,t} \quad \geq \quad \underline{RES}_{p,t} \tag{31}$$

$$\forall t \in T, p \in P$$

2) Minimum Generation Quotas:

a) Constraint-based formulation (CON): Similar to modeling hydro-thermal coordination and mid-term maintenance scheduling, the requirement that each generator achieves a minimum amount of generation over the course of the year introduces a large coupling constraint. Extending the time horizon T to an entire year would require significant simplifications to remain tractable, therefore we propose minimum generation constraints on aggregated similar cost units. This allows, for example, a unit to not be committed during the model horizon without violating its annual quota. As similar cost units, the result from an aggregate constraint on production over horizon T should not differ significantly from imposing the constraint on each individual generator over the year, with the possible exception of underestimating commitment costs. The method for determining similar cost units for the given power system is described in Sec IV-A. This first step of simplifying the set of generators will be known as “aggregation” to distinguish from “clustering” introduced below.

Additionally, we do not assume that the constraint is constant throughout the year. In particular, as cogeneration units must be committed to provide district heating in the roughly six-month winter heating season, non-cogeneration units are preferentially committed in summer to maintain comparable generation amounts over the year. There is a trade-off when selecting the length of the time horizon T such that, on one hand, sufficient flexibility is allowed for meeting the quota constraint and, on the other, the problem remains tractable. We describe a method in Sec IV-D for our test system to construct these quotas at a weekly level from annual data. In its general form, this constraint is given by:

$$\sum_{g \in G_{p,k}} \sum_{t \in T} \boldsymbol{y}_{g,t} \geq Q_{p,k} \cdot |T| \cdot \sum_{g \in G_{p,k}} \overline{P}_g, \ \forall p \in P, k \in K \tag{32}$$

where $k \in K$ indexes clustered generators with minimum generation quotas, $Q_{p,k}$ is the capacity factor quota for generators of type k in province p, and $G_{p,k}$ is the set of these generators.

b) Penalty formulation (PEN): The quota introduces large coupling constraints across the entire time horizon. We therefore test an alternative formulation with a penalty in the objective, replacing (1) with:

$$\min \quad \sum_{g \in G} \sum_{t \in T} \left(p_g^{su} \boldsymbol{v}_{g,t}^{up} + p_g^{var} \boldsymbol{y}_{g,t} \right) + \sum_{p,k} p^Q q_{p,k} \tag{33}$$

where p^Q is a (large) penalty for failing to meet the quota, and the deviation $q_{p,k}$ is given by:

$$q_{p,k} \geq \sum_{g \in G_{p,k}} \sum_{t \in T} \left(Q_{p,k} \overline{P}_g - \boldsymbol{y}_{g,t} \right) \tag{34}$$

$$q_{p,k} \geq 0 \tag{35}$$

$$\forall p \in P, k \in K$$

E. Clustering

Improving computational performance of UC models is a major area of research with aims of coupling operational technical characteristics in planning models, incorporating uncertainty and other high-dimension methods, and improving solution reliability of current practices [11]–[19]. The key complicating feature is the trio of binary commitment variables ($\boldsymbol{u}_{g,t}$, $\boldsymbol{v}_{g,t}^{up}$, $\boldsymbol{v}_{g,t}^{dn}$) that historically required Lagrangian relaxation methods and is now commonly solved using branch-and-bound methods. Efforts that aim to preserve feasibility while improving solution times involve tightening bounds by finding valid inequalities [12], [16] and reducing the number of binary variables [17].

In addition to speeding up computation for sensitivity analysis with uncertain political parameters, the appropriate UC model for China must consider the complicating constraint arising from the annual generation quota. Considering the entire year would be intractable, and even at shorter time scales, this coupling constraint can slow convergence, in ways analogous to incorporating unit startup/shutdown decisions into expansion planning models. Reduction techniques generally fall into categories of time dimension reduction through the use of representative days and weeks [15], [18] and homogeneous or similar unit clustering [13], [19].

We employ here a formulation based on [13] that clusters multiple binary commitment variables of similar units into integer variables over the combined cluster of generators. Here, we refer to "clustering" as transforming the set of decision variables for multiple units into a new, smaller set for a cluster. We distinguish this from "aggregation" introduced in Section III-D2, which only refers to making similar units identical across some set of generation parameters, but which does not alter the set of decision variables. We extend the original formulation to a multi-node system, testing the validity of this approximation in Sec IV. Indexing over generator types $k \in K$ within each $p \in P$, the basic structure involves the variable transformations:

$$(\boldsymbol{y}_{g,t},\, \boldsymbol{w}_{g,t}) \rightarrow (\mathbf{Y}_{p,k,t},\, \mathbf{W}_{p,k,t}) \tag{36}$$

$$(\boldsymbol{u}_{g,t},\, \boldsymbol{v}_{g,t}^{up},\, \boldsymbol{v}_{g,t}^{dn}) \rightarrow (\mathbf{U}_{p,k,t},\, \mathbf{V}_{p,k,t}^{up},\, \mathbf{V}_{p,k,t}^{dn}) \tag{37}$$

$$(\boldsymbol{r}_{g,t},\, \boldsymbol{s}_{g,t}) \rightarrow (\mathbf{R}_{p,k,t},\, \mathbf{S}_{p,k,t}) \tag{38}$$

where $\mathbf{U}$, $\mathbf{V}^{up}$, $\mathbf{V}^{dn} \in \mathbb{Z}_{\geq 0}$. Wind and hydropower are aggregated at the provincial level for all formulations and do not require special treatment. Generator parameters such as ($\underline{P}_g$, $\overline{P}_g$) are converted to their clustered equivalent.

This is mostly a "drop-in" formulation, where variables and summations in (1), (2), (11), (19)-(24), (30)-(35) are transformed as (36)-(38) and summed over indices $k \in K$ (and $p \in P$ as appropriate). Integer commitment variables are constrained by the number of clustered units:

$$\mathbf{U}_{p,k,t} \leq |G_{p,k}|\,, \forall p \in P, k \in K, t \in T \tag{39}$$

Furthermore, some modifications to the commitment state equations (16)-(18) are required:

$$\mathbf{U}_{p,k,t} \geq \sum_{t'=t-MU_k}^{t} \mathbf{V}^{up}_{p,k,t'} \tag{40}$$

$$|G_{p,k}| - \mathbf{U}_{p,k,t} \geq \sum_{t'=t-MD_k}^{t} \mathbf{V}^{dn}_{p,k,t'} \tag{41}$$

$$\mathbf{U}_{p,k,t} - \mathbf{U}_{p,k,t-1} = \mathbf{V}^{up}_{p,k,t} - \mathbf{V}^{dn}_{p,k,t} \tag{42}$$

$$\forall p \in P, k \in K, t \in T$$

Ramping constraints (13)-(15) are modified to account for startups:

$$\mathbf{W}_{p,k,t} = \mathbf{Y}_{p,k,t} - \underline{P}_k \mathbf{U}_{p,k,t} \tag{43}$$

$$\mathbf{W}_{g,t} - \mathbf{W}_{g,t-1} \leq \mathbf{U}_{p,k,t} RU_k + \mathbf{V}^{up}_{p,k,t} \underline{P}_k \tag{44}$$

$$\mathbf{W}_{g,t-1} - \mathbf{W}_{g,t} \leq \mathbf{U}_{p,k,t} RD_k + \mathbf{V}^{dn}_{p,k,t} \underline{P}_k \tag{45}$$

$$\forall p \in P, k \in K, t \in T$$

Finally, some reserve constraints (19)-(20), (23)-(24), (30)-(31) are transformed $g \to (p,k)$ in a similar manner as above. The remainder (21)-(22) require a slight modification to account for the number of committed units:

$$\mathbf{R}_{p,k,t} \leq \mathbf{U}_{p,k,t} RU_k \tag{46}$$

$$\mathbf{S}_{p,k,t} \leq \mathbf{U}_{p,k,t} RD_k \tag{47}$$

$$\forall p \in P, k \in K, t \in T$$

IV. Experimental Setup

We demonstrate these formulations on the Northeast China Grid (NE), one of five major grid regions in the State Grid Corporation of China, consisting of four distinct service territories: Heilongjiang, Jilin, Liaoning, and Eastern Inner Mongolia. The NE grid is recognized for its high degree of operational inflexibility, owing to the large penetration of coal-fired CHP, relative lack of flexible generation such as hydropower and natural gas, and overcapacity in thermal generation [46]. Overcapacity would tend to increase the relative importance of quotas as more thermal generators must be accommodated, and collectively these indicate signs of coupled political and technical constraints tractable with the modified UC model. Among other outward signs of operational difficulties, the NE has consistently high amounts of wind curtailment, reaching 32%, 21% and 10% in Jilin, Heilongjiang and Liaoning provinces in 2015 [47]. In our formulation, the NE grid is considered to be isolated from neighboring grids, based on its limited external interconnections: only 21.5 TWh, or 5% of total generation, was exported to North China Grid in 2014 [48].

The experimental setup consists of running the UC model over a one-week horizon (168-hour time step) in the winter season, averaging results over different scenarios of wind production keeping all other inputs (e.g., demand) constant. Periodic boundary conditions are assumed and the entire week's results are kept. The horizon should be longer than typical operational scales of 1-3 days in order to justify the assumption of aggregating production quotas

Table I

GENERATING CAPACITIES IN NORTHEAST GRID AT THE END OF 2010 (MW). SOURCE: [49]

	HL	JL	LN	IME
Thermal	16,901	13,647	27,649	11,774
Hydro	888	4,185	1,817	0
Wind	1,861	2,209	3,381	3,232

across like units as in (32)-(35). One week also aligns with the typical minimum commitment times in practice across multiple regions in China [34]. Longer model horizons, such as a month, would be more accurate in terms of the quota and monthly generation plans, but would increase problem size and could lead to complications in terms of considering maintenance scheduling of units.

The models are implemented in GAMS and solved using ILOG CPLEX 12.6.2. Each scenario is run using up to 8 parallel threads on a dual-socket 12-core 2.5 GHz Intel Xeon machine with 128 GB RAM. The MIP optimality tolerance is set to 10^{-3} and resource limit to 360 minutes.

A. Generator Composition and Characteristics

The Northeast China Grid in our reference year has three basic types of generators: thermal, hydropower, and wind (see Table I). Thermal units are assumed to be entirely coal as biomass and natural gas-fired generators have very little penetration in the NE. Our reference year is also prior to the installation of any nuclear plants in the region. Coal-fired units in China range in size from 6 MW up to 1000 MW for the most advanced units. This wide distribution of unit sizes impacts efficiency and generator constraints important for commitment and dispatch schedules, and will be the main source of variation in production costs for the system. As smaller units are typically older and slated first for early retirement under strong energy efficiency policy incentives, the exact composition is changing each year. An historical database of thermal plant-level data for 2011 (calculated at end of 2010) was chosen as an authoritative source for this analysis [49].

These plant level data were further converted to unit-level data (for the case of multiple units inside the fence) and identified as electricity-only or CHP through information obtained in the plant names and cross-checking with generation company websites and other public sources. The total thermal capacity obtained matches aggregate provincial statistics [50]. While there are some published aggregate statistics on fractions of CHP [46], [51], these are highly varied, likely depending on whether all units inside the fence or just the active subset are considered CHP, and the method of assigning cut-offs for CHP. We feel our unit-level approach is consistent with the purposes of the modeling exercise, and run sensitivities varying the effective must-run outputs. This resulted in 507 thermal units.

Next, we clustered thermal units into six different sizes we observed frequently during the above cross-checking: 25, 50, 135, 200, 350 and 600 MW. Combined with the binary CHP indicator, this leads to 12 clusters per province (see Table VIII). As the specific cogeneration technology – extraction-condensing or backpressure – is not available for all units, 25 MW was used a cut-off for backpressure, above which all are assumed extraction-condensing. These sizes were used to determine the heat rates of the various generators, in particular, heat rates are assigned in all formulations based on the clustered generator type, not on individual heat rates, which are unobserved. See Table IX for full generator details.

In previous clustering studies [52], size thresholds were defined for each fuel type (e.g., small, medium, and large) and all units within each class were homogenized into a single type of generator with size equal to the mean of the aggregated units. This approach would lead to smaller average capacities relative to the thresholds.

Instead, in this work, units were clustered according to the closest size threshold (either above or below). In order to have comparable unit types across provinces, we let the homogenized unit be the average capacity of units in all provinces for a given type (see Table IX). This leads to deviations in total capacity by province compared to the full set of units, the implications of which we explore below.

Finally, by fixing the minimum generating outputs of must-run cogeneration units, we must specify exogenously which units will be committed. This can be calculated from the total must-run capacity for each province. Again, previous work [46], [51] can guide this determination, but are also subject to data availability concerns. As our reference, we remove cogeneration units from each province roughly equally across sizes in order to achieve an approximate 80% commitment rate. We explore the implications of this choice with an additional sensitivity.

B. Network

Identifying each province as a node, inter-provincial transmission is modeled using a transport model satisfying Kirchhoff's first law. Connections between major nodes in the NE grid range from 300-800 km and are primarily at 500kV [53]. Resistive loss coefficients $\mu_{p,p'}$ were estimated by summing in parallel the resistive losses for each line type, as given by typical lines of given voltage and whether they are AC or DC [54], [55]: for example, a 500-kV line with 1000-MW loading has a loss rate of 1.3% per 100 miles (161 km) [54] .

Publicly-available government and grid company sources typically list only voltages and numbers of transmission lines. Using 900-MW as maximum loading for a 500 kV line and 200-MW for a 200 kV line at a distance of 500 km [56], interconnection capacities were estimated as in Table II. Inter-regional transmission from NE to North China Grid was ignored because of the low interchanges.

In practice, as transmission capacity is allocated on an annual basis, with difficulties to respond to short-term system conditions, these transmission interconnections are not utilized to their full capacity. Based on historical transmission amounts, the effective transmission interconnection capacities $\overline{F}^*_{p,p'}$ under provincial dispatch (see Sec III-D1) were estimated as in Table III by assuming constant loading throughout the year. Furthermore, in accordance with clear export/import relationships in government documents [57], some transmission interconnections were assumed to be uni-directional. We note that Heilongjiang and Liaoning, which do not share an interconnection, have an export/import relationship in transmission pricing and summary statistics [57], [58]. This presumably reflects coordination in the long-term energy and transmission plan allocation process of over-generation in Heilongjiang and under-generation in Liaoning, as politicians are primarily concerned with total energy transfers. Rather than

Table II

ESTIMATED INTER-PROVINCIAL TRANSMISSION CAPACITIES $\overline{F}_{p,p'}$ (MW) IN 2011. (HL = HEILONGJIANG, JL = JILIN, LN = LIAONING, IME = EASTERN INNER MONGOLIA)

	HL	JL	LN	IME
HL	0	4500	0	1800
JL	4500	0	3600	600
LN	0	3600	0	8000
IME	1800	600	8000	0

Table III

MODELED EFFECTIVE INTER-PROVINCIAL TRANSMISSION CAPACITIES $\overline{F}^{*}_{p,p'}$ UNDER PROVINCIAL DISPATCH (MW). SOURCE OF EXPORTS (2011): [57].

	Exports (PWh)	Avg. power (MW)	$\overline{F}^{*}_{p,p'}$ (MW)
HL $\rightarrow$ JL	0.119	14	0
HL $\rightarrow$ LN	5.257	600	600
HL $\rightarrow$ IME	0.426	49	0
JL $\rightarrow$ LN	2.579	294	300
IME $\rightarrow$ LN	10.622	1213	1200

constrain the demand balances at these two nodes to meet energy transfer requirements,we simplify by establishing an artificial direct transmission link, whose distance is equivalent to the sum of the intermediate paths. As only Kirchhoff's first law is modeled in our network representation (and the series resistances are equal for these two lines), these two approaches are equivalent.

Reserve requirements are held constant over the week, set at the province by its peak load, wind capacity and largest unit as contingency. Additional reserve requirements due to wind power are difficult to estimate because of inaccurate wind power forecast error models, and correlations with other imbalances (i.e., load) in the system [59]. In a meta-survey for systems of up to 10% wind penetration by energy, 4% of wind capacity was the reasonable upper end of additional procurement necessary (excluding one outlier study at 15%) [60]. Regulation and load-following reserves in total are set at 3% of peak load:

$$\begin{aligned} \overline{RES}_p &= 3\% \cdot MaxLoad_p + 4\% \cdot WindCap_p \\ &\quad + LargestUnit_p \end{aligned} \tag{48}$$

$$\underline{RES}_p = 3\% \cdot MaxLoad_p + 4\% \cdot WindCap_p \tag{49}$$

where $LargestUnit_p = 700$ MW is a contingency reserve, $\overline{RES}_p$ is the up reserve requirement for p, and $\underline{RES}_p$ is the down reserve requirement. This results in up reserve requirements of $6.1\% \sim 17.4\%$ depending on the province (see Table X). Region-wide reserves, in scenarios where reserves can be shared across borders, are the sums of and replace provincial requirements (9.1% for up reserves).

C. Demand, Wind and Hydropower Resource Profiles

A single representative week of electricity load from March 2011 for each province is used for all scenarios, shown in Figure 1. This was reconstructed from daily consumption data [61] to represent daily and weekly variation in the winter of the NE grid, similar to the procedure described in [62].

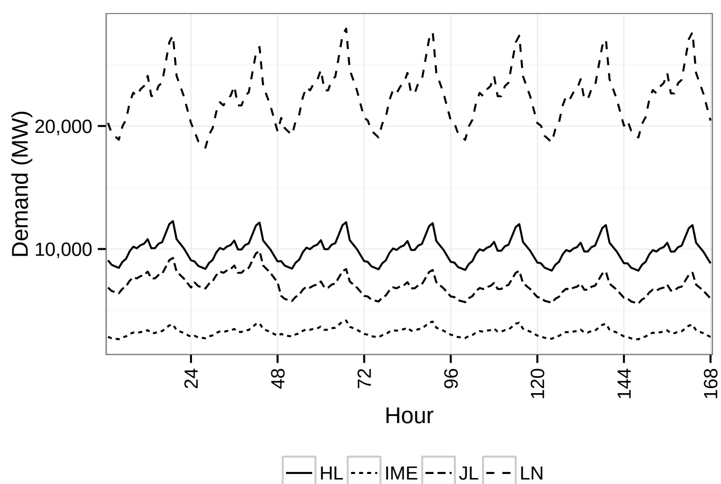

Figure 1. Demand profiles

Province-wide average wind capacity factors were generated as in [62] as follows. Wind resource profiles were calculated using Modern Era Retrospective-analysis for Research and Applications (MERRA) boundary layer flux data, a high temporal resolution (one hour) atmospheric dataset with $0.5°$ latitude by $0.67°$ longitude spatial resolution (approx. 56 km x 61 km at mid-latitudes). Forests, urban areas, slopes greater than 10% and geographic features such as lakes and rivers were excluded and the power curve of a Sinovel 1.5-MW wind turbine with 82-meter hub height, common in Chinese onshore applications, was used. Smoothed province-wide wind capacity factors for each hour were then constructed by an average of hour production in all cells within the province weighted by the available land area.

To capture variability of wind resources, six weeks from a recent year of average annual wind in the dataset (2009) were chosen: three each from January and March, winter months when wind is most plentiful and constraints from CHP generation are largest. The minimum of these six was also generated to aid with finding an initial feasible integer solution.

Hydropower generators are modeled by (25)-(29) using historical generation data from 2001-2014 for the Northeast provinces [63]. H_g, the mean inflow of generator g over the problem timestep, is given by dividing

Table IV

MODELED MINIMUM WINTER THERMAL GENERATION QUOTAS Q_p BY PROVINCE.

	Annual CF	Assumed Max Summer CF	Estimated Min Winter CF	Q_p
HL	47%	80%	14%	14%
JL	39%	80%	-2%	0%
LN	50%	80%	20%	20%
IME	58%	80%	36%	36%

generation equally throughout the winter months (Jan-Mar). $HL_{g,1} = HL_{g,|T|} \gg H_g$ are the fixed initial and final levels. The minima of hydro generation in each month over the period were used, accounting for 1.6% of generation. The main results are robust to taking the maxima of hydro generation over the period: increasing hydro availability in the model decreases total production costs but does not change the relative impacts of regulatory formulations on objectives or wind outcomes.

D. Generation Quota

The quota system is designed to ensure equitable dispatch and revenue sufficiency for generators. These assignments can vary substantially across provinces. For example, as the NE has excess thermal capacity, the generation that can be allocated per coal plant is depressed, which would have the effect of reducing quotas.

The quota is an annual minimum constraint on individual generators, but clustering allows us to consider seasonal averages over sums of similar type units as in (32). To calculate the heating season ("winter") average, we assume that CHP units will achieve most of their quota during the high must-run winter months. By contrast, electricity-only units will predominate in the non-winter heating season ("summer"). If the maximum capacity factor they can achieve in summer is 80% due to availability of units and common loadings, then we can roughly approximate the minimum capacity factor they must achieve in the winter based on annual averages (see Table (IV)). Due to data availability, we use average thermal capacity factors in 2012 [64] and adjust for inter-annual changes. We also let the quotas be constant across all sizes (i.e., $Q_{p,k} = Q_p$).

V. RESULTS

A. Solution Performance

The binary formulations, with 125k discrete variables and 344k constraints following presolve, have varied performance in terms of solution times and optimality gaps when the coupling quota constraint is activated. We adjusted the scale of the problem using SCALEOPT in GAMS and changed the following CPLEX options, resulting in significant gains:

Table V

SOLUTION TIMES FOR BINARY (FULL UNITS), AGGREGATED-BINARY (12-TYPE) AND AGGREGATED-INTEGER (CLUSTERED) FORMULATIONS. (MINUTES)

RUN	FULL UNITS	12 TYPE	CLUSTER
R	18.19	12.90	1.59
P	22.37	15.26	1.35
RT	160.00	480.52*	2.32
PT	101.14	1996.62*	1.84
RQ		317.68	4.46
PQ		774.21*	40.16
RTQ		627.33*	9.42
PTQ		2522.04*	69.53

R: Regional reserves. P: Provincial reserves.

T: Limited transmission. Q: Quota.

*One or more wind scenarios did not solve to optimality.

- MIPSTART. Solve first the minimum wind scenario and use this as initial feasible integer solution for each of the wind profiles.
- Barrier algorithm for linear subproblems. In contrast to the default dual simplex method, the barrier method (interior point) reached optimality more quickly, likely because of the large, sparse constraint matrices.
- Relaxation Induced Neighborhood Search (RINS). After a specified number of nodes explored (we tested 100), the solver searches in the neighborhood of the current incumbent for an improved solution, thus potentially paring down lengthy search trees.

Solution times for all formulations are shown in Table V, inclusive of solving the initial minimum wind scenario. In moving from the full binary formulation (considering each unit's reported minimum and maximum output, and scaled ramp rates) to the aggregated binary formulation (units are homogenized into one of the twelve categories), solution times unexpectedly increase dramatically for the constrained transmission case. In addition, several wind scenarios do not converge, and a handful do not even find a feasible solution. In the clustered formulation (integer commitments), times reduced by 30-1300x compared to the aggregated-binary.

The poor performance of the aggregated formulation compared to the full binary is likely attributable to the degeneracies of similar units, and further work could examine in what circumstances we might expect greater time penalties. The addition of transmission losses also affects solution times – increasing times in the aggregated case, but *decreasing* solution times by an order of magnitude or larger for the provincial-reserves transmission-constrained (PT) full binary.

B. Effects of Aggregation and Clustering

The standard model results, in the absence of regulatory constraints, show the aggregated model (binary decision variables) produces high capacity factors from must-run cogeneration units, wind, and high-efficiency coal (coal600).

All other generators are relatively unused, and production from low-efficiency non-cogeneration units are basically zero (see Figure 2).

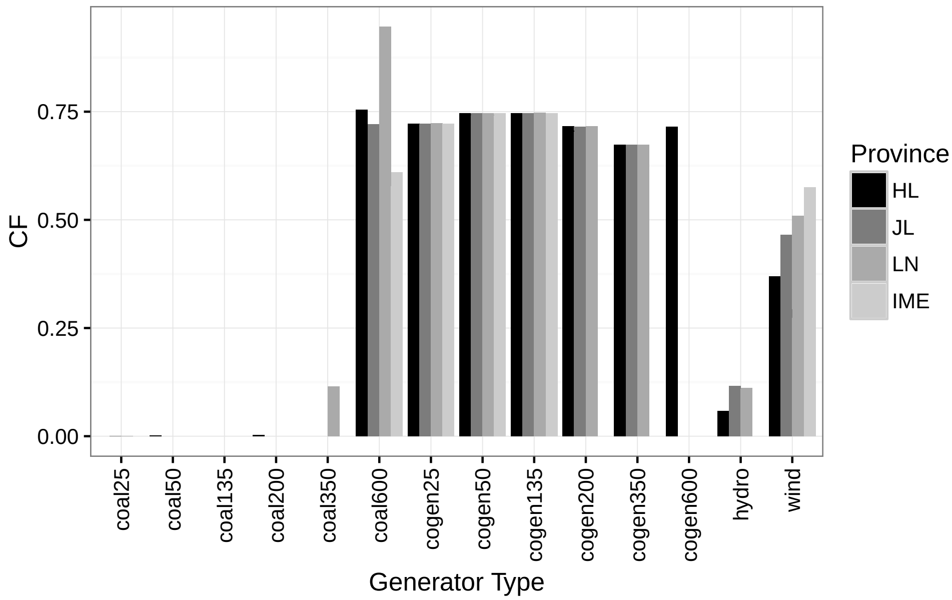

Figure 2. Capacity factors of generation types by province in 12-type aggregation, reference case.

Examining the effects of aggregation and clustering, these two sequential simplification steps have a small impact on two outcome variables of interest: objective and wind total. Comparing the aggregated binary (12-type) and aggregated integer (Clustered) formulations, the errors introduced with respect to using the full set of units and unmodified capacities are very small over the entire system: objectives are within 0.02%, and wind totals within 0.14% (see Figure 3). These errors are magnified at the individual provincial node in the objective, ranging from $-1.4\% \sim +2.4\%$ for the 12-type and $-2.1\% \sim +3.1\%$ for the clustered formulations. Wind totals at the province are within $\pm 0.75\%$. Collectively, these demonstrate that clustering can be used on this simple network with the given set of generator parameters (in particular, heat rates are assigned in all formulations based on the aggregated generator type, not on unobserved individual heat rates).

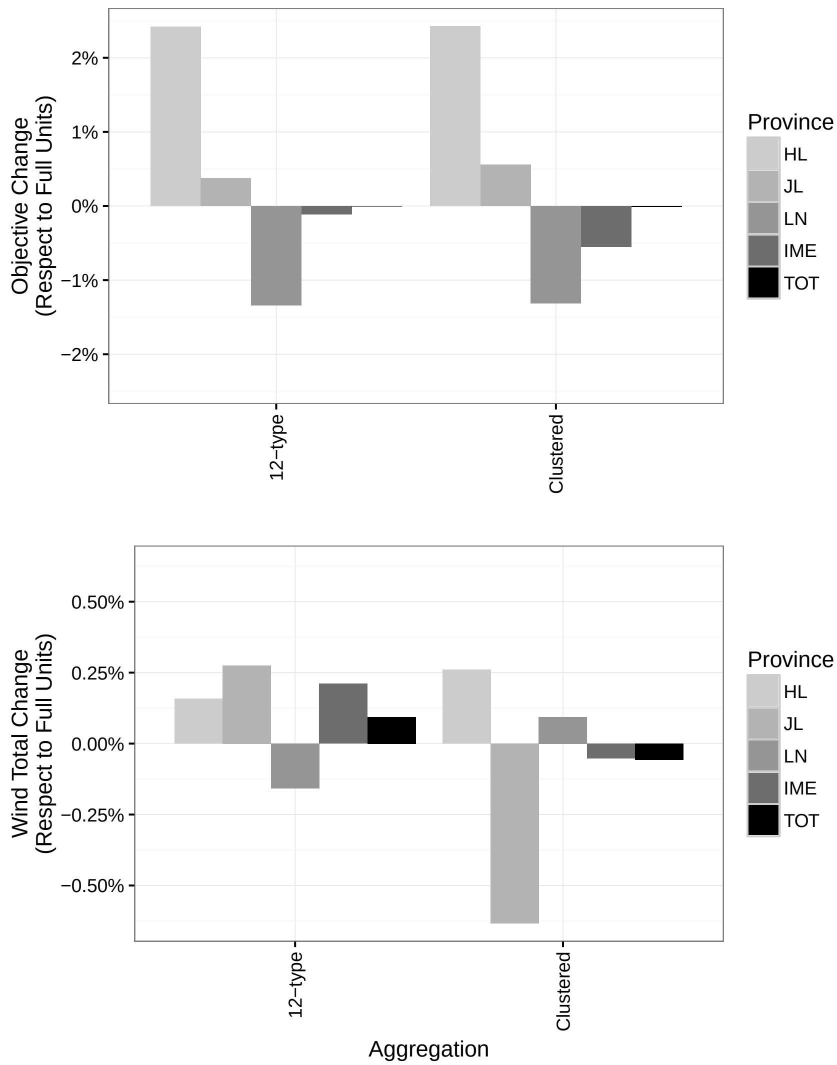

Figure 3. Aggregation errors of objective and wind totals by province for aggregated-binary (12-type) and aggregated-integer (Clustered), reference case.

C. Comparing Quota Formulations

Comparing the two formulations for the quota – constraint-driven (CON) given by (32), and penalty-driven (PEN) given by (33)-(35) – we find that PEN leads to lower solution times and better convergence in some constrained cases (e.g., provincial reserves) (see Table VI). In the aggregated-integer (Cluster) formulation, PEN also generally shows better solution times. Methodologically, the various aggregation models achieve similar results, but there are large time savings (30-1300x) in the integer formulation, demonstrating a tractable tool for policy analysis of the quota coupling constraint.

D. Effects of Political Constraints

In the next several sections are results addressing the key policy questions presented in this paper, namely the effects of three aspects of China's partial liberalization on our outcomes of interest, objective and wind generation.

Table VI

SOLUTION TIMES FOR QUOTA FORMULATIONS: CONSTRAINT-DRIVEN (CON) AND PENALTY-DRIVEN (PEN). (MINUTES)

RUN	12 TYPE	CLUSTER
RQ-CON	317.68	4.46
RQ-PEN	354.77	6.20
RTQ-CON	627.33*	9.42
RTQ-PEN	631.74*	9.58
PQ-CON	774.21*	40.16
PQ-PEN	483.12	23.66
PTQ-CON	2522.04*	69.53
PTQ-PEN	2522.13*	54.07

R: Regional reserves. P: Provincial reserves.

T: Limited transmission. Q: Quota.

*One or more wind scenarios did not solve to optimality.

As rates of curtailment are more useful from a policy perspective, we first highlight differences in these rates across the political constraint scenarios. Next, we highlight the specific impacts of constraining transmission interconnection capacity (long-term barriers to trade in our parlance). We conclude with two sensitivities: examining the impact of changing the quota, and changing the amount of must-run CHP units.

As we add these constraints, we observe total production costs increasing, though the magnitude depends highly on the interaction of the imposed regulations (see Figure 4). For example, implementing restrictions that disallow inter-provincial reserve sharing while leaving other aspects untouched (P in Figure 4) does not differ from the reference case (R). However, the combination of within-province reserve requirements and limited transmission interconnection (PT) decreases flexibility, causing increased costs and higher wind curtailment. The specific requirement to meet winter-adjusted generation quotas for electricity-only coal plants also increases costs, but does not significantly affect wind curtailment for any combination of other parameters.

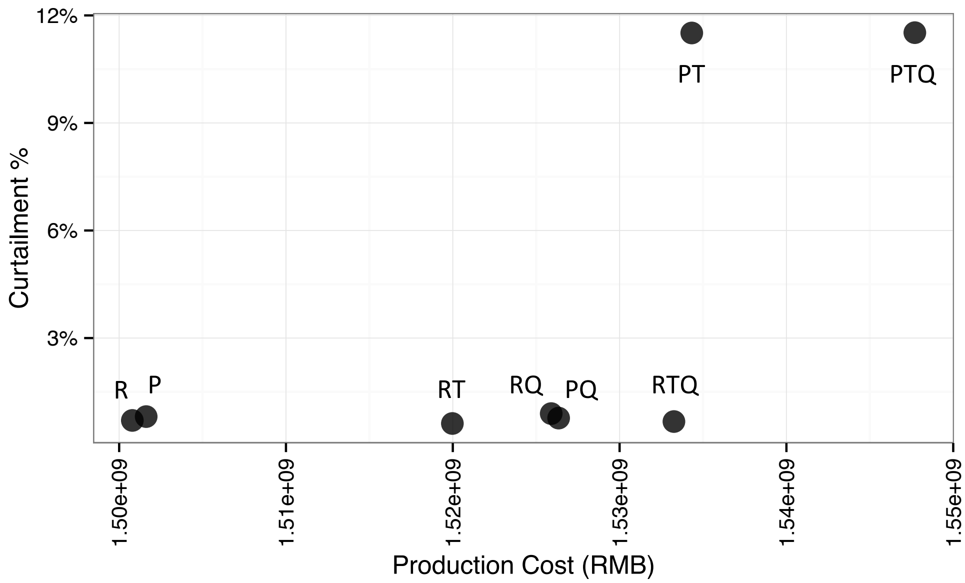

Figure 4. Objective and wind curtailment by regulatory formulations (Clustered). R=Regional reserves, P=Provincial reserves, T=Limited transmission, Q=Quota.

E. Experiments with Quota Parameters

Clustering commitment variables into integers allows us to test the effect of varying regulatory parameters over a wider range, both as sensitivities as well as to identify implications of policy changes. We show this for the case of modifying the quota in Figure 5. In it, we change uniformly for each province the ratio of the quota with respect to the default quotas in Table IV, i.e. a ratio of 1.0 is the default, and 0.0 is the absence of a quota.

As the quota increases, the effect of limited transmission on the objective decreases, so that we see convergence of RTQ, PQ and RQ under high quotas. The interaction of transmission and within-province reserves is robust, however, to changes in quota. The effect on wind curtailment is essentially flat for all values of the quota, demonstrating robustness of the results in Figure 4.

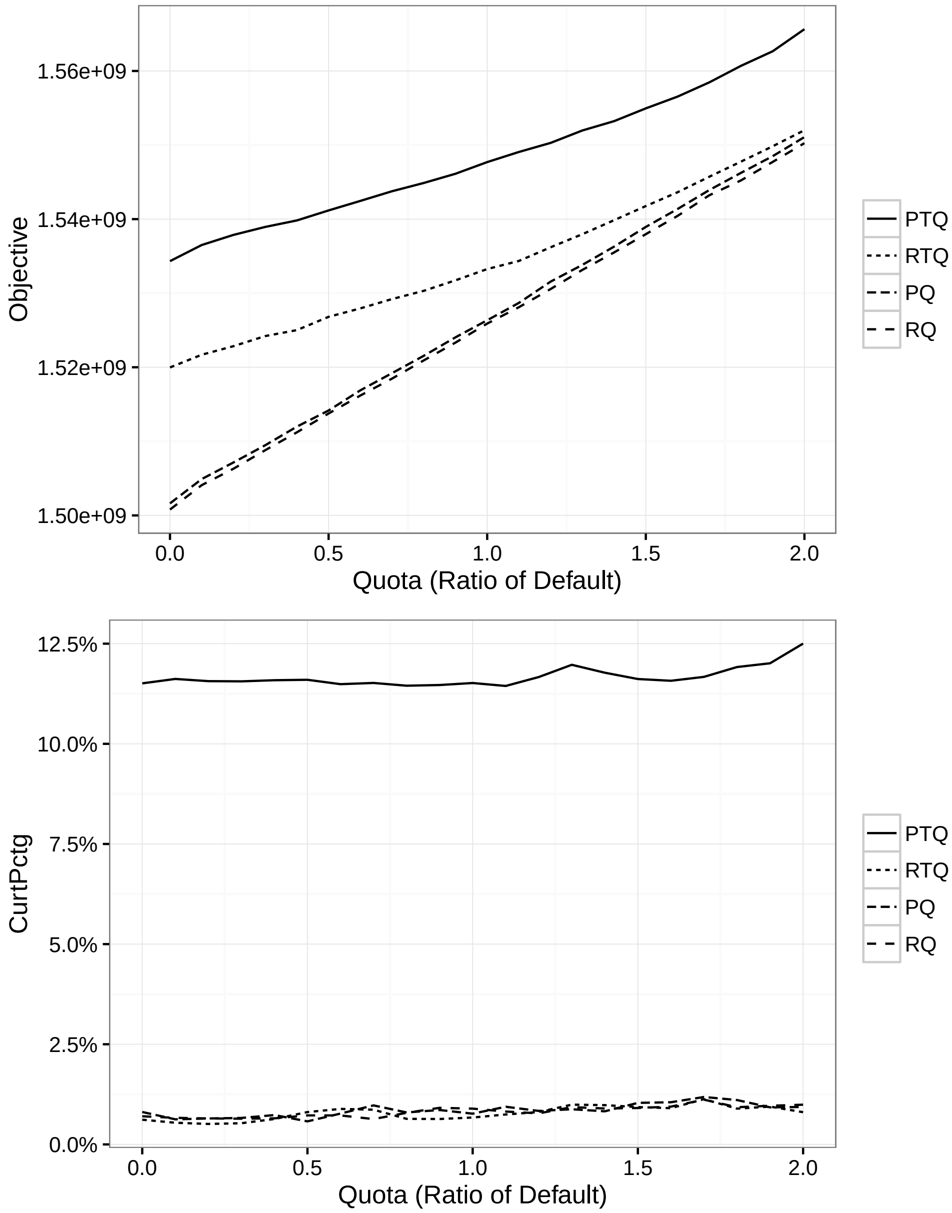

Figure 5. Objective and wind curtailment as a function of quota. R=Regional reserves, P=Provincial reserves, T=Limited transmission, Q=Quota.

Another way to examine the impact of the quota is to adjust the penalty p^Q in (33)-(35) to be a compensation for the unmet quota. This could be facilitated through, e.g., the ongoing generation rights trading (GRT) programs in several Chinese provinces that allow some generators to sell their quota [39]. Generator switching and increased total welfare occurs for penalties less than the difference in marginal costs of substituted generator types. In Figure 6, coal use is plotted with vertical lines equal to the difference in cost with respect to the most efficient unit coal600.

We see that benefits of this program in terms of reduced coal use can be achieved if compensation for foregone quotas is ~50 Yuan/MWh. Though, these benefits are reduced in half if provincial trade restrictions persist: 10 ktce compared to 20 ktce if these regulations are removed. These support findings from government reports that indicate provincial trading barriers as significant challenges to exchanging quotas [39]. Further work could compare these to opportunity costs of various generation types to determine what, if any, of these trades will be Pareto improving for the inefficient generators.

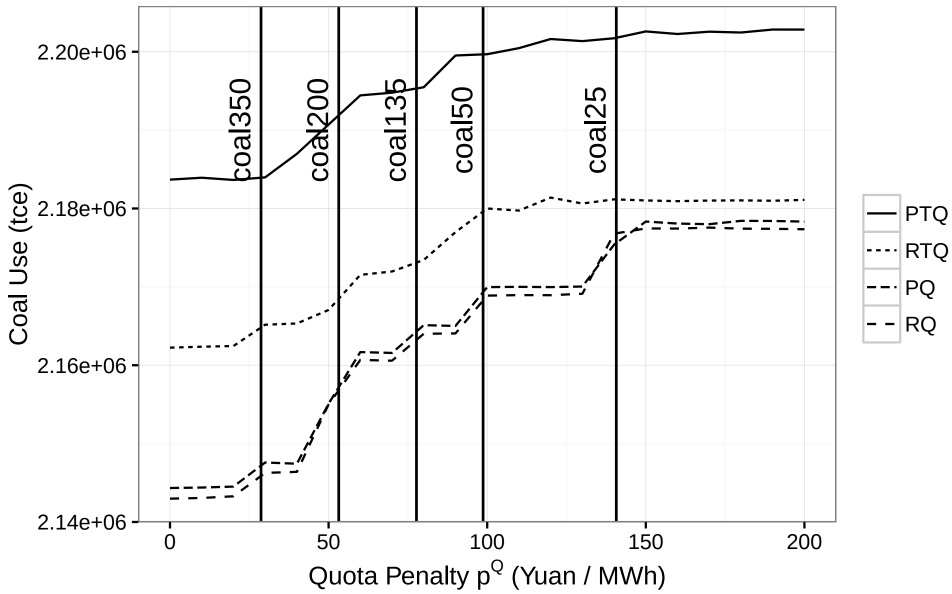

Figure 6. Coal use (tons-coal-equivalent) as a function of quota penalty/compensation . Vertical lines are differences in marginal costs of generator types with respect to coal600.

F. Sensitivity to Must-Run Cogeneration

As must-run levels are difficult to verify and have a large impact on system flexibility, we perform a sensitivity on the commitment rate of must-run cogeneration. Compared to our reference (~80% commitments), a higher must-run threshold of ~90% commitments increases the fraction of the minimum load in provinces that must be met by cogeneration units: in the most extreme case, Jilin, this ratio rises from 78% to 87%. Costs increase as more generation is substituted away from high-efficiency electricity-only generators to smaller cogeneration units. However, wind curtailment is relatively insensitive to adjusting this parameter over this range (see Figure 7). Must-run thresholds and minimum loads for each province are in Table VII.

Table VII
MUST-RUN THRESHOLDS BY COMMITMENT RATE, AND MINIMUM LOAD BY PROVINCE (MW)

	80%	90%	Min Load
HL	3790	4194	8241
JL	4334	4831	5571
LN	4567	5201	18236
IME	556	553	2657

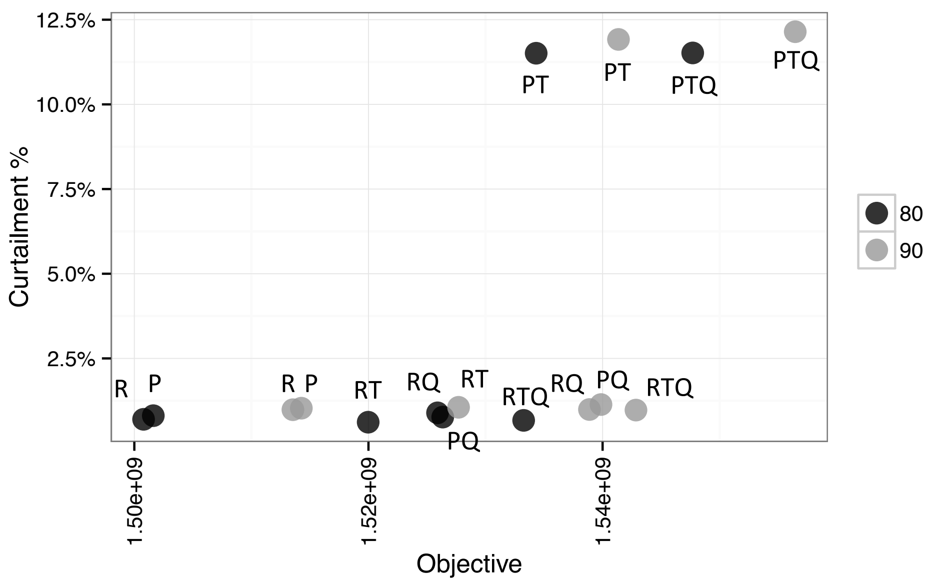

Figure 7. Objective and wind curtailment by regulatory formulations for two must-run parameterizations, 80% and 90% commitment rates (Clustered). R=Regional reserves, P=Provincial reserves, T=Limited transmission, Q=Quota.

VI. DISCUSSION

Restructuring an electricity sector entails a complex realignment of political and economic institutions, which may both delay and distort the achievement of satisfactorily competitive conditions, and therefore efficiency. In research and planning for policy interventions in power systems under these varied regulatory environments, typical models assuming ideal operations may neglect important interactions between techno-economic criteria and political constraints, leading to poor understanding of underlying causes of inefficiency and to inappropriate recommendations. In this work, we have developed tractable formulations of a “sub-optimal” unit commitment problem that endogenize important political factors in a major grid region of China, demonstrating the importance of these on operations, and providing a set of options for researchers to explore further pathways for China’s ongoing power system reforms.

A quintessential feature of operations in China is the quota system, which allocates on an annual basis minimum generation amounts to generators that must be met in an equitable manner by the grid companies' dispatch centers, with implications for transmission capacity allocations and other decisions. This is a long-term coupling constraint similar to hydro-thermal coordination or maintenance scheduling, giving rise to similar modeling trade-offs in terms of time horizons and numbers of decision variables. After demonstrating the adequacy of our aggregation technique, we find that the quota is not the primary political factor driving wind integration challenges. Under cost-minimizing dispatch, wind curtailment increases dramatically when inter-provincial trade is constrained in both the short-term (reserves) and the long-term (effective interconnection). Just one of these two sources of inflexibility alone is insufficient to change significantly wind integration outcomes. This highlights interactive effects of technical and political constraints that can only be captured in a unified model such as the one presented.

A well-known cause of inflexibility is the high must-run threshold of cogeneration units during the winter heating season. We show that while this has clear impacts on total objective costs, it does not fully explain poor flexibility leading to high wind curtailment either. Improving inter-provincial transmission can help add flexibility during key winter heating hours while still satisfying generation quotas. Conversely, advancing restructuring efforts may be a necessary prerequisite to achieve proposed benefits of increased heat-electricity system flexibility, such as introducing flexible cogeneration and heat storage [65].

Both of the simplified unit commitment models presented (aggregated and aggregated-clustered) provide several valuable avenues for further research. The models, while respecting the constraints of grid operators, assume perfect forecasts and a single optimizing agent. This should be seen as the best-case scenario for operating in political context, and additional studies into actual dispatch practices can create heuristics that capture the larger inflexibilities observed. Intra-provincial transmission constraints, ignored in this analysis, may be binding in some regions with rapid wind expansion and could be considered with a more detailed network.

Examination of the impact of individual political constraints can provide guidance for the relative importance of reform options under consideration to achieve near-efficient outcomes and facilitate other policy priorities such as renewable energy integration. Quantifying the benefits of, for example, improving coal unit commitment scheduling and minimum generation outputs can highlight the cost-benefit trade-offs inherent in modifying the current primarily administrative scheduling practices. Future work can expand to other network and generator configurations, and explore optimal unit aggregation techniques.

Acknowledgement

The authors would like to thank A. Ramos for advice on enhancing solution performance, and F. Kahrl and A. Papavasiliou for valuable feedback on this manuscript.

We thank the founding sponsors of the China Energy and Climate Project, Eni S.p.A., ICF International, Shell International Limited, and the French Development Agency (AFD), for supporting this work. We are grateful to the Energy Information Administration, U.S. Department of Energy for supporting this work through a cooperative agreement.

REFERENCES

[1] P. Joskow, “Lessons learned from electricity market liberalization,” *The Energy Journal*, vol. 29, no. 2, pp. 9–42, 2008.

[2] T. Jamasb, “Between the state and market: Electricity sector reform in developing countries,” *Utilities Policy*, vol. 14, pp. 14–30, Mar. 2006.

[3] J. Williams and R. Ghanadan, “Electricity reform in developing and transition countries: A reappraisal,” *Energy*, vol. 31, no. 6-7, pp. 815–844, 2006.

[4] Correlje and de Vries, “Hybrid Electricity Markets: The Problem of Explaining Different Patterns of Restructuring,” in *Competitive electricity markets: design, implementation, performance*, Amsterdam: Elsevier, 2008.

[5] C. Teixeira, M. Albano, A. Skou, L. P. Dueñas, F. Antonacci, R. Ferreira, K. Lotzfeldt Pedersen, and S. Scalari, “Convergence to the European Energy Policy in European Countries: Case Studies and Comparison,” *Social Technologies*, vol. 4, no. 1, pp. 7–24, 2014.

[6] R. Mullin, “CAISO Revised Western RTO Plan Highlights State Authority,” July 2016.

[7] P. Andrews-Speed, “Reform Postponed: The Evolution of China’s Electricity Markets,” *Evolution of Global Electricity Markets: New Paradigms, New Challenges, New Approaches*, pp. 531–567, 2013.

[8] F. Kahrl, M. Dupuy, and W. Xuan, “Issues in China Power Sector Reform: Generator Dispatch,” tech. rep., Regulatory Assistance Project (RAP), Beijing, 2016.

[9] S. Council, “Opinion Regarding Deepening Electricity Sector Reform,” Tech. Rep. 9, State Council, Beijing, Mar. 2015.

[10] NEA, “Notice for Comment Regarding Completing Work for Electricity Market Construction,” tech. rep., National Energy Administration, Beijing, 2016.

[11] N. Padhy, “Unit Commitment: A Bibliographical Survey,” *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1196–1205, 2004.

[12] 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, pp. 39–46, Feb. 2012.

[13] B. Palmintier and M. Webster, “Heterogeneous Unit Clustering for Efficient Operational Flexibility Modeling,” *IEEE Transactions on Power Systems*, vol. 29, pp. 1089–1098, May 2014.

[14] A. Ramos, S. Cerisola, and J. M. Latorre, “A Decision Support System for Generation Planning and Operation in Electricity Markets,” *Handbook of Power Systems Ii*, pp. 315–335, 2010.

[15] A. Papavasiliou, S. Oren, and R. O’Neill, “Reserve Requirements for Wind Power Integration: A Scenario-Based Stochastic Programming Framework,” *IEEE Transactions on Power Systems*, vol. 26, pp. 2197–2206, Nov. 2011.

[16] Morales-Espana, J. M. Latorre, and A. Ramos, “Tight and Compact MILP Formulation for the Thermal Unit Commitment Problem,” *IEEE Transactions on Power Systems*, vol. 28, pp. 4897–4908, Nov. 2013.

[17] M. Carrión and J. M. Arroyo, “A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem,” *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1371–1378, 2006.

[18] F. J. de Sisternes and M. D. Webster, “Selection of Sample Weeks for Approximating the Net Load in Generation Planning Problems,” ESD Working Paper, Massachusetts Institute of Technology, 2013.

[19] S. Cerisola, A. Baíllo, J. Fernández-López, A. Ramos, and R. Gollmer, “Stochastic Power Generation Unit Commitment in Electricity Markets: A Novel Formulation and a Comparison of Solution Methods,” *Operations Research*, vol. 57, pp. 32–46, Feb. 2009.

[20] F. Kahrl and X. Wang, “Integrating Renewables into Power Systems in China: A Technical Primer - Power System Operations,” tech. rep., The Regulatory Assistance Project, Beijing, 2014.

[21] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn, *Spot Pricing of Electricity*. The Kluwer International Series in Engineering and Computer Science, Power Electronics & Power Systems, Boston, MA: Springer, 1988.

[22] I. J. Pérez-Arriaga and C. Meseguer, “Wholesale marginal prices in competitive generation markets,” *Power Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 710–717, 1997.

[23] I. Aravena and A. Papavasiliou, “Renewable Energy Integration in Zonal Markets,” *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[24] C. Ruiz, A. J. Conejo, J. D. Fuller, S. A. Gabriel, and B. F. Hobbs, “A tutorial review of complementarity models for decision-making in energy markets,” *EURO Journal on Decision Processes*, vol. 2, pp. 91–120, Oct. 2013.

[25] S. A. Gabriel, A. J. Conejo, C. Ruiz, and S. Siddiqui, “Solving discretely constrained, mixed linear complementarity problems with applications in energy,” *Computers & Operations Research*, vol. 40, pp. 1339–1350, May 2013.

[26] C. Ma and L. He, "From state monopoly to renewable portfolio: Restructuring China's electric utility," *Energy Policy*, vol. 36, pp. 1697–1711, May 2008.

[27] S. Xu and W. Chen, "The reform of electricity power sector in the PR of China," *Energy Policy*, vol. 34, pp. 2455–2465, Nov. 2006.

[28] C.-E. Bai and Y. Qian, "Infrastructure development in China: The cases of electricity, highways, and railways," *Journal of Comparative Economics*, vol. 38, pp. 34–51, Mar. 2010.

[29] S. Council, "Opinion Regarding Issues Deepening Electricity Sector Reform (No. 146)," tech. rep., State Council, Beijing, 1998.

[30] S. Council, "Electricity Sector Reform Plan," tech. rep., State Council, Beijing, 2002.

[31] J. L. Ma, "On-grid electricity tariffs in China: Development, reform and prospects," *Energy Policy*, vol. 39, pp. 2633–2645, May 2011.

[32] F. Kahrl, J. H. Williams, and J. Hu, "The political economy of electricity dispatch reform in China," *Energy Policy*, vol. 53, no. 0, pp. 361–369, 2013.

[33] I. Pérez-Arriaga, ed., *Regulation of the Power Sector*. London: Springer, 2013.

[34] M. R. Davidson, C.-T. Li, and V. J. Karplus, "Grid Operations and Renewable Energy Integration in China. Mimeo (based on interviews conducted in Northwest China and Inner Mongolia), forthcoming as a Working Paper," tech. rep., MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA, 2017.

[35] CNSE, "El Funcionamiento del Mercado Eléctrico en el Año 1998," tech. rep., Comisión Nacional del Sistema Eléctrico, Madrid, 2000.

[36] NDRC, MEP, SERC, and NEWG, "Pilot Measures for Energy Efficient Dispatch," tech. rep., National Development and Reform Commission, Beijing, Aug. 2007.

[37] SERC, "2010 Annual Electricity Regulation Report," tech. rep., State Electricity Regulatory Commission, Beijing, 2011.

[38] J. Tong, X. Zhang, J. Ren, Y. Tang, X. Xu, X. Jiang, and N. Li, "Energy-Efficient Dispatch Implementation Analysis and Policy Recommendations," tech. rep., Energy Foundation, 2011.

[39] SERC, "2012h1 Inter-Provincial Energy Trading and Generation Rights Trading Report," tech. rep., State Electricity Regulatory Commission, Beijing, 2012.

[40] Y.-c. Xu, *Electricity Reform in China, India and Russia: The World Bank template and the politics of power*. Cheltenham: Edward Elgar, 2005.

[41] D. Z. Fitiwi, L. Olmos, M. Rivier, F. de Cuadra, and I. J. Pérez-Arriaga, "Finding a representative network losses model for large-scale transmission expansion planning with renewable energy sources," *Energy*, vol. 101, pp. 343–358, Apr. 2016.

[42] X. Zhao, F. Wang, and M. Wang, "Large-scale utilization of wind power in China: Obstacles of conflict between market and planning," *Energy Policy*, vol. 48, pp. 222–232, Sept. 2012.

[43] R. Bartnik and Z. Buryn, *Conversion of Coal-Fired Power Plants to Cogeneration and Combined-Cycle*. London: Springer London, 2011.

[44] A. Y. Rong, H. Hakonen, and R. Lahdelma, "An efficient linear model and optimisation algorithm for multi-site combined heat and power production," *European Journal of Operational Research*, vol. 168, Jan. 2006.

[45] NEA, "2015 National Electricity Dispatch Exchange and Market Operations Supervision Report," tech. rep., National Energy Administration, Beijing, 2016.

[46] X. Zhao, S. Zhang, R. Yang, and M. Wang, "Constraints on the effective utilization of wind power in China: An illustration from the northeast China grid," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 4508–4514, Sept. 2012.

[47] NEA, "Status of the Wind Industry (2012-2015)," tech. rep., National Energy Administration, 2016.

[48] NECG, "Northeast Grid 2014 Electricity Market Exchange Report," tech. rep., Northeast China Grid, Beijing, 2015.

[49] CEC, "2010 Electricity Industry Statistical Collection," tech. rep., China Electricity Council, Beijing, 2011.

[50] CEC, "Provincial Electricity Generating Capacities (2010)," tech. rep., CEIC Data, 2016.

[51] X. Zhao, S. Zhang, Y. Zou, and J. Yao, "To what extent does wind power deployment affect vested interests? A case study of the Northeast China Grid," *Energy Policy*, vol. 63, pp. 814–822, Dec. 2013.

[52] B. S. Palmintier and M. D. Webster, "Heterogeneous Unit Clustering for Efficient Operational Flexibility Modeling for Strategic Models," tech. rep., Massachusetts Institute of Technology, 2013.

[53] D. Gao, Q. Li, Y. Jin, J. Yu, and J. Zhang, "Analysis and Research of Operational Characteristics of Wind Power Generation in Northeast Power Grid," *Electric Power Technology*, vol. 19, no. 2, 2010.

[54] P. Interconnection, "A Survey of Transmission Cost Allocation Issues, Methods and Practices," Tech. Rep. Docket No. ER05-121-006, Federal Energy Regulatory Commission, 2010.

[55] L. Weimers, "Bulk power transmission at extra high voltages, a comparison between transmission lines for HVDC at voltages above 600 kV DC and 800 kV AC."

[56] AEP, "Transmission facts," tech. rep., AEP.

[57] State Grid, "2011 Power Exchange Report," tech. rep., State Grid, Beijing, 2012.

[58] NDRC, "Notice on the adjustment of electricity generation company on-grid tariffs (No. 1942)," tech. rep., National Development and Reform Commission, 2013.

[59] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Gomez-Lazaro, and N. Detlefsen, "Methodologies to Determine Operating Reserves Due to Increased Wind Power," *IEEE Transactions on Sustainable Energy*, vol. 3, pp. 713–723, Oct. 2012.

[60] H. Holttinen, M. R. Milligan, and E. Ela, *Design and operation of power systems with large amounts of wind power: final report, IEA Wind Task 25, Phase One, 2006-2008*. Espoo, Finland: VTT Technical Research Centre of Finland, 2009.

[61] State Electricity Regulatory Commission (SERC), "Daily Electricity Consumption by Province," tech. rep., CEIC Database, Oct. 2013.

[62] M. R. Davidson, D. Zhang, W. Xiong, X. Zhang, and V. J. Karplus, "Modelling the potential for wind energy integration on China's coal-heavy electricity grid," *Nature Energy*, vol. 1, p. 16086, 2016.

[63] National Bureau of Statistics (NBS), "Monthly Electricity Production (2001-2014)," tech. rep., CEIC Data, 2015.

[64] China Electricity Council (CEC), "Power Generating Equipment Utilization: Thermal Power," tech. rep., CEIC Database, 2014.

[65] X. Chen, C. Kang, M. O'Malley, Q. Xia, J. Bai, C. Liu, R. Sun, W. Wang, and H. Li, "Increasing the Flexibility of Combined Heat and Power for Wind Power Integration in China: Modeling and Implications," *IEEE Transactions on Power Systems*, vol. 30, pp. 1848–1857, July 2015.

NOMENCLATURE

Sets:

$g \in G$: generators

$t \in T$: time periods

$p \in P$: provincial nodes

$k \in K$: clustered generator types

$G_p \subset G$: generators in province p

$G_{p,k} \subset G_p$: generators of cluster type k in province p

$G_{wind} \subset G$: wind generators

$G_{hydro} \subset G$: hydro generators

$G_{res} \subset G$: generators providing reserves

$G_{thermal} \subset G$: thermal generators

$G_{CHP} \subset G_{thermal}$: combined heat and power generators

Decision Variables:

$\boldsymbol{u}_{g,t} \in \{0,1\}$: commitment status of generator g at time t

$\boldsymbol{v}^{up}_{g,t} \in \{0,1\}$: startup of generator g at time t

$\boldsymbol{v}^{dn}_{g,t} \in \{0,1\}$: shutdown of generator g at time t

$\boldsymbol{y}_{g,t} \geq 0$: production of generator g at time t

$\boldsymbol{r}_{g,t},\, \boldsymbol{s}_{g,t} \geq 0$: available up and down reserve capability of generator g at time t

$\boldsymbol{w}_{g,t}$: auxiliary ramping variable of generator g at t

$\boldsymbol{f}_{p,p',t}$: flow from p to p' at time t

$\boldsymbol{f}^+_{p,p',t},\, \boldsymbol{f}^-_{p,p',t}$: positive and negative components of $\boldsymbol{f}_{p,p',t}$

$\boldsymbol{l}_{p,p',t}$: transmission losses due to flow $\boldsymbol{f}_{p,p',t}$

$\boldsymbol{j}_{p,p',t,s}$: sth piece-wise segment of the flow $\boldsymbol{f}_{p,p',t}$

$\boldsymbol{h}_{g,t}$: hydro reservoir level of hydro generator g at t, in units of generation

$\mathbf{Y}_{p,k,t} \geq 0$: production of cluster k in p at time t

$\mathbf{W}_{p,k,t}$: auxiliary ramping variable, cluster k in p at time t

$(\mathbf{U}_{p,k,t}, \mathbf{V}^{up}_{p,k,t}, \mathbf{V}^{dn}_{p,k,t}) \in (\mathbb{Z}_{\geq 0})^3$: commitment variables in clustered formulation

$\mathbf{R}_{p,k,t}$, $\mathbf{S}_{p,k,t} \geq 0$: up and down reserve capabilities in clustered formulation

Parameters:

$d_{p,t}$: demand at p at time t

p^{var}_g: variable cost of generator g

p^{su}_g: startup cost of generator g

$\underline{P}_g$, $\overline{P}_g$: minimum and maximum outputs of generator g

$\overline{F}_{p,p'}$: transmission flow limit from p to p'

$\mu_{p,p'}$: quadratic resistive loss coefficient of path p to p'

$W_{g,t}$: available wind power of generator g at time t

RD_g, RU_g: down and up ramp rates of generator g

MD_g, MU_g: minimum down and up times of generator g

$\underline{RES}_t$, $\overline{RES}_t$: down and up reserve requirements at time t

$\underline{RES}_{p,t}$, $\overline{RES}_{p,t}$: down and up provincial reserve requirements in p at time t

H_g: mean hydro inflow of generator g over a timestep

$HL_{g,t}$, $t = \{1, |T|\}$: initial and final levels of generator g

$Q_{p,k}$: minimum generation quota at p for generator type k

p^Q: cost penalty for unmet minimum generation quota

Table VIII

CLUSTERED UNITS AND CAPACITIES (MW) OF THERMAL GENERATORS (MW)

	HL		JL		LN		IME	
	Units	Capacity	Units	Capacity	Units	Capacity	Units	Capacity
coal25	50	822	32	526	39	641	2	33
coal50	11	627	6	342	7	399	2	114
coal135	1	136	4	544	3	408	2	272
coal200	9	1,827	1	203	14	2,842	5	1,015
coal350	1	326	1	326	2	652	3	978
coal600	11	6,713	6	3,661	23	14,036	14	8,543
cogen25	65	1,410	8	174	69	1,497	9	195
cogen50	11	583	2	106	5	265	1	53
cogen135	4	508	4	508	11	1,397	5	635
cogen200	3	600	20	4,000	10	2,000	0	0
cogen350	10	3,060	10	3,060	10	3,060	0	0
cogen600	1	600	0	0	0	0	0	0
Totals	177	17,212	94	13,450	193	27,197	43	11,839

Table IX

GENERATOR PARAMETERS FOR AGGREGATED TYPES

	$\underline{P}_g$ (MW)	$\overline{P}_g$ (MW)	RD_g, RU_g (MW / h)	MU_g, MD_g h	p_g^{var} Yuan/MWh	p_g^{su} kYuan
coal25	9	16	2	3	350	10
coal50	31	57	9	3	308	34
coal135	73	136	20	6	287	82
coal200	110	203	30	6	263	122
coal350	176	326	49	12	238	196
coal600	330	610	92	12	209	366
cogen25	16	22	3	3	350	13
cogen50	39	53	8	3	308	32
cogen135	94	127	19	6	287	76
cogen200	140	200	30	6	263	120
cogen350	201	306	46	12	238	184
cogen600	360	600	90	12	209	360

Table X

PROVINCIAL RESERVE REQUIREMENTS (MW)

	Down		Up	
	MW	%ge of peak load	MW	%ge of peak load
HL	424	3.6%	1124	9.6%
JL	325	4.1%	1025	13.0%
LN	952	3.5%	1652	6.1%
IME	302	5.2%	1002	17.4%
Total	2004	3.8%	4804	9.1%

MIT CEEPR

MIT Center for Energy and Environmental Policy Research

MIT Center for Energy and Environmental Policy Research
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139
USA

Website: ceepr.mit.edu

MIT CEEPR Working Paper Series is published by the MIT Center for Energy and Environmental Policy Research from submissions by affiliated researchers.

Email ceepr@mit.edu
Phone (617) 253-3551
Fax (617) 253-9845