Annals of Operations Research 96 (2000) 167–189

Unit commitment in power generation – a basic model and some extensions

Ralf Gollmer [a], Matthias P. Nowak [b], Werner Römisch [b] and Rüdiger Schultz [a]

[a] *Department of Mathematics, Gerhard-Mercator University Duisburg, Lotharstr. 65, D-47048 Duisburg, Germany*

[b] *Institute of Mathematics, Humboldt-University Berlin, Unter-den-Linden 6, D-10099 Berlin, Germany*

For the unit commitment problem in the hydro-thermal power system of VEAG Vereinigte Energiewerke AG Berlin we present a basic model and discuss possible extensions where both primal and dual solution approaches lead to flexible optimization tools. Extensions include staggered fuel prices, reserve policies involving hydro units, nonlinear start-up costs, and uncertain load profiles.

1. Introduction

Optimization problems in the power industry have attracted researchers from engineering, operations research and mathematics for many years. The complex nature of generation, transmission, and distribution of electric power implies ample opportunity of improvement towards the optimal. Mathematical models have proven indispensable in deepening the understanding of these optimization problems. The progress in algorithms and implementations has an essential share in widening the abilities to solve these optimization problems on hardware that is permanently improving.

In the present paper we address unit commitment in power operation planning. This problem concerns the scheduling of start-up/shut-down decisions and operation levels for power generation units such that the fuel costs over some time horizon are minimal. The diversity of power systems regarding technological design and economic environment leads to a variety of issues potentially occurring in mathematical models of unit commitment. The ongoing liberalization of electricity markets will add to this by shifting the objective in power planning from fuel cost minimization to revenue maximization. For an introduction into basic aspects of unit commitment the reader is referred to the book by Wood and Wollenberg [35]. A literature synopsis on various traditional methodological approaches has been compiled by Sheble and Fahd [29]. In our paper, we present some of the more recent issues in modeling and algorithms for unit commitment.

The present paper grew out of a collaboration with the German utility VEAG Vereinigte Energiewerke AG Berlin whose generation system comprises conventional coal and gas fired thermal units as well as pumped-storage plants. An important

feature of that system is its amount of installed pumped-storage capacity that enables the inclusion of pumped-storage plants into the optimization. This contrasts with other utilities where pumped-storage energy mainly serves reserve purposes and has only limited impact on smoothing the load curve.

The mentioned technological and economic diversity of power planning in general has its natural counterparts in the more specific setting of unit commitment. However, there are quite a few basic features that unit commitment problems in hydro-thermal systems like the VEAG one have in common. In our paper, this is reflected by the model we call the basic one. It is a large-scale mixed-integer optimization problem coupled both in time and with respect to the different generation units. For its solution we employ primal and dual algorithms, and it will turn out that both approaches have their merits and shortcomings. These pros and cons become essential when heading for model extensions. A main purpose of the present paper is to report on our experiences with model extensions and to indicate which approach (primal vs. dual) offers advantages for a given model extension.

Our paper is organized as follows. In section 2 we present our basic model together with the basics of the primal and the dual solution methods. Section 3 is devoted to extensions where primal solution techniques offer sufficient flexibility while dual ones have to be modified substantially. We address reserve policies beyond spinning reserve involving pumped-storage energy and include staggered fuel prices depending on total fuel consumption. In section 4 we elaborate extensions where dual approaches are the more flexible ones. Here we emphasize accurate modeling of unit start-up costs and we extend the model towards the inclusion of uncertain load profiles.

2. Basic model and solution approaches

2.1. Model

The planning horizon for unit commitment, although in principle continuous in time, gives rise to a discretization into (hourly, half-hourly, or even shorter) subintervals. This is due to the availability of load data, the execution time for scheduling decisions, and, last but not least, the algorithmic capabilities which prevent handling the complex mixed-integer decision problem in continuous time. Therefore, unit commitment models typically are in discrete time.

Let T denote the number of subintervals of the optimization horizon and suppose there are I thermal as well as J pumped-storage hydro units. The variable $\boldsymbol{u}_i^t \in \{0,1\}$, $i = 1,\ldots,I$; $t = 1,\ldots,T$, indicates whether the thermal unit i is in operation at time t. Variables $\boldsymbol{p}_i^t$, $\boldsymbol{s}_j^t$, $\boldsymbol{w}_j^t$, $i = 1,\ldots,I$; $j = 1,\ldots,J$; $t = 1,\ldots,T$, are the output levels for the thermal units, the hydro units in generation and in pumping modes, respectively. The variables $\boldsymbol{l}_j^t$ denote the fill (in energy) of the upper dam of the hydro unit j at the end of interval t, $j = 1,\ldots,J$; $t = 1,\ldots,T$.

Generically, the objective function to be minimized can be expressed as

$$\sum_{t=1}^{T}\sum_{i=1}^{I} C_i\big(\boldsymbol{p}_i^t, \boldsymbol{u}_i^t\big) + \sum_{t=1}^{T}\sum_{i=1}^{I} S_i^t(\boldsymbol{u}_i), \tag{1}$$

where C_i denotes the fuel costs and S_i^t denotes the start-up costs for the thermal unit i.

In our basic model, the fuel costs C_i are affine linear in $\boldsymbol{p}_i^t$. A more detailed modeling here leads to convex piecewise linear, convex quadratic or even nonconvex piecewise linear continuous curves, although it has to be admitted that all these curves are fairly close to straight lines.

The start-up costs S_i^t of the basic model incur a unit-dependent but down-time-independent cost a_i per start-up. This can be expressed by $a_i \max\{\boldsymbol{u}_i^t - \boldsymbol{u}_i^{t-1}, 0\}$, and, in a minimization framework, this cost function can be transformed in a standard way into a linear one plus additional linear constraints. In contrast to the fuel costs where nonlinearities are mostly negligible, here an essential nonlinearity has been excluded – the down-time dependence of start-up costs, which leads to exponential terms or at least to step functions. We will see that the latter prevents application of our primal approach but can be handled via the dual method.

When modeling the constraints, again our emphasis is on mixed-integer linear terms, although sometimes there are elegant nonlinear alternatives. This is motivated by the far more powerful algorithmic tools that are available for mixed-integer linear problems. In light of the dimensionality of our mixed-integer decision problem, it currently seems utopic to handle nonlinearities directly.

The power output of units and the fill of the upper dams have to fit the following bounds:

$$\begin{aligned} p_{it}^{\min}\boldsymbol{u}_i^t \leqslant \boldsymbol{p}_i^t \leqslant p_{it}^{\max}\boldsymbol{u}_i^t, &\quad i = 1,\dots,I,\ t = 1,\dots,T,\\ 0 \leqslant \boldsymbol{s}_j^t \leqslant s_{jt}^{\max}, &\quad j = 1,\dots,J,\ t = 1,\dots,T,\\ 0 \leqslant \boldsymbol{w}_j^t \leqslant w_{jt}^{\max}, &\quad j = 1,\dots,J,\ t = 1,\dots,T,\\ 0 \leqslant \boldsymbol{l}_j^t \leqslant l_j^{\max}, &\quad j = 1,\dots,J,\ t = 1,\dots,T. \end{aligned} \tag{2}$$

Here, $p_{it}^{\min}$, $p_{it}^{\max}$, $s_{jt}^{\max}$, $w_{jt}^{\max}$ denote minimal and maximal outputs, respectively, and $l_j^{\max}$ is the maximal fill of the upper dam.

Load coverage is modeled by the constraints

$$\sum_{i=1}^{I} \boldsymbol{p}_i^t + \sum_{j=1}^{J} \big(\boldsymbol{s}_j^t - \boldsymbol{w}_j^t\big) \geqslant D^t, \quad t = 1,\dots,T, \tag{3}$$

where D^t denotes the electrical load at time t.

Reserve management is an important security issue in power system optimization. In case of failure of generation equipment or of sudden load peaks, the power system will be able to mobilize reserves within a prescribed time schedule. The latter being quite sophisticated reserve capacities, they are usually not included to full extent into

the optimization. Some basic issues, however, have to be incorporated, of which the following spinning reserve R^t of the thermal units is mandatory:

$$\sum_{i=1}^{I} \left(\boldsymbol{u}_i^t p_{it}^{\max} - \boldsymbol{p}_i^t\right) \geqslant R^t, \quad t = 1, \ldots, T. \tag{4}$$

Our basic model will restrict the reserve management to the above conditions. The peculiarity of the VEAG system with its considerable share of pumped-storage capacity, however, necessitates an inclusion of pumped-storage units into the reserve scheme. This leads to further constraints that couple different units, and, as we shall see later, such coupling constraints over units, of which our basic model will contain just (3) and (4), provide major challenges in the dual approach.

For the pumped-storage plants we have the following balances that interconnect different time intervals:

$$\begin{aligned} &\boldsymbol{l}_j^t = \boldsymbol{l}_j^{t-1} - \left(\boldsymbol{s}_j^t - \eta_j \boldsymbol{w}_j^t\right), \quad j = 1, \ldots, J; \\ &\boldsymbol{l}_j^0 = l_j^{\text{in}}, \quad \boldsymbol{l}_j^T = l_j^{\text{end}}, \qquad\quad t = 1, \ldots, T. \end{aligned} \tag{5}$$

Here, $l_j^{\text{in}}, l_j^{\text{end}}$ are the initial and final fills (in energy) of the upper dams, η_j denote the pumping efficiencies. The latter are known to be nonlinear functions of the fill in the upper dam, and our above argument on finding a proper compromise between accurate modeling of nonlinearities and the ability to tackle the complex decision problem applies at this place. Test runs for the VEAG system have confirmed the above approximation to be tolerable.

Constraints avoiding simultaneous generation and pumping in the hydro plants are dispensable since it can be shown that such a deficiency can not occur in optimal points.

To avoid excessive thermal stresses in the coal fired blocks, they have to adhere to minimum down times τ_i (and sometimes also to minimum up times). For the VEAG system down times are relevant. They are modeled via

$$\begin{aligned} &\boldsymbol{u}_i^{t-1} - \boldsymbol{u}_i^t \leqslant 1 - \boldsymbol{u}_i^l, \quad i = 1, \ldots, I; \; t = 2, \ldots, T-1; \\ &\quad l = t+1, \ldots, \min\{t + \tau_i - 1, T\}. \end{aligned} \tag{6}$$

Of course, these conditions have to be adapted for units that are off-line at time $t = 1$ and still within their minimum down time.

This completes the description of our basic model. An issue left out above but often considered to be basic is ramping, i.e., imposing bounds on the speed of load changes for a given unit in consecutive time steps. In the present paper, the finest time discretization will be an hourly one. At least for the VEAG system ramping then is no longer critical. This is different, of course, for half-hourly or even finer discretizations.

2.2. Primal approach

LP-based branch-and-bound is among the earliest mathematical approaches to unit commitment, cf. [29]. Early branch-and-bound approaches to unit commitment suffered from the comparatively poor mathematical methodology and software technology at that time. In the meantime, this has changed drastically, and general purpose codes like the CPLEX Callable Library [4] combine latest LP-methodology with a variety of options for arranging the branch-and-bound. The major advantage of the primal approach via LP-based branch-and-bound is that model enrichment is possible as long as this is expressible in mixed-integer linear terms. In particular, further internal coupling of the model caused by additional constraints has no structural implications for the algorithm. On the other hand, the full model has to be handled, which may become prohibitive even if advanced methods are used for the LP-relaxations.

The above statements are illustrated by our test runs with the basic model, cf. tables 1, 2, and figure 1. These were performed with real-life VEAG data for hourly discretized time horizons of 1 week, 1 month, and 6 months. The generation system comprises 34 thermal and 7 hydro units. Since the system includes some thermal units of identical design, the basic model was slightly changed by aggregating the Boolean variables for these units into proper integer variables. To indicate the increase in effort when enriching the model we ran instances where the basic 1-step start-up costs are replaced with a 3-step function. Accuracy of solutions was measured by the relative gap between the best feasible solution and the minimum lower LP-bound at active nodes in the search tree.

Table 1
Model dimensions.

Model dimensions	Variant with groups of aggregated units and 1-step start-up costs			Variant with individual units and a 3-step function for start-up costs		
	1 week	1 month	6 months	1 week	1 month	6 months
Integer variables	2112	8184	56472	5420	20832	130320
Real variables	9781	37867	217608	15210	65442	383033
Constraints	8053	31237	204576	22902	83364	594619
Nonzeroes	31448	121877	760110	196803	749430	6363009

Table 2
Computing times on a HP 9000 (770/J180) and accuracy bounds.

CPU-time and accuracy	Variant with groups of aggregated units and 1-step start-up costs			Variant with individual units and a 3-step function for start-up costs		
	1 week	1 month	6 months	1 week	1 month	6 months
CPU-time (min)	0:58.9	7:40.9	234:02.9	7:44.3	161:32.9	out of
accuracy bound (%)	0.086	0.073	0.133	0.391	0.389	memory

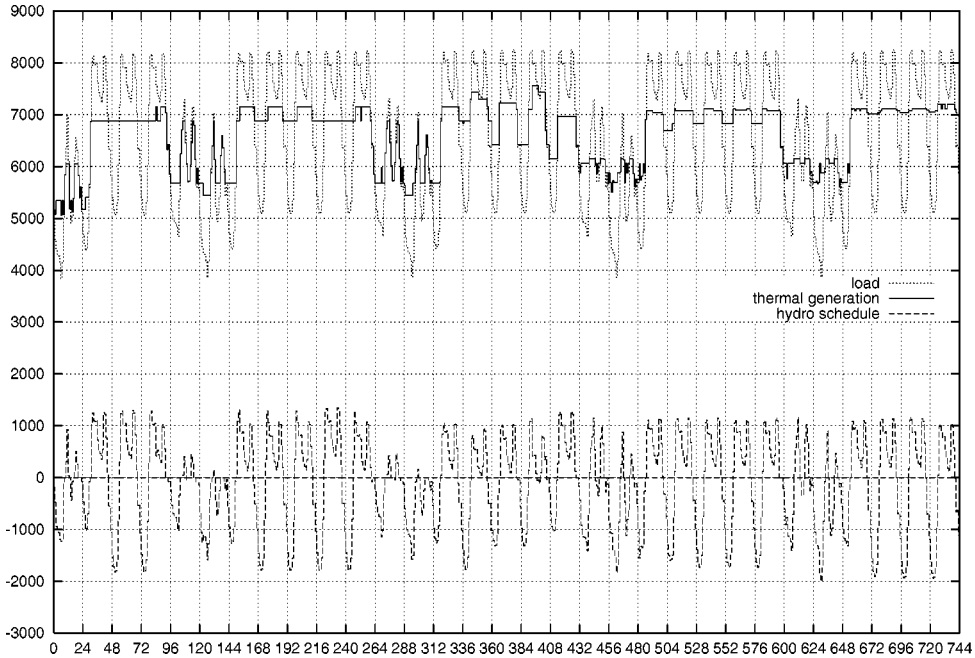

Figure 1. Solution of the primal method for 1 month.

2.3. Dual approach

The constraints from the basic model fall into two groups: (2), (5), and (6) concern single units and possibly interconnect time intervals; (3) and (4) concern single time intervals but interconnect units. The basic idea of the dual approach is to perform a Lagrangian relaxation of load and reserve constraints (3), (4) leading to a decomposition of the model into single-unit subproblems.

In unit commitment, Lagrangian relaxation is very popular and has a long history [29,30]. Recently, three aspects have made Lagrangian relaxation even more attractive and applicable to large-scale instances: the algorithmic progress in solving the nondifferentiable Lagrangian dual, the usually small relative duality gap and the progress in fast Lagrangian heuristics for good primal feasible solutions.

Early approaches to the dual problem were based on subgradient methods and smoothing techniques (cf. [29]). During the last decade more refined and efficient methods became available: variants of cutting plane and bundle methods for convex nondifferentiable minimization (cf. [13]). Here, dynamically constrained cutting plane methods [15], bundle-trust algorithms [21], reduced complexity bundle methods [20], variable metric bundle methods [19], and proximal bundle methods [6,9,10,12] have to be mentioned. Moreover, dual convergence properties of proximal bundle methods are exploited in [9] to derive new Lagrangian heuristics for thermal systems, and [18] provides a novel qualitative study of the duality gap for several Lagrangian relaxation schemes.

For the basic model our Lagrangian relaxation approach associates Lagrange multipliers with the load constraints (3) and the modified reserve constraints

$$\sum_{i=1}^{I} \boldsymbol{u}_i^t p_{it}^{\max} + \sum_{j=1}^{J} \left(\boldsymbol{s}_j^t - \boldsymbol{w}_j^t\right) \geqslant D^t + R^t \quad (t = 1, \ldots, T). \tag{7}$$

The dual problem reads

$$\max_{(\boldsymbol{\lambda},\boldsymbol{\mu}) \in \mathbb{R}_+^T \times \mathbb{R}_+^T} d(\boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{8}$$

where $\boldsymbol{\lambda}, \boldsymbol{\mu}$ are the Lagrange multipliers. The function d is defined by the infimum of the Lagrangian with respect to $(\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{s}, \boldsymbol{w})$ under (2), (5), (6). It has the separable form

$$d(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \sum_{i=1}^{I} d_i(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{j=1}^{J} \widetilde{d}_j(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{t=1}^{T} \left[\boldsymbol{\lambda}^t D^t + \boldsymbol{\mu}^t \left(D^t + R^t\right)\right], \tag{9}$$

where the functions d_j, $\widetilde{d}_j$ are optimal values of single-unit thermal and hydro subproblems:

$$d_i(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \min_{\boldsymbol{u}_i} \left\{ \sum_{t=1}^{T} \left[S_i^t(\boldsymbol{u}_i) - \boldsymbol{\mu}^t \boldsymbol{u}_i^t p_{it}^{\max} + \min_{\boldsymbol{p}_i^t} \left\{ C_i\left(\boldsymbol{p}_i^t, \boldsymbol{u}_i^t\right) - \boldsymbol{\lambda}^t \boldsymbol{p}_i^t \right\} \right] : \right.$$
$$\left. \boldsymbol{u}_i^t, \boldsymbol{p}_i^t \text{ satisfy (2) and (6)} \right\},$$

$$\widetilde{d}_j(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \min_{\boldsymbol{s}_j, \boldsymbol{w}_j} \left\{ \sum_{t=1}^{T} \left(\boldsymbol{\lambda}^t + \boldsymbol{\mu}^t\right)\left(\boldsymbol{w}_j^t - \boldsymbol{s}_j^t\right) : \ \boldsymbol{s}_j^t, \boldsymbol{w}_j^t \text{ satisfy (2) and (5)} \right\}.$$

The inner minimization of the thermal subproblems with respect to $\boldsymbol{p}_i^t$ is done explicitly while the outer minimization with respect to $\boldsymbol{u}_i$ is done by dynamic programming. For the hydro subproblems a fast descent algorithm from [22] is used. Since for the concave dual function d subgradients are available, a powerful bundle-type algorithm [16] is used for solving the Lagrangian dual (8). The optimal value of the dual provides a lower bound for the minimal costs of the basic model and with the optimal multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ we have solutions to the thermal and hydro subproblems. In general these solutions may violate the load and reserve constraints (3) and (7) such that a low-cost (primal) feasible solution has to be determined by Lagrangian heuristics.

The Lagrangian relaxation approach to unit commitment thus basically consists of two steps whose realizations are both essential for the total performance of the method: the solution of the dual (8) including initialization of the multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and the determination of a primal feasible solution by a suitable heuristic.

Initialization of the multiplier $\boldsymbol{\lambda}$ is done via a priority list of thermal units in ascending order of their relative costs at maximum output. The initial $\boldsymbol{\lambda}^t$ is set as

the relative costs of the most expensive on-line thermal unit when switching on in list order just as many units as needed for covering the load D^t. The initial $\boldsymbol{\mu}^t$ are zero in all intervals.

The proximal bundle method generates a sequence $(\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$ converging to some optimal multiplier as well as trial points $(\overline{\boldsymbol{\lambda}}_k, \overline{\boldsymbol{\mu}}_k)$ starting with $(\overline{\boldsymbol{\lambda}}_1, \overline{\boldsymbol{\mu}}_1) = (\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1)$. The trial points are used for evaluating subgradients $g(\overline{\boldsymbol{\lambda}}_k, \overline{\boldsymbol{\mu}}_k)$ of the dual function d and its polyhedral upper approximation $\widehat{d}_k(\boldsymbol{\lambda}, \boldsymbol{\mu})$ defined by

$$\min_{j \in J_k} \big\{ d(\overline{\boldsymbol{\lambda}}_j, \overline{\boldsymbol{\mu}}_j) + g(\overline{\boldsymbol{\lambda}}_j, \overline{\boldsymbol{\mu}}_j)^T (\boldsymbol{\lambda} - \overline{\boldsymbol{\lambda}}_j, \boldsymbol{\mu} - \overline{\boldsymbol{\mu}}_j) \big\},$$

where J_k is a subset of $\{1, \ldots, k\}$. At iteration k the next trial point $(\overline{\boldsymbol{\lambda}}_{k+1}, \overline{\boldsymbol{\mu}}_{k+1})$ is selected to belong to

$$\operatorname{argmax} \big\{ \widehat{d}_k(\boldsymbol{\lambda}, \boldsymbol{\mu}) - \tfrac{1}{2}\sigma_k \big\| (\boldsymbol{\lambda} - \boldsymbol{\lambda}_k, \boldsymbol{\mu} - \boldsymbol{\mu}_k) \big\|^2 \colon\ (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^T_+ \times \mathbb{R}^T_+ \big\},$$

where σ_k is a proximity weight. An ascent step to $(\boldsymbol{\lambda}_{k+1}, \boldsymbol{\mu}_{k+1}) = (\overline{\boldsymbol{\lambda}}_{k+1}, \overline{\boldsymbol{\mu}}_{k+1})$ occurs if $d(\overline{\boldsymbol{\lambda}}_{k+1}, \overline{\boldsymbol{\mu}}_{k+1}) \geqslant d(\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k) + \alpha v_k$, where $\alpha \in (0, 1)$ is fixed and $v_k = \widehat{d}_k(\overline{\boldsymbol{\lambda}}_{k+1}, \overline{\boldsymbol{\mu}}_{k+1}) - d(\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$. Otherwise a null step $(\boldsymbol{\lambda}_{k+1}, \boldsymbol{\mu}_{k+1}) = (\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$ improves the next polyhedral function $\widehat{d}_{k+1}$. General strategies for updating σ_k and choosing J_{k+1} are discussed in [16,17]. The method is implemented in [17] such that the cardinality of J_k is bounded and that it terminates if v_k is less than a given (relative) optimality tolerance.

We developed two Lagrangian heuristics for the unit commitment problem. Both of them start with the outcome of the dual optimization and head for proper modifications such that the reserve constraint (7) is fulfilled. Of course, then also the demand constraint (3) can be met.

The first heuristic, LH1 (see [12]), starts with fixing the $\boldsymbol{u}$-components of solutions $(\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{s}, \boldsymbol{w})$ corresponding to optimal multipliers. Then the schedule of the hydro plants is modified with the aim of reducing the value $D^t + R^t + \sum_{j=1}^J [\boldsymbol{w}_j^t - \boldsymbol{s}_j^t]$ for intervals t where (7) is violated. Afterwards, the hydro variables are fixed, and following [36] we search for binary variables $\boldsymbol{u}_i^t$ fulfilling $\sum_{i=1}^I \boldsymbol{u}_i^t p_{it}^{\max} \geqslant D^t + R^t + \sum_{j=1}^J [\boldsymbol{w}_j^t - \boldsymbol{s}_j^t]$. This is done by selecting the interval t where this condition is violated most and computing the increase of $\boldsymbol{\mu}^t$ that, via the minimization behind $d_i(\boldsymbol{\lambda}, \boldsymbol{\mu})$, enforces just as many additional start-ups as necessary to meet the reserve constraint at time t. This is repeated until (7) holds in all intervals. Finally, the economic dispatch problem, i.e., the problem with $\boldsymbol{u}$ fixed, is solved.

The second heuristic, LH2, works as follows: the $\boldsymbol{u}$-components of solutions $(\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{s}, \boldsymbol{w})$ corresponding to optimal multipliers and values $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ in their vicinity are screened. Test runs showed that only a few of these variables change. Fixing the remaining binary decisions drastically reduces dimension and leads to a quite tractable problem. Now either the remaining problem allows for a direct solution with the primal branch-and-bound method from subsection 2.2 or the following heuristic ideas are employed. Again by increasing the $\boldsymbol{\mu}^t$ a start-up vector $\boldsymbol{u}$ is enforced such that

(7) holds in all time intervals. This vector forms the first member in a decreasing sequence of switching decisions. In each step a period t is selected where the available reserve $\sum_{i=1}^{I}(\boldsymbol{u}_i^t p_{it}^{\max} - \boldsymbol{p}_i^t) - R^t$ is large. Then, by a dynamic programming step with the additional constraint $\boldsymbol{u}_i^t = 0$, preceding and consecutive periods are determined where units can be switched off without violating the reserve constraints. For each member of the sequence the economic dispatch problem is solved by a modification of the descent method from [22], and the member with least optimal value determines the output of the heuristic LH2. For further details regarding both heuristics see [23].

The results in table 3 and figure 2 are based on the same data and hardware as for the primal method. The weekly results in table 3 were obtained with the heuristic LH2, and the remaining results with LH1. Compared with the primal approach we have savings in the computing times with wider but still acceptable accuracy bounds.

At this place, let us remark that further test runs indicated substantial improvements over the results in table 3 when refining the linear fuel costs towards convex piecewise linear functions, for details see [23].

Table 3
CPU-time in minutes on HP 9000 (770/J180) and upper bound of the duality gap of the dual method.

NOA 3.0 optimality	Optimization horizon		
tolerance: 10^{-4}	1 week	1 month	6 months
CPU-time (min)	0:19	2:36	60:15
Bound of gap (%)	0.44	0.93	0.84

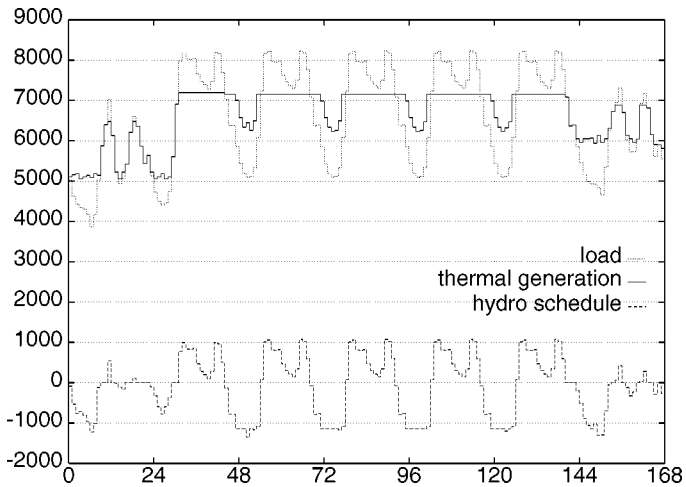

Figure 2. Solution of the dual method for 1 week.

3. Model extensions – primal solution techniques

3.1. Reserve policies

As pointed out in section 2, reserve management is indispensable in power optimization. The spinning reserve constraint (4) reflects the thermal units' ability to contribute to the reserve requirement: on-line units have to maintain a "safety margin". This "margin" is the most expensive part of the capacity put on-line, and in the worst case, the load is already covered but an additional (expensive) unit has to be switched on-line only for maintaining the reserve.

In hydro-thermal systems like the VEAG one, therefore, reserve requirements are distributed among all units including the hydro ones. The essential difference between spinning reserve and reserve in the pumped-storage plants is that the former is available as long as the unit is on-line while the latter can be utilized only with a sufficient fill in the upper dam and hence for a limited time period. Modeling of pumped-storage reserve thus has to involve book-keeping over time. Pumped-storage plants contribute to the reserve in a twofold manner: either by leaving a "margin" towards the hourly maximum output that is in tune with the total fill of the upper dam or by reducing the pumping.

The share of the reserve power that has to be maintained by the pumped-storage plants may either be fixed a priori or left as a variable. In what follows we pursue the first of the two alternatives and denote the share by $R_W^t,\ t = 1,\dots,T$. The hydro units' limitation to contribute to the reserve only for a certain time period implies the necessity to replace hydro reserve with thermal reserve by units that can be started up quickly. The start-up capabilities of these units determine a number h_W of consecutive time intervals in which one may resort to hydro reserve. Furthermore, we have variables $\boldsymbol{\varsigma}_j^t$ for the reserve gained by increased generation and $\boldsymbol{\zeta}_j^t$ for the reserve gained by reduced pumping.

Then the following constraints model the total hydro reserve and the bounds for the generation increases and the pumping decreases:

$$\begin{aligned}\sum_{j\in J}\left(\boldsymbol{\varsigma}_j^t+\boldsymbol{\zeta}_j^t\right)&\geqslant R_W^t,\\ \boldsymbol{s}_j^t+\boldsymbol{\varsigma}_j^t&\leqslant \boldsymbol{s}_j^{\max},\\ \boldsymbol{\zeta}_j^t&\leqslant \boldsymbol{w}_j^t,\quad j\in J,\ t=1,\dots,T.\end{aligned}$$

The book-keeping over an horizon of h_W intervals starting from each time step t in each of the hydro units is reflected by

$$\begin{aligned}\boldsymbol{l}_j^{t-1}+\sum_{l=0}^{k}\left[-\left(\boldsymbol{s}_j^{t+l}+\boldsymbol{\varsigma}_j^{t+l}\right)+\eta_j\left(\boldsymbol{w}_j^{t+l}-\boldsymbol{\zeta}_j^{t+l}\right)\right]&\geqslant 0,\\ j\in J,\ t=1,\dots,T,\ k=0,\dots,h_W.&\end{aligned}$$

Table 4
Model dimensions.

Dimensions	1 week	1 month	6 months
Integer variables	2112	8184	56472
Continuous variables	11313	48283	278424
Constraints	15084	58000	360939
Nonzeroes in matrix	70746	293475	1763308

Table 5
Computing times on a HP 9000 model 770/J180 and accuracy bounds.

	1 week	1 month	6 months
CPU-time (min)	6:06	17:32	1126:59
Accuracy bound (%)	0.083%	0.08%	0.24%

This model extension does not increase the number of integer variables. On the other hand, it increases both the number of continuous variables and, due to the last group of constraints, the number of nonzeroes in the constraint matrix. This leads to increased solution times for the LP-algorithm used within the branch-and-bound procedure.

Table 4 displays the increase of model dimension. While the number of integer variables coincides with the counterpart in table 1, the number of continuous variables goes up by 30% and the number of matrix nonzeroes more than doubles (in the largest model ranging over 6 months). Computing times are listed in table 5. The test runs were performed for the variant with groups of aggregated units and 1-step start-up costs (cf. tables 1 and 2) and with $h_W = 3$.

3.2. Staggered fuel prices

Delivery contracts of power utilities with fuel suppliers often involve discounted fuel prices. The bigger the fuel purchase, the lower the price per unit of fuel. Different units of the generation system may require different qualities of fuel, and usually there is a distinction between the fuels used for operation and for the start-up of a thermal unit. Several generating units may use the same type of fuel.

In our basic model in section 2 we operate with constant fuel prices such that the minimization of fuel costs coincides with the minimization of fuel consumption. Now fuel prices are staggered and, moreover, there is an additional coupling among the units using the same sort of fuel. We assume that the price per unit of fuel follows a decreasing step function such that the fuel costs become piecewise linear and concave (cf. figures 3 and 4).

Suppose that for a given sort of fuel we have prices $f_\imath$, $\imath = 1, \ldots, \mathcal{I}$, holding on intervals $[\xi_{\imath-1}, \xi_\imath]$, $\imath = 1, \ldots, \mathcal{I}$, with $\xi_0 = 0$. For a fuel consumption $\boldsymbol{\xi} \in [\xi_{\bar{\imath}-1}, \xi_{\bar{\imath}}]$

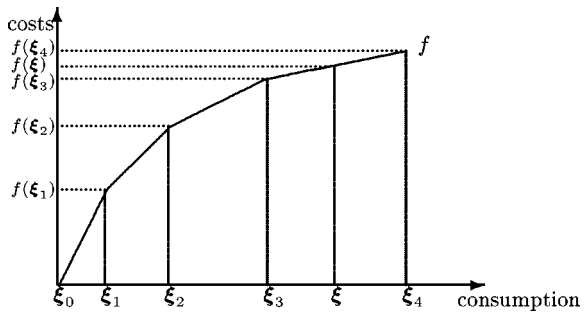

Figure 3. Fuel costs with staggered prices.

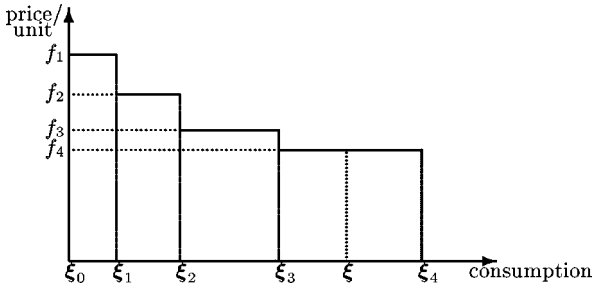

Figure 4. Staggered prices per unit of fuel.

the fuel costs $f(\xi)$ then compute as

$$f(\xi) = \sum_{i=1}^{\bar{i}-1} (\xi_i - \xi_{i-1}) f_i + (\xi - \xi_{\bar{i}}) f_{\bar{i}}.$$

Introducing $\alpha_i = \max\{0, \xi - \xi_i\},\ i = 0, \ldots, \mathcal{I}$, we obtain the equality

$$f(\xi) = \sum_{i=1}^{\mathcal{I}} (\alpha_{i-1} - \alpha_i) f_i \quad \text{for all } \xi \in [0, \xi_{\mathcal{I}}]. \tag{10}$$

Introducing variables $\delta_i \in \{0, 1\}$ and a constant $M \geqslant \xi_{\mathcal{I}}$ we have that

$$\alpha_i = \max\{0, \xi - \xi_i\}, \quad i = 0, \ldots, \mathcal{I},$$

if and only if there exists a solution (α, δ) to the system

$$\begin{aligned} \delta_i M + \xi - \xi_i &\geqslant \alpha_i, \\ M - \delta_i M &\geqslant \alpha_i, \\ \xi - \xi_i &\leqslant \alpha_i, \\ 0 &\leqslant \alpha_i, \\ \delta_i &\in \{0, 1\}, \quad i = 0, \ldots, \mathcal{I}. \end{aligned} \tag{11}$$

Together with the right-hand side of (10) the above system provides a mixed-integer linear model for the fuel costs $f(\boldsymbol{\xi})$ that can be incorporated into our basic model. Although we here restrict ourselves to piecewise linear concave costs f the same model also works for general piecewise linear continuous costs.

Suppose the utility consumes $\jmath = 1, \ldots, \mathcal{J}$ different sorts of fuel which may be used for production only, for start-ups only or for both purposes simultaneously. For each of the sorts $\jmath$ let $\mathsf{I}_\jmath^p, \mathsf{I}_\jmath^s$ denote the subsets of thermal units that use $\jmath$ for production and start-ups, respectively. We allow some of these sets to be empty, for instance, if some fuel is used for production or start-ups exclusively. Moreover, there are variables $\boldsymbol{\xi}_\jmath$, $\jmath = 1, \ldots, \mathcal{J}$, for the total consumption of the fuel $\jmath$. This consumption computes as

$$\boldsymbol{\xi}_\jmath = \sum_{t=1}^{T} \sum_{i \in \mathsf{I}_\jmath^p} \widehat{C}_i\big(\boldsymbol{p}_i^t, \boldsymbol{u}_i^t\big) + \sum_{t=1}^{T} \sum_{i \in \mathsf{I}_\jmath^s} \widehat{S}_i^t(\boldsymbol{u}_i), \quad \jmath = 1, \ldots, \mathcal{J}. \tag{12}$$

Here $\widehat{C}$, $\widehat{S}$ are the fuel consumptions. They are computed by dividing the fuel costs C, S from the basic model by the constant fuel prices adopted there.

The extension of the basic model towards staggered fuel prices is established as follows. For each of the fuels we set up the cost model given by the system (11) and the right-hand side in (10). Each system (11) enters the constraints, and the sum of the expressions from (10) forms the objective of our extended model. The links between total consumption of each of the fuels and the basic model are established by the equations in (12) which enter the constraints. The model is completed by the constraints (2)–(6) already occurring in the basic model.

The transformation of the maximum terms behind $\widehat{S}$, cf. section 2, into linear expressions is accomplished in the same way as described in section 2.

Of course, staggered fuel prices are relevant for longer time periods only. Therefore we designed test problems with an optimization horizon of 6 months (with a 4-hourly time discretization). Of the 11 different fuels there are staggered prices for $\mathcal{J} = 3$ sorts each of them with $\mathcal{I} = 3$ price steps. After roughly an hour, the primal method produced feasible points that later turned out to be within 1% of the optimum. Verifying the quality of the solutions, however, is very expensive, and we observed computing times of up to another 40 hours.

4. Model extensions – dual solution techniques

4.1. Nonlinear start-up costs

In contrast to what is assumed in our basic model, start-up costs of coal fired thermal units essentially depend on the preceding down time of the block. This dependence follows an exponential saturation curve towards a finite constant. In practice,

after some finite down time, the start-up costs are constant (cold start). The following description reflects this dependence on state variables:

$$S_i^t(\boldsymbol{u}_i) = \max_{\tau=0,\ldots,\tau_i^c} c_i^\tau \left(\boldsymbol{u}_i^t - \sum_{\kappa=1}^{\tau} \boldsymbol{u}_i^{t-\kappa} \right),$$

where $c_i^0 = 0$ and c_i^τ for $\tau = 0, \ldots, \tau_i^c$ are fixed increasing cost coefficients, τ_i^c is the time the unit i needs to cool down, and $c_i^{\tau_i^c}$ is the cost for cold start. Choosing $c_i^0 = 0$ ensures that $S_i^t(\boldsymbol{u}_i)$ is non-negative. For $\tau > 0$ the second factor equals 1 if the unit is on-line at time t and has been off-line the τ preceding time periods. Table 1 shows the increase in dimension and constraint matrix fill, and table 2 displays our computational results when imposing a (rough) 3-step approximation of the curve. In conclusion, the primal approach is principally able to handle this situation, but the effort is unacceptable.

Fortunately, this situation improves considerably when employing the dual approach: with an equidistant time discretization and a finite time to enter the cold start phase the dynamic programming procedure for the thermal subproblems behind $d_j(\boldsymbol{\lambda}, \boldsymbol{\mu})$ can be adapted to handle the exponential saturation curve (in fact, to handle any nonlinear curve). The state space, so far determined by the on/off decisions $\boldsymbol{u}_i^t$ only, is extended by the consecutive time $\boldsymbol{o}_i^t$ unit i has been off-line at time step t. Since we have an equidistant discretization and reach the cold start phase in finite time, there are only finitely many different states for the $\boldsymbol{o}_i^t$.

Let M_i denote the number of time intervals for reaching the cold start phase and recall that τ_i is the minimum down time of unit i. Then there are $M_i + 1$ different states per unit and time step, namely $(\boldsymbol{u}_i^t, \boldsymbol{o}_i^t) \in \{(1,0), (0,1), (0,2), \ldots, (0, M_i)\}$, and the possible transitions are

$$\begin{array}{ll}
(1,0) \to (1,0), & (1,0) \to (0,1), \\
\big(0,t'\big) \to \big(0,t'+1\big) & \text{for } t' < \tau_i, \\
\big(0,t'\big) \to \big(0,t'+1\big), & \big(0,t'\big) \to (1,0) \quad \text{for } \tau_i \leqslant t' \leqslant M_i - 1, \\
(0,M_i) \to (0,M_i), & (0,M_i) \to (1,0).
\end{array}$$

For each of these finitely many state transitions the corresponding start-up cost is readily computed via the functional dependence in the saturation curve.

Table 6 shows our computational results for $M_i = 39$. Again, for the weekly computations the heuristic LH2 was employed and for the remaining computations LH1 (cf. subsection 2.3). Although there is an increased effort due to the state space extension in the dynamic programming, this is far more tolerable than the enormous increase with the primal approach.

4.2. Inclusion of uncertainties

Uncertainty is a major issue in power optimization. Among the main sources of uncertainty there are load profiles, generator outages, stream flows in water units,

Table 6
Nonlinear start-up costs, CPU-time in minutes on HP 9000 (770/J180) and upper bound of the duality gap of the dual method.

NOA 3.0 optimality tolerance: 10^{-4}	Optimization horizon		
	1 week	1 month	6 months
CPU-time (min)	0:18	3:04	63:04
Bound of gap (%)	0.28	0.98	0.73

and prices or market situations in general. Liberalization in the power industry has fostered the mathematical analysis of power systems under uncertainty. Although liberalization of power markets is out of the scope of the present unit commitment paper, the mathematical machinery from stochastic programming that we are going to utilize next has a considerable potential for addressing liberalization issues as well.

In the literature there is a growing number of contributions to power optimization under uncertainty with emphasis on modeling aspects and solution methods. For instance, the papers [14,25] address optimization models for hydroelectricity production and their solution by nested Benders decomposition techniques. Models for hydrothermal generation systems under uncertain electrical load and/or electricity prices are considered in [5,27,31–33], and variants of (augmented) Lagrangian decomposition methods are proposed for their solution.

In the present paper we will focus on the issue of planning a unit commitment schedule under uncertainty of power demand. We assume that the electrical load $\{D^t\colon t = 1, \ldots, T\}$ is a random variable. At the beginning of the optimization horizon we are facing incomplete information in that we only know the probability distribution of D and not its precise outcomes. Nevertheless we are forced to take decisions. Stochastic programming offers deterministic equivalents for optimizing decisions in situations like the one we are in. In fact, we are exposed to a multi-stage scheme of alternating decisions and observations. Assuming complete information on the load at time $t = 1$ we decide on all variables of the first interval, then we observe the outcome of D^2 and take the decisions for $t = 2$, afterwards we observe D^3, take the decisions for $t = 3$, and so on. Here, the issue of non-anticipativity is crucial: decisions must not depend on future realizations of the random components. A proper criterion for optimization in this context would be to minimize the sum of the direct costs caused by the decisions at $t = 1$ plus the expectation of the costs caused by all the future non-anticipative decisions. In this way, we end up with a linear mixed-integer multi-stage stochastic program. For a detailed introduction to multi-stage stochastic programs that adds mathematical rigour to the above sketch we refer to the textbook [1].

The multi-stage stochastic program for unit commitment sketched above is described in more detail in [7]. It embodies an operational model that is very demanding from the computational point of view. Nevertheless, it is possible to tackle the problem by a stochastic variant of the Lagrangian relaxation approach from subsection 2.3

(cf. [7,24]). The idea is to associate stochastic Lagrange multipliers to the stochastic load and reserve constraints. The full model then decomposes into single-unit multi-stage stochastic programs that are coordinated by the Lagrangian dual. Again, specialized algorithms (stochastic dynamic programming, descent algorithm) for the stochastic single-unit (thermal, storage) subproblems, concave nondifferentiable maximization of the (stochastic) dual, and heuristics for regaining the relaxed load and reserve constraints are crucial. For details of the algorithm and preliminary numerical experience for the weekly VEAG power generation model under uncertain load we refer to [23] and [24].

In what follows, we adopt a planning rather than an operational point of view. In some sense this will lead us to a two-stage approximation of the above multi-stage program. We assume that, in advance, we have to decide for the whole time horizon on those variables that reflect decisons which cannot be employed as short-term corrective actions. The vector of all these variables forms our first-stage decision, the remaining variables are in the second stage, i.e., they depend on the outcome of the random variable D. The multi-stage modeling approach then reduces to a two-stage one.

Starting up a coal fired block involves some time delay before the block becomes available for electricity generation. Therefore, switching decisions for these units have to be taken well in advance and cannot be employed as short-term corrective actions. This motivates us to put the $\boldsymbol{u}$-components belonging to the coal fired blocks into the first stage. Indeed, all the remaining decisions in the basic model from section 2, namely switching of gas turbines as well as operation of the on-line thermal and hydro units, involve only minor delay that is feasible for short-term corrective actions. The two-stage stochastic program then yields an implementable weekly plan of on/off decisons for the coal fired units. This plan minimizes the sum of the direct (start-up) costs plus the expected value of the costs that arise after having observed the load profile and optimized the second-stage corrective actions.

A crucial ingredient of a stochastic program is the probability distribution underlying the random data. Its extraction from statistical data is highly non-trivial and a field of active research, see, e.g., [8,26,34]. The availability of statistical data itself often may be a problem. Fortunately, this is not the case with load profiles in power industry where utilities maintain rich data collections. Behind the expectation entering the two- and multi-stage stochastic programs addressed above, there is a multivariate integral over an implicitly given integrand. Numerically, it is thus hopeless to operate with multivariate continuous probability distributions at this place. Therefore we assume that the random load profile D follows a discrete distribution with finite support, realizations D^ω, $\omega \in \Omega$, and probabilities π^ω, $\omega \in \Omega$. Following a usual convention, the realizations will be called scenarios.

To formalize our extension of the basic model from section 2 to the two-stage stochastic program mentioned above, we have to distinguish between coal and gas fired thermal units. Deviating from the notation in section 2 we therefore denote by $i = 1, \ldots, I$ the coal fired thermal units only, and we introduce $k = 1, \ldots, K$ for the gas turbines.

The variables $\boldsymbol{u}_i^t \in \{0,1\}$, $i = 1,\dots,I$, $t = 1,\dots,T$, now denote the first-stage decisions. The second-stage variables depend on the scenarios and thus carry an additional index ω: $\boldsymbol{u}_k^{t\omega} \in \{0,1\}$, $k = 1,\dots,K$, $t = 1,\dots,T$, $\omega \in \boldsymbol{\Omega}$, are the start-ups of the gas turbines and $\boldsymbol{p}_i^{t\omega}$, $\boldsymbol{p}_k^{t\omega}$, $\boldsymbol{s}_j^{t\omega}$, $\boldsymbol{w}_j^{t\omega}$, $i = 1,\dots,I$, $k = 1,\dots,K$, $j = 1,\dots,J$, $t = 1,\dots,T$, $\omega \in \boldsymbol{\Omega}$, are the output levels for the coal and gas fired thermal units, the hydro units in generation and in pumping modes, respectively. Finally, we have $\boldsymbol{l}_j^{t\omega}$ for the fills which are second-stage variables as well.

As with the basic model, these variables have to fulfill the constraints (2)–(5), and (6). In (3) the scenarios $D^{t\omega}$ enter at the right-hand side.

In the basic model we have affine expressions for the fuel costs of the thermal units in operation. With the above variables these expressions now read $c_i\boldsymbol{p}_i^{t\omega} + c_i^o$ for the coal fired and $c_k\boldsymbol{p}_k^{t\omega} + c_k^o$ for the gas fired units. Here, c_i, c_i^o, c_k, c_k^o are suitable constants. The objective function of our two-stage stochastic program is then given by

$$\sum_{t=1}^{T}\sum_{i=1}^{I} a_i \max\{\boldsymbol{u}_i^t - \boldsymbol{u}_i^{t-1}, 0\} + \boldsymbol{E}_\omega\left[\sum_{t=1}^{T}\sum_{k=1}^{K} a_k \max\{\boldsymbol{u}_k^{t\omega} - \boldsymbol{u}_k^{(t-1)\omega}, 0\}\right]$$
$$+ \boldsymbol{E}_\omega\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{I} \boldsymbol{u}_i^t\left(c_i\boldsymbol{p}_i^{t\omega} + c_i^o\right) + \sum_{k=1}^{K} \boldsymbol{u}_k^{t\omega}\left(c_k\boldsymbol{p}_k^{t\omega} + c_k^o\right)\right)\right].$$

The box constraints (2) allow to replace in the above expression $\boldsymbol{u}_i^t(c_i\boldsymbol{p}_i^{t\omega} + c_i^o)$ by $c_i\boldsymbol{p}_i^{t\omega} + c_i^o\boldsymbol{u}_i^t$, and accordingly for the gas turbines. Altogether we end up with a linear mixed-integer objective function.

The above model is elaborated in detail in [3]. To study its principal features we rewrite the model as

$$\min\left\{ c\boldsymbol{x} + \sum_{\nu=1}^{r} \pi^\nu q\boldsymbol{y}^\nu\colon\ A\boldsymbol{x} \leqslant b,\ \boldsymbol{x} \in X,\ T\boldsymbol{x} + W\boldsymbol{y}^\nu \leqslant h^\nu,\ \boldsymbol{y}^\nu \in Y,\right.$$
$$\left.\nu = 1,\dots,r \right\}. \tag{13}$$

Here $\boldsymbol{x}$, $\boldsymbol{y}$ refer to the first- and second-stage variables and X, Y denote restrictions requiring some or all of the variables to be binary. Accordingly, the data vectors and matrices are derived, with the mentioned remodeling of the maximum expressions in the objective. In particular, r denotes the cardinality of the support of the random variable D.

Problem (13) is a large-scale mixed-integer linear program whose constraint matrix obeys the block-angular structure depicted in figure 5.

Of course, this model is far too big to be tackled directly. Again decomposition will be helpful. To this end, we rewrite (13) by introducing the copies $\boldsymbol{x}^1,\dots,\boldsymbol{x}^r$ and adding the constraints $\boldsymbol{x}^1 = \cdots = \boldsymbol{x}^r$:

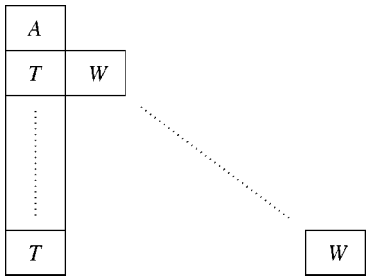

Figure 5. Constraints matrix structure of (13).

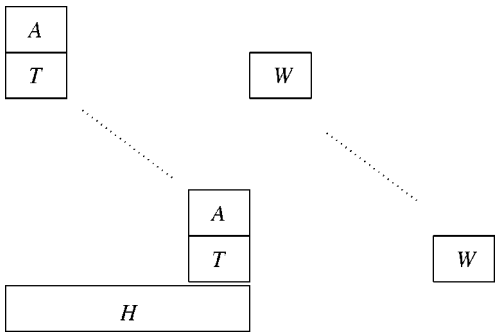

Figure 6. Constraints matrix of the scenario formulation (14).

$$\min\Bigg\{\sum_{\nu=1}^{r}\pi^{\nu}\big(c\boldsymbol{x}^{\nu}+q\boldsymbol{y}^{\nu}\big)\colon\ A\boldsymbol{x}^{\nu}\leqslant b,\ \boldsymbol{x}^{\nu}\in X,\ T\boldsymbol{x}^{\nu}+W\boldsymbol{y}^{\nu}\leqslant h^{\nu},\ \boldsymbol{y}^{\nu}\in Y,$$
$$\nu=1,\ldots,r,\ \boldsymbol{x}^{1}=\cdots=\boldsymbol{x}^{r}\Bigg\}. \tag{14}$$

The equations $\boldsymbol{x}^1 = \cdots = \boldsymbol{x}^r$ then explicitly represent the non-anticipativity that before was modeled implicitly by the scenario independence of the first-stage variables. For convenience we express the equations $\boldsymbol{x}^1 = \cdots = \boldsymbol{x}^r$ by the constraint $\sum_{\nu=1}^{r} H^{\nu}\boldsymbol{x}^{\nu} = 0$ where $H = (H^1, \ldots, H^r)$ is a suitable matrix. This leads to the block structure depicted in figure 6.

Figure 6 suggests a Lagrangian relaxation of the non-anticipativity constraint $\sum_{\nu=1}^{r} H^{\nu}\boldsymbol{x}^{\nu} = 0$ since the latter is the only condition in (14) that interconnects single-scenario subproblems. In this way, we obtain the Lagrangian dual

$$\max_{\lambda} d(\boldsymbol{\lambda}), \tag{15}$$

where

$$d(\boldsymbol{\lambda}) := \sum_{\nu=1}^{r} d_{\nu}(\boldsymbol{\lambda}) \tag{16}$$

and

$$d_\nu(\boldsymbol{\lambda}) := \min\bigl\{\pi^\nu\bigl(c\boldsymbol{x}^\nu + q\boldsymbol{y}^\nu\bigr) + \boldsymbol{\lambda}\bigl(H^\nu\boldsymbol{x}^\nu\bigr)\colon\ A\boldsymbol{x}^\nu \leqslant b,\ \boldsymbol{x}^\nu \in X,\\ T\boldsymbol{x}^\nu + W\boldsymbol{y}^\nu \leqslant h^\nu,\ \boldsymbol{y}^\nu \in Y\bigr\}$$

for $\nu = 1, \ldots, r$.

In principle, we are now in the same situation as in subsection 2.3. The Lagrangian dual is a non-smooth concave maximization problem for whose solution we apply the proximal bundle method from [16,17]. Function values and subgradients of d are given by (16) where we can exploit separability into single-scenario subproblems. Note that the latter are very close to our basic model from section 2. The only difference is the term $\boldsymbol{\lambda}(H^\nu\boldsymbol{x}^\nu)$ in the objective. Experience in solving the basic model (by either primal or dual methods) hence may be exploited directly when solving the single-scenario subproblems.

In the end, the optimal value of the Lagrangian dual (15) provides us with a lower bound to the optimal value of our stochastic program (13). The presence of integer variables and hence the missing convexity is the reason for these values to be different in general. Therefore, the solutions to the single-scenario subproblems for the optimal $\boldsymbol{\lambda}$ in general violate the relaxed non-anticipativity constraint $\sum_{\nu=1}^{r} H^\nu\boldsymbol{x}^\nu = 0$, and a Lagrangian heuristic for regaining non-anticipativity has to be set up. Compared with the situation in subsection 2.3, however, now the relaxed constraint is a pretty simple identity of the $\boldsymbol{x}$-components of the single-scenario solutions. This immediately gives rise to heuristics: consider the $\boldsymbol{x}$-components of the single-scenario solutions as proposals and decide for one of them by either averaging (and rounding to the next integer) or by the frequency of occurrence or some other criterion.

Altogether, the above dual approach to the stochastic program (13) leads to a solution for which the lower bound from the Lagrangian dual provides a quality certificate (gap). A major algorithmic advantage is in reducing the solution of the model extension (13) to instances that are very close to our basic model.

To reduce the gap even further, a branch-and-bound scheme in the spirit of global optimization is placed on top of the above procedure, cf. [2]. The link to global optimization becomes evident when rewriting (13) as

$$\min\bigl\{c\boldsymbol{x} + Q(\boldsymbol{x})\colon\ A\boldsymbol{x} \leqslant b,\ \boldsymbol{x} \in X\bigr\}, \tag{17}$$

where

$$Q(\boldsymbol{x}) = \boldsymbol{E}_\omega\phi\bigl(h^\omega - T\boldsymbol{x}\bigr)$$

and

$$\phi(s) = \min\{q\boldsymbol{y}\colon\ W\boldsymbol{y} \leqslant s,\ \boldsymbol{y} \in Y\}.$$

Indeed, the above function $Q(\boldsymbol{x})$ is lower semicontinuous in general [28] such that it makes sense to tackle (17) by branch-and-bound. Branching is performed by subdividing the set $\{\boldsymbol{x}\colon\ A\boldsymbol{x} \leqslant b,\ \boldsymbol{x} \in X\}$, upper and lower bounds are obtained as above, and the critical property to be established is non-anticipativity.

To outline the algorithm let $\mathcal{P}$ denote the list of current problems and $z_{LD} = z_{LD}(P)$ be a lower bound associated with problem $P \in \mathcal{P}$. Then we proceed as follows:

Step 1 *Initialization:* Set $\overline{z} = +\infty$ and let $\mathcal{P}$ consist of problem (14).

Step 2 *Termination:* If $\mathcal{P} = \emptyset$ then the solution $\widehat{\boldsymbol{x}}$ that yielded $\overline{z} = c\widehat{\boldsymbol{x}} + Q(\widehat{\boldsymbol{x}})$ is optimal.

Step 3 *Node selection:* Select and delete a problem P from $\mathcal{P}$ and solve its Lagrangian dual. If the optimal value $z_{LD}(P)$ hereof equals $+\infty$ (infeasibility of a subproblem) then go to step 2.

Step 4 *Bounding:* If $z_{LD}(P) \geqslant \overline{z}$ go to step 2 (this step can be carried out as soon as the value of the Lagrangian dual rises above $\overline{z}$).

(i) The scenario solutions $\boldsymbol{x}^\nu$, $\nu = 1, \ldots, r$, are identical: if $c\boldsymbol{x}^\nu + Q(\boldsymbol{x}^\nu) < \overline{z}$ then let $\overline{z} = c\boldsymbol{x}^\nu + Q(\boldsymbol{x}^\nu)$ and delete from $\mathcal{P}$ all problems P' with $z_{LD}(P') \geqslant \overline{z}$. Go to step 2.

(ii) The scenario solutions $\boldsymbol{x}^\nu$, $\nu = 1, \ldots, r$, differ: compute the average $\overline{\boldsymbol{x}} = \sum_{\nu=1}^{r} \pi^\nu \boldsymbol{x}^\nu$ and round it by some heuristic to obtain $\overline{\boldsymbol{x}}^R$. If $c\overline{\boldsymbol{x}}^R + Q(\overline{\boldsymbol{x}}^R) < \overline{z}$ then let $\overline{z} = c\overline{\boldsymbol{x}}^R + Q(\overline{\boldsymbol{x}}^R)$ and delete from $\mathcal{P}$ all problems P' with $z_{LD}(P') \geqslant \overline{z}$. Go to step 5.

Step 5 *Branching:* Select a component $\boldsymbol{x}_{(m)}$ of $\boldsymbol{x}$ and add two new problems to $\mathcal{P}$ obtained from P by setting $\boldsymbol{x}_{(m)} = 0$ and $\boldsymbol{x}_{(m)} = 1$.

The model extension (13) is solved quite satisfactorily by the above methods. Details are reported in [3]. We ran both instances with binary on/off-variables for thermal units and with integer variables in case of identical blocks (cf. subsection 2.2).

Table 7 displays problem dimensions for the deterministic equivalent (13). All problem instances are based on an optimization horizon of $T = 168$ hourly intervals. The power system comprises $I = 17$ coal fired blocks, $K = 8$ gas turbines, and $J = 7$ pumped storage plants. The columns correspond to the number of scenarios, of constraints, of variables in total, of integer variables, and to the dimension of the multiplier λ.

Table 8 shows the quality certificates (gaps) obtained after 10 minutes of CPU-time at a Digital Alpha Personal Workstation with 500 MHz processor. There are two different instances of the uncertain electrical load, one caused by generator failure, the other by inaccurate load forecast. The columns "with NOA 3.0" correspond to the full algorithm outlined above, the columns "without NOA 3.0" concern the algorithmic shortcut where instead of maximizing in the Lagrangian dual (15) we just computed $d(\lambda)$ for $\lambda = 0$. On the one hand, this allowed far more iterations of the branch-and-bound scheme, on the other hand, lower bounds became so inferior that the final gaps could not compete with those from the full algorithm.

Table 7
Problem sizes.

Model	Scen.	Constr.	Var.	Int.	Mult.
Binary	4	47159	47327	7560	11424
	10	113639	109775	14616	28560
	16	180119	172223	21672	45696
Integer	4	32049	37257	5880	4704
	10	78369	89625	12936	11760
	16	124689	141993	19992	18816

Table 8
Quality certificates.

Model formulation	Scen.	Gap			
		Generator failure instances		Inaccurate load forecast instances	
		without NOA 3.0	with NOA 3.0	without NOA 3.0	with NOA 3.0
Binary	4	3.2%	0.8%	1.4%	0.5%
	10	11.1%	0.8%	8.3%	3.3%
	16	9.0%	2.8%	10.1%	2.7%
Integer	4	3.2%	0.1%	4.1%	0.1%
	10	1.7%	0.2%	3.1%	0.3%
	16	2.2%	0.3%	2.5%	0.4%

Acknowledgement

This work would never have been possible without the outstanding cooperation with VEAG Vereinigte Energiewerke AG Berlin over many years. In particular we would like to thank G. Schwarzbach, J. Thomas, and J. Krause. Further thanks are due to C.C. Carøe (Unibank Copenhagen, formerly with the University of Copenhagen) for the collaboration in unit commitment under uncertainty and to K.C. Kiwiel (Polish Academy of Sciences, Warsaw) for the permission to use the NOA 3.0 package.

References

[1] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming* (Springer, New York, 1997).
[2] C.C. Carøe and R. Schultz, Dual decomposition in stochastic integer programming, Operations Research Letters 24 (1999) 37–45.
[3] C.C. Carøe and R. Schultz, A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal system, Preprint SC 98-11, Konrad-Zuse-Zentrum für Informationstechnik Berlin (1998). Downloadable as SC 98-11 from `http://www.zib.de/bib/pub/pw/index.en.html`.
[4] *Using the CPLEX Callable Library*, CPLEX Optimization, Inc. 1994.
[5] P. Carpentier, G. Cohen, J.-C. Culioli and A. Renaud, Stochastic optimization of unit commitment, a new decomposition framework, IEEE Transactions on Power Systems 11 (1996) 1067–1073.

[6] D. Dentcheva, R. Gollmer, A. Möller, W. Römisch and R. Schultz, Solving the unit commitment problem in power generation by primal and dual methods, in: *Progress in Industrial Mathematics at ECMI 96*, eds. M. Brøns, M.P. Bendsøe and M.P. Sørensen (Teubner, Stuttgart, 1997) pp. 332–339.
[7] D. Dentcheva and W. Römisch, Optimal power generation under uncertainty via stochastic programming, in: *Stochastic Programming Methods and Technical Applications*, eds. K. Marti and P. Kall, Lecture Notes in Economics and Mathematical Systems, Vol. 458 (Springer, Berlin, 1998) pp. 22–56.
[8] J. Dupačová, G. Consigli and S.W. Wallace, Scenarios for multistage stochastic programs, Working Paper, Charles University Prague (1998). Submitted to Annals of Operations Research.
[9] S. Feltenmark and K.C. Kiwiel, Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems, SIAM Journal on Optimization (to appear).
[10] S. Feltenmark, K.C. Kiwiel and P.O. Lindberg, Solving unit commitment problems in power production planning, in: *Operations Research Proceedings 1996*, eds. U. Zimmermann et al. (Springer, Berlin, 1997) pp. 236–241.
[11] M. Fischer, Gestaffelte Brennstoffpreise in der Kraftwerkseinsatzoptimierung, Diplomarbeit, Technische Universität Berlin (1997).
[12] R. Gollmer, A. Möller, W. Römisch, R. Schultz, G. Schwarzbach and J. Thomas, Optimale Blockauswahl bei der Kraftwerkseinsatzplanung der VEAG, in: *Optimierung in der Energieversorgung II* (VDI-Berichte 1352, 1997) pp. 71–85.
[13] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II* (Springer, Berlin, 1993).
[14] J. Jacobs, G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus and J. Stedinger, SOCRATES: A system for scheduling hydroelectric generation under uncertainty, Annals of Operations Research 59 (1995) 99–133.
[15] N. Jiménez Redondo and A.J. Conejo, Short-term hydro-thermal coordination by Lagrangian relaxation, Solution of the dual problem, IEEE Transactions on Power Systems 14 (1999) 89–95.
[16] K.C. Kiwiel, Proximity control in bundle methods for convex nondifferentiable minimization, Mathematical Programming 46 (1990) 105–122.
[17] K.C. Kiwiel, User's Guide for NOA 2.0/3.0: A Fortran package for convex nondifferentiable optimization, Polish Academy of Science, System Research Institute, Warsaw (1993/1994).
[18] C. Lemaréchal and A. Renaud, Dual equivalent convex and nonconvex problems, Research Report, INRIA, Rocquencourt (1996).
[19] C. Lemaréchal, C. Sagastizábal, F. Pellegrino and A. Renaud, Bundle methods applied to the unit commitment problem, in: *System modelling and optimization*, eds. J. Doležal and J. Fiedler (Chapman & Hall, London, 1996) pp. 395–402.
[20] P.B. Luh, R.N. Tomastik and D. Zhang, An algorithm for solving the dual problem of hydrothermal scheduling, IEEE Transactions on Power Systems 13 (1998) 593–600.
[21] A. Möller and W. Römisch, A dual method for the unit commitment problem, Preprint Nr. 95-1, Humboldt-Universität Berlin, Institut für Mathematik (1995). Downloadable from `http://taylor.mathematik.hu-berlin.de/publ/pre/1995/p-95-1.ps`.
[22] M.P. Nowak, A fast descent method for the hydro storage subproblem in power generation, Working Paper WP-96-109, IIASA, Laxenburg (Austria) (1996). Downloadable from `http://www.iiasa.ac.at/Publications/Documents/WP-96-109.ps`.
[23] M.P. Nowak, Stochastic Lagrangian relaxation applied to power scheduling of a hydro-thermal system under uncertainty, Forthcoming dissertation, Humboldt-Universität Berlin, Institut für Mathematik (1999).
[24] M.P. Nowak and W. Römisch, Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty, Preprint Nr. 98-24, Humboldt-Universität Berlin, Institut für Mathematik (1998). Submitted to Annals of Operations Research. Downloadable from `http://www.mathematik.hu-berlin.de/publ/pre/1998/M-98-24`.

[25] M.V.F. Pereira and L.M.V.G. Pinto, Multi-stage stochastic optimization applied to energy planning, Mathematical Programming 52 (1991) 359–375.
[26] G.Ch. Pflug and A. Świetanowski, Optimal scenario tree generation for multiperiod financial optimization, Technical Report, AURORA TR 1998-22, Universität Wien (1998). Downloadable from ftp://ftp.par.univie.ac.at/projects/aurora/reports/auroratr1998-22.ps.gz.
[27] W. Römisch and R. Schultz, Decomposition of a multi-stage stochastic program for power dispatch, ZAMM – Zeitschrift für Angewandte Mathematik und Mechanik 76(3) (1996) 29–32.
[28] R. Schultz, On structure and stability in stochastic programs with random technology matrix and complete integer recourse, Mathematical Programming 70 (1995) 73–89.
[29] G.B. Sheble and G.N. Fahd, Unit commitment literature synopsis, IEEE Transactions on Power Systems 9 (1994) 128–135.
[30] S. Takriti, The unit commitment problem, in: *Operations Research in Industry*, eds. T.A. Ciriani, S. Gliozzi and E.L. Johnson (Macmillan Press, 1999) (to appear).
[31] S. Takriti, J.R. Birge and E. Long, A stochastic model for the unit commitment problem, EEE Transactions on Power Systems 11 (1996) 1497–1508.
[32] S. Takriti, B. Krasenbrink and L.S.-Y. Wu, Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem, Operations Research (to appear).
[33] S. Takriti, C. Supatgiat and L.S.-Y. Wu, Coordinating fuel inventory and electric power generation under uncertainty, IBM Research Report RC 21152, Yorktown Heights, New York (1998). Downloadable from http://domino.watson.ibm.com/library/CyberDig.nsf/Papers?SearchView&Query=RC21152.
[34] I. Wegner, Erzeugung von Szenariobäumen für die Kraftwerks-Einsatzoptimierung, Diplomarbeit, Humboldt-Universität Berlin, Institut für Mathematik (1999).
[35] A.J. Wood and B.F. Wollenberg, *Power Generation, Operation and Control*, 2nd ed. (Wiley, New York, 1996).
[36] F. Zhuang and F.D. Galiana, Towards a more rigorous and practical unit commitment by Lagrangian relaxation, IEEE Transactions on Power Systems 3 (1988) 763–773.