Optimization and Engineering, 3, 355–378, 2002

A Two-Stage Planning Model for Power Scheduling in a Hydro-Thermal System Under Uncertainty

ROBERT NÜRNBERG, WERNER RÖMISCH*
Institute of Mathematics, Humboldt University Berlin, D-10099 Berlin, Germany

Received August 7, 2000; Revised October 5, 2001

Abstract. A two-stage stochastic programming model for the short- or mid-term cost-optimal electric power production planning is developed. We consider the power generation in a hydro-thermal generation system under uncertainty in demand (or load) and prices for fuel and delivery contracts. The model involves a large number of mixed-integer (stochastic) decision variables and constraints linking time periods and operating power units. A stochastic Lagrangian relaxation scheme is designed by assigning (stochastic) multipliers to all constraints that couple power units. It is assumed that the stochastic load and price processes are given (or approximated) by a finite number of realizations (scenarios). Solving the dual by a bundle subgradient method leads to a successive decomposition into stochastic single unit subproblems. The stochastic thermal and hydro subproblems are solved by a stochastic dynamic programming technique and by a specific descent algorithm, respectively. A Lagrangian heuristics that provides approximate solutions for the primal problem is developed. Numerical results are presented for realistic data from a German power utility and for numbers of scenarios ranging from 5 to 100 and a time horizon of 168 hours. The sizes of the corresponding optimization problems go up to 400.000 binary and 650.000 continuous variables, and more than 1.300.000 constraints.

Keywords: stochastic programming, Lagrangian relaxation, unit commitment

1. Introduction

The optimal management of electric power generation systems requires operational and planning models. Operational optimization models provide schedules for the actual operation of power units and of electricity trading while planning models are needed to support mid- or long-term decisions (e.g. by estimating the yearly sales trend or fuel consumption). Both types of optimization models are often large-scale, mixed-integer and stochastic. The latter aspect mostly concerns uncertainties of electrical load forecasts, of generator failures, of flows to hydro reservoirs or plants, and of fuel or electricity prices.

In the present paper we develop a stochastic *planning model* for a thermal or thermally dominated generation system that allows the computation of realistic production costs for a short- or mid-term time horizon. Realistic production schedules of a power system typically consist of a composition of piecewise (optimal) schedules for parts of the whole time interval. Such a composition of schedule pieces is due to system re-optimizations after data (e.g. load) changes or further unforeseeable events. Moreover, the power system has to be run such that it is always able to satisfy all system constraints. This behaviour is

*Corresponding author.

modelled in Section 2 by introducing a *two-stage stochastic programming model*, where the first-stage decision corresponds to the above composite schedule and the stochastic second-stage decision reflects possible scenarios of the optimal system schedule that refer to future recourse actions of each unit in response to the environment created by the first-stage decision and by the data scenarios. Earlier attempts of developing such two-stage models appear in Carøe and Schultz (1998) and Dentcheva and Römisch (1998). While the model in Dentcheva and Römisch (1998) may be regarded as a starting point of the present paper, the modelling idea of Carøe and Schultz (1998) is to consider the binary variables as first-stage decisions and the unit outputs as recourse or second-stage actions. The modelling novelty of the present paper consists in the introduction of *compatibility constraints* for first- and second-stage actions.

More specifically, we consider a model for a generation system comprising thermal units and pumped hydro storage plants under uncertain electrical load and prices for fuel and electricity. The relevant mathematical optimization model contains a large number of binary and continuous variables, constraints and stochasticity appearing in right-hand sides and cost coefficients. The time horizon ranges from one week up to one year. Computational results are given for a time period of one week as it is needed for the efficient weekly production planning of hydro-thermal systems involving weekly load and pumping cycles. Planning results for longer time periods could also be obtained approximately by solving the model for several selected characteristic weeks. Our model and solution techniques are validated on the system of the German utility VEAG Vereinigte Energiewerke AG. Its total capacity is about 13,000 megawatts (MW) including a hydro capacity of 1,700 MW; the system peak loads are about 8,600 MW. Test runs were performed for a typical configuration of the VEAG system with 25 thermal units and 7 pumped hydro plants.

Since our stochastic programming model contains mixed-integer decisions in both stages and is large-scale, new questions on the design of solution algorithms are raised. Nowadays, solution methods are well developed for linear two-stage stochastic programs without integrality constraints (see Birge, 1997; Ruszczyński, 1997). Most of them are based on discrete approximations of the stochastic data process. Recently, some algorithmic progress has also been achieved in mixed-integer stochastic programming models and applications to operations research and power optimization. The following algorithmic approaches to such models appear in the literature: (a) Scenario decomposition by splitting methods combined with suitable heuristics (Takriti and Birge, 2000), (b) scenario decomposition combined with branch and bound (Carøe and Schultz, 1998, 1999), (c) stochastic (augmented) Lagrangian relaxation of coupling constraints (Carpentier et al., 1996; Dentcheva and Römisch, 1998; Gröwe-Kuska et al., 2002; Nowak, 2000; Takriti et al., 2000). The approaches in (a) and (b) are based on a successive decomposition of the stochastic program into finitely many deterministic (or scenario) programs that may be solved by available conventional techniques. The approach of (c) hinges on a successive decomposition into finitely many smaller stochastic subproblems for which (efficient) solution techniques must be developed eventually. Due to the nonconvexity of the underlying stochastic program, the successive decompositions in (a)–(c) have to be combined with certain global optimization techniques (branch-and-bound, heuristics, etc.).

Recently, a number of new applications of integer stochastic programming models appeared in different areas. We mention here the multi-period investment model for capacity expansion in an uncertain environment (Ahmed et al., 2002), a two-stage simultaneous power production and day-ahead trading at a power exchange under price and load uncertainty (Nowak et al., 2002) and a capacity planning model for a semiconductor manufacturing process under demand uncertainty (Swaminathan, 2000). The numerical solution techniques proposed in these papers are based on LP relaxation and branch-and-bound, on the method (b) above and on classical non-probabilistic Lagrangian relaxation, respectively.

The algorithmic approach followed in the present paper consists in a stochastic version of the classical Lagrangian relaxation idea (Lemaréchal, 1992), which is very popular in power optimization (cf. Bertsekas et al., 1983; Feltenmark and Kiwiel, 2000; Gollmer et al., 1999; Lemaréchal and Renaud, 2001; Sheble and Fahd, 1994; Zhuang and Galiana, 1988). Since the corresponding coupling constraints contain random variables, stochastic multipliers are needed for the dualization, and the dual problem represents a nondifferentiable stochastic program. Subsequently, the approach is based on the same, but stochastic, ingredients as in the classical case: a solver for the nondifferentiable dual, subproblem solvers, and a Lagrangian heuristics. It turns out that, due to the availability of a state-of-the-art bundle method for solving the dual, efficient stochastic subproblem solvers based on a specific descent algorithm and stochastic dynamic programming, respectively, and a specific Lagrangian heuristics for determining a nearly optimal primal solution, this *stochastic Lagrangian relaxation* algorithm becomes efficient.

The paper is organized as follows. In Section 2 a detailed description of the hydro-thermal generation system is given and the stochastic programming model is developed. Section 3 describes the stochastic Lagrangian relaxation approach together with its components. Finally, numerical results for the stochastic Lagrangian relaxation based algorithm are reported in Section 4 for realistic data of the VEAG system.

2. Model

We consider a power generation system comprising (coal-fired and gas-burning) thermal units, pumped hydro storage plants and delivery contracts. We describe a model for its mid-term cost-optimal power production planning under uncertainty on the electrical load and on the electricity and fuel prices. Let T denote the number of time intervals obtained from a uniform discretization of the operation horizon. Let I and J denote the number of thermal and pumped hydro storage units in the system, respectively. Delivery contracts are regarded as particular thermal units.

The decision variables in the model comprise the outputs of all units, i.e., the electric power generated or consumed by each unit of the system. They are denoted by u_{it}, p_{it}, $i = 1, \ldots, I$, and s_{jt}, w_{jt}, $j = 1, \ldots, J$, $t = 1, \ldots, T$, where $u_{it} \in \{0, 1\}$ is the on/off decision and p_{it} is the production level of the thermal unit i during the time period t. Thus, $u_{it} = 0$ and $u_{it} = 1$ mean that the unit i is off-line and on-line during period t, respectively. The generation and pumping levels of the pumped hydro storage plant j during period t are specified by s_{jt} and w_{jt}. Further, we denote the storage level (or volume) in the upper reservoir of plant j at the end of the interval t by ℓ_{jt}.

All variables mentioned above have finite upper and lower bounds representing unit limits and reservoir capacities of the generation system:

$$p_{it}^{\min} u_{it} \leq p_{it} \leq p_{it}^{\max} u_{it}, \quad u_{it} \in \{0, 1\}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, I, \tag{1a}$$

$$0 \leq s_{jt} \leq s_{jt}^{\max}, \quad 0 \leq w_{jt} \leq w_{jt}^{\max}, \quad 0 \leq \ell_{jt} \leq \ell_{jt}^{\max}, \qquad t = 1, \ldots, T, \ j = 1, \ldots, J. \tag{1b}$$

The constants $p_{it}^{\min}$, $p_{it}^{\max}$, $s_{jt}^{\max}$, $w_{jt}^{\max}$ and $\ell_{jt}^{\max}$ denote the minimal and maximal outputs of the units and the maximal storage levels in the upper reservoirs, respectively. The dynamics of the storage level, which is measured in electrical energy, is modelled by the equations:

$$\begin{aligned} &\ell_{jt} = \ell_{j,t-1} - s_{jt} + \eta_j w_{jt}, \quad t = 1, \ldots, T, \\ &\ell_{j0} = \ell_j^{in}, \quad \ell_{jT} = \ell_j^{end}, \quad j = 1, \ldots, J. \end{aligned} \tag{2}$$

Here, ℓ_j^{in} and ℓ_j^{end} denote the initial and final levels in the upper reservoir, respectively, and η_j is the cycle (or pumping) efficiency of plant j. The cycle efficiency is defined as the quotient of the generated and the pumped energy that correspond to the same amount of water. Together with the upper and lower bounds for ℓ_{jt} the Eq. (2) mean that certain reservoir constraints have to be maintained for all storage plants during the whole time horizon. Moreover, they show that there occur no in- or outflows in the upper reservoirs, and, hence, that the storage plants of the system operate with a constant amount of water. Further single-unit constraints are minimum up- and down-times and possible must-on/off constraints for each thermal unit. Minimum up- and down-time constraints are imposed to prevent thermal stress and high maintenance costs due to excessive unit cycling. Denoting the minimum up- and down-times of unit i by σ_i and τ_i, respectively, the corresponding constraints are described by the inequalities:

$$u_{it} - u_{i,t-1} \leq u_{i\sigma}, \quad \sigma = t+1, \ldots, \min\{t + \sigma_i - 1, T\}, \tag{3a}$$

$$u_i, t_{-1} - u_{it} \leq 1 - u_{i\tau}, \quad \tau = t+1, \ldots, \min\{t + \tau_i - 1, T\}, \qquad t = 1, \ldots, T-1, \ i = 1, \ldots, I. \tag{3b}$$

The next constraints are coupling across the units: the load and reserve constraints. The first constraints are essential for the operation of the power system and express that the sum of the output powers is greater than or equal to the load demand in each time period. Denoting by d_t the electrical load during period t, the load constraints are described by the inequalities:

$$\sum_{i=1}^{I} p_{it} + \sum_{j=1}^{J} (s_{jt} - w_{jt}) \geq d_t, \quad t = 1, \ldots, T. \tag{4}$$

In order to compensate unexpected events (e.g. sudden load increases or decreases, outages of units) within a specified short time period, a spinning reserve describing the total amount of generation available from all thermal units synchronized on the system minus the present load is prescribed. The corresponding constraints are given by the following inequalities:

$$\sum_{i=1}^{I} \left(p_{it}^{\max} u_{it} - p_{it}\right) \geq r_t, \quad t = 1, \ldots, T, \tag{5}$$

where $r_t > 0$ is the spinning reserve in period t.

The objective function is given by the total costs of operating all the units. Since the operating costs of hydro plants are usually negligible, the total system costs are given by the sum of the startup and operating costs of all thermal units over the whole scheduling horizon, i.e.:

$$\sum_{i=1}^{I} \sum_{t=1}^{T} [C_{it}(p_{it}, u_{it}) + S_{it}(u_i)], \tag{6}$$

where C_{it} are the costs for the operation of the thermal unit i during period t and S_{it} are the startup costs for getting the unit on-line in this period. We assume that each C_{it} is piecewise linear convex, strictly monotonically increasing and of the form

$$C_{it}(p, u) = \max_{l=1,\ldots,\bar{l}} \{\alpha_{ilt} p + \beta_{ilt} u\}, \tag{7}$$

where α_{ilt} and β_{ilt} are fixed cost coefficients. The startup costs of unit i depend on its downtime. They can vary from a maximum cold-start value to a smaller value when the unit is still relatively close to the operating temperature. Here we will neglect this dependence and assume constant costs γ_{it} for starting up unit i at time t. This simplification could be removed at the expense of higher model complexity. In particular, memory requirements for dynamic structures, which will be introduced in Section 3.2, would grow considerably. This would lead to a severe limitation of the performance of the developed algorithm. Hence, the description of the startup costs is given by

$$S_{it}(u_i) := \gamma_{it} \max\{u_{it} - u_{i,t-1}, 0\}, \tag{8}$$

where $u_i := (u_{it})_{t=0}^{T}$ and $u_{i0} \in \{0, 1\}$ is a given initial value.

Altogether minimizing the objective function (6) subject to the constraints (1)–(5) leads to a cost-optimal schedule for all units of the power system during the specified time horizon. It is worth mentioning that a cost-optimal schedule has the following two interesting properties, both of which are a consequence of the strict monotonicity of the cost functions. If a schedule (u, p, s, w) is optimal, then the load constraints (4) are typically satisfied with equality and we have $s_{jt} w_{jt} = 0$ for all $j = 1, \ldots, J,\ t = 1, \ldots, T$, i.e., generation and pumping do not occur simultaneously (cf. Gröwe et al., 1995).

The minimization problem (1)–(6) represents a mixed-integer program with linear constraints and IT binary and $(I + 2J)T$ continuous decision variables. For a configuration of the VEAG system with $I = 25$, $J = 7$ and $T = 168$ (i.e., 7 days with hourly discretization), this amounts to 4200 binary and 6552 continuous variables. Figure 1 shows a typical load

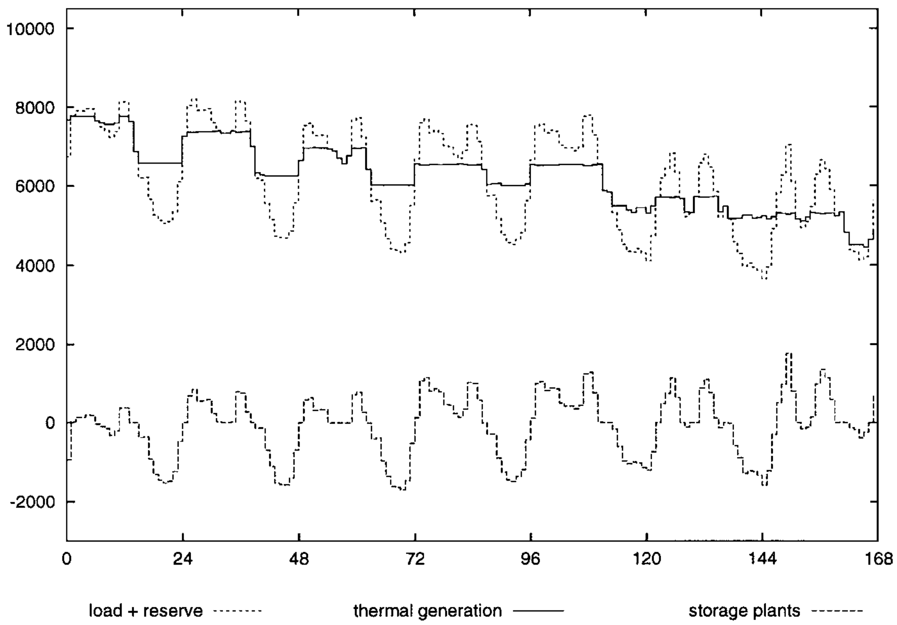

Figure 1. Load curve and hydro-thermal schedule.

curve of a peak load week and a corresponding cost-optimal hydro-thermal schedule. The load curve in the figure exhibits two overlapping cycles: a daily and a weekly cycle. Pumped hydro plants are designed to exploit these two cycles by transporting electrical energy in time. They help to save fuel costs through pumping during off-peak periods (e.g. nights and weekends) in order to refill the reservoir on the one hand, and serving with hydro-energy during peak-load periods on the other hand. The hydro schedule in figure 1 reflects this typical operation of pumped hydro storage plants. The remainder of the demand, i.e., the difference between the original system load and the hydro schedule, shows a more uniform structure than the original load. This portion of the load is covered by the total output of the thermal units.

The model elaborated so far covers the case where we are faced with deterministic data and thus with deterministic decision variables only. In power production planning this approach soon becomes futile when one considers time periods lying far in the future. In order to derive solutions that hedge against uncertainty it is necessary to incorporate the randomness of the data into the model. So far this has mainly been done for operational models (cf. Dentcheva and Römisch (1998) and Gröwe-Kuska et al. (2002) and references therein).

In electric utilities schedulers forecast the electrical load for the required time span. For this purpose they make use of historical data, meteorological parameters, experience and statistical methods. Since we regard future planning periods (e.g. next week or year), we assume that the quality of available information on the load uncertainty does not depend on time, i.e., the uncertainty does not increase with the length of the planning horizon.

Furthermore, the load and the prices are stochastic right from the beginning of the considered time period. The stochastic behaviour of the load $\boldsymbol{d}_t$, the spinning reserve $\boldsymbol{r}_t$ and the price for fuel and electricity—characterized by its coefficients $\boldsymbol{a}_t$, $\boldsymbol{b}_t$ and $\boldsymbol{c}_t$—is represented by a discrete-time stochastic process

$$\{\boldsymbol{\xi}_t := (\boldsymbol{d}_t, \boldsymbol{r}_t, \boldsymbol{a}_t, \boldsymbol{b}_t, \boldsymbol{c}_t)\}_{t=1}^{T}$$

on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. By $\boldsymbol{\xi}^{\omega}$ we denote the realizations of the stochastic process for all scenarios $\omega \in \Omega$. For any stochastic variable $\mathbf{x}$ we will use $\mathbf{x}^{\omega}$ accordingly, i.e., in order to denote its realization under scenario ω. Please note that throughout the paper we will use bold characters to emphasize stochastic elements.

Now, the decision process consists of two stages where the first-stage decisions correspond to the here-and-now schedules for all power generation units. The second-stage decisions, on the other hand, correspond to future compensation or recourse actions of each unit in each time period. The latter naturally depend on the environment created by the chosen first-stage decisions and the load and price realization in that specific time period. Hence, the aim of such a two-stage planning model can be formulated as follows. Find an optimal schedule for the whole power system and planning horizon such that the uncertain demand can be compensated by the system, all system constraints are satisfied and the sum of the total generation costs and the expected recourse costs is minimal.

In order to give a mathematical formulation of the two-stage model, let (u, p, s, w) denote the first-stage scheduling decisions as before. Furthermore, let $(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w})$ be the stochastic second-stage decisions having the components $\mathbf{u}_{it}, \mathbf{p}_{it}, \mathbf{s}_{it}, \mathbf{w}_{jt}, i = 1, \ldots, I, j = 1, \ldots, J, t = 1, \ldots, T$, which correspond to the recourse actions of each unit at time period t. In addition to the (non-stochastic) constraints for (u, p, s, w), i.e., the capacity limits (1), the storage dynamics (2) and the minimum up- and down-times (3), we have to require that the recourse actions also satisfy the system constraints. These are the operating ranges of all units, the minimum up/down-time requirements for the thermal units and the reservoir capacity bounds:

$$p_{it}^{\min}\mathbf{u}_{it} \le \mathbf{p}_{it} \le p_{it}^{\max}\mathbf{u}_{it}, \quad \mathbf{u}_{it} \in \{0, 1\}, \tag{9a}$$

$$\begin{aligned} &\mathbf{u}_{it} - \mathbf{u}_{i,t-1} \le \mathbf{u}_{i\sigma}, \quad \sigma = t+1, \ldots, \min\{t + \sigma_i - 1, T\}, \\ &\mathbf{u}_{i,t-1} - \mathbf{u}_{it} \le 1 - \mathbf{u}_{i\tau}, \tau = t+1, \ldots, \min\{t + \tau_i - 1, T\}, \\ &\qquad t = 1, \ldots, T-1, i = 1, \ldots, I, \ \mathbb{P}\text{—a.s.} \end{aligned} \tag{9b}$$

$$\begin{aligned} &0 \le \mathbf{s}_{jt} \le s_{jt}^{\max}, \quad 0 \le \mathbf{w}_{jt} \le w_{jt}^{\max}, \quad 0 \le \boldsymbol{\ell}_{jt} \le \ell_{jt}^{\max}, \\ &\qquad t = 1, \ldots, T, \quad j = 1, \ldots, J, \ \mathbb{P}\text{—a.s.} \end{aligned} \tag{9c}$$

$$\begin{aligned} &\boldsymbol{\ell}_{jt} = \boldsymbol{\ell}_{j,t-1} - \mathbf{s}_{jt} + \eta_j \mathbf{w}_{jt}, \quad t = 1, \ldots, T, \\ &\boldsymbol{\ell}_{j0} = \ell_j^{in}, \quad \boldsymbol{\ell}_{jT} = \ell_j^{end}, \quad j = 1, \ldots, j, \ \mathbb{P}\text{—a.s.} \end{aligned} \tag{9d}$$

Here some remarks concerning the interplay of the two stages are due. The first-stage solutions act as a basis for the recourse actions, which have to satisfy the second-stage constraints in a cost-optimal way. To this end we have to guarantee that the transition from the first to the second stage is feasible. While the static constraints (9a) and (9c) need no further consideration, we neglect the possible impact of the constraints (9d). Since we are confronted with purely thermal to thermally dominated power generation systems, this simplification is justified. More general systems could be incorporated into the model, making it more complex, though. The minimum up- and down-times constraints (9b) for the thermal units, on the other hand, need some refinement. In order to enforce compatibility between the first- and second-stage decisions, we introduce constraints that relate the scheduling behaviours of the two stages to each other. This means that we prevent a thermal unit from being switched on or off in the second stage if the scheduling history in the first stage prohibits that. The same canonical dependence is required in the other direction as well, i.e., we restrict switching in the first stage subject to the constraints set by the second-stage scheduling. To allow for more generality we further introduce probability quantiles $\pi_{it} \in (0, 1]$, $i = 1, \ldots, I$, $t = 1, \ldots, T$, that give the probability with which unit i has to fulfil the described restrictions at time t. Thus we have the constraints:

$$u_{it} - u_{i,t-1} \leq 1 - \left(\mathbf{u}^{\omega}_{i,\sigma-1} - \mathbf{u}^{\omega}_{i\sigma}\right), \quad \sigma = t, \ldots, \min\{t + \sigma_i - 1, T\}, \tag{10a}$$

$$u_{i,t-1} - u_{it} \leq 1 - \left(\mathbf{u}^{\omega}_{i\tau} - \mathbf{u}^{\omega}_{i,\tau-1}\right), \quad \tau = t, \ldots, \min\{t + \tau_i - 1, T\}, \tag{10b}$$

$$\mathbf{u}^{\omega}_{it} - \mathbf{u}^{\omega}_{i,t-1} \leq 1 - (u_{i,\sigma-1} - u_{i\sigma}), \quad \sigma = t, \ldots, \min\{t + \sigma_i - 1, T\}, \tag{10c}$$

$$\mathbf{u}^{\omega}_{i,t-1} - \mathbf{u}^{\omega}_{it} \leq 1 - (u_{i\tau} - u_{i,\tau-1}), \quad \tau = t, \ldots, \min\{t + \tau_i - 1, T\}, \tag{10d}$$

$$\forall \omega \in A_{it}, A_{it} \in \mathcal{A}, \mathbb{P}(A_{it}) \geq \pi_{it}, \quad i = 1, \ldots, I, \quad t = 1, \ldots, T - 1.$$

Observe the consequences of the *compatibility constraints* (10). The inequality (10b), for instance, represents a constraint for the second stage if and only if unit i is switched off in the first stage at time t. In this case it enforces that the thermal unit will not be switched on in the second stage as long as the unit is cooling for its minimum down-time in the first stage. The remaining inequalities have similar effects.

Furthermore, we introduce a subdivision of the set $\mathcal{I} := \{1, \ldots, I\}$ of all thermal units into two subsets $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $\mathcal{I}_1 \cup \mathcal{I}_2 = \mathcal{I}$ and the conditions

$$u_{it} = \mathbf{u}_{it}, \quad i \in \mathcal{I}_2, \quad t = 1, \ldots, T, \ \mathbb{P}\text{—a.s.},$$

are satisfied. This means that only some of the available thermal units may change their on/off state when compensating the uncertain data. From a modelling point of view this approach leads to a reduction of the number of binary variables corresponding to a unit $i \in \mathcal{I}_2$. Moreover, the case $\mathcal{I}_2 = \mathcal{I}$ conforms with the view taken in Carøe and Schultz (1998). There, all on/off decisions of the thermal units are regarded as long-term decisions and thus belong to the first stage only. This is a somewhat pessimistic attitude since it does not allow for any recourse action that includes switching-on or -off decisions. Therefore, the best objective value for this case is in general greater than the one for the original model, i.e., for the case $\mathcal{I}_2 = \emptyset$ (see Section 4 for details). Observe that (10) is clearly satisfied for all $i \in \mathcal{I}_2$.

The loading constraints (5) have to be adapted to the new situation. Here we distinguish between the two stages. As mentioned before we are looking for a solution to the here-and-now decisions that satisfies the uncertain demand with a certain probability and, moreover, allows an optimal scheduling in each of the scenarios. That is why the first-stage power outputs of all generation units have to meet at least the expected load, while the second-stage power outputs are required to satisfy the load $\boldsymbol{d}_t$ with probability one. Hence, the (modified) loading constraints are given by the following inequalities:

$$\sum_{i=1}^{I} p_{it} + \sum_{j=1}^{J} (s_{jt} - w_{jt}) \geq \mathbb{E}(\boldsymbol{d}_t), \quad t = 1, \ldots, T, \tag{11a}$$

$$\sum_{i=1}^{I} \mathbf{p}_{it} + \sum_{j=1}^{J} (\mathbf{s}_{jt} - \mathbf{w}_{jt}) \geq \boldsymbol{d}_t, \quad t = 1, \ldots, T, \ \mathbb{P}\text{—a.s.} \tag{11b}$$

A variant of (11a) arises when the term $\mathbb{E}(\boldsymbol{d}_t)$ is replaced by a probability quantile like $F_{d_t}^{-1}(\bar{\pi}_t)$, where F_{d_t} is the distribution function of $\boldsymbol{d}_t$ and $\bar{\pi}_t \in (0, 1)$ is a given probability. In both cases the constraint (11) means that the sum of the first-stage output power satisfies a certain predicted or approximated load and the second-stage decisions take care of satisfying the stochastic load with probability one. The reserve constraints (6) are modified in the same way:

$$\sum_{i=1}^{I} \left(p_{it}^{\max} u_{it} - p_{it}\right) \geq \mathbb{E}(\boldsymbol{r}_t), \quad t = 1, \ldots, T, \tag{12a}$$

$$\sum_{i=1}^{I} \left(p_{it}^{\max} \mathbf{u}_{it} - \mathbf{p}_{it}\right) \geq \boldsymbol{r}_t, \quad t = 1, \ldots, T, \ \mathbb{P}\text{—a.s.} \tag{12b}$$

Again the second-stage decisions cover the specified amount with probability one, while the first-stage spinning reserve meets at least the expected reserve.

Finally we incorporate the stochastic fuel and electricity prices into the model. To this end we define the random functions $\boldsymbol{C}_{it}$, which describe the costs for operating the thermal unit i in the second-stage during time period t, in the following way:

$$\boldsymbol{C}_{it}(p, u) := \max_{l=1,\ldots,\bar{l}} \{\boldsymbol{a}_{ilt} p + \boldsymbol{b}_{ilt} u\}, \tag{13}$$

where $\boldsymbol{a}_{ilt}, \boldsymbol{b}_{ilt}$ are components of the random variable $\boldsymbol{\xi}_t$. They represent stochastic cost coefficients such that $\boldsymbol{C}_{it}(\cdot, 1)$ is $\mathbb{P}$-almost surely convex and increasing on $\mathbb{R}_+$. We define the cost functions C_{it} for the first stage accordingly, taking the expected values of the price coefficients, i.e.,

$$C_{it}(p, u) := \max_{l=1,\ldots,\bar{l}} \{\mathbb{E}(\boldsymbol{a}_{ilt}) p + \mathbb{E}(\boldsymbol{b}_{ilt}) u\}. \tag{14}$$

Observe that (14) corresponds to $\alpha_{ilt} = \mathbb{E}(\boldsymbol{a}_{ilt})$, $\beta_{ilt} = \mathbb{E}(\boldsymbol{b}_{ilt})$ in (7). The effect of stochastic prices on the startup costs is modelled in a similar way. More precisely, taking $\gamma_{it} = \mathbb{E}(\boldsymbol{c}_{it})$ in (8), we have

$$\boldsymbol{S}_{it}(\mathbf{u}_i) := \boldsymbol{c}_{it} \max\{\mathbf{u}_{it} - \mathbf{u}_{i,t-1}, 0\}, \tag{15a}$$

$$S_{it}(u_i) := \mathbb{E}(\boldsymbol{c}_{it}) \max\{u_{it} - u_{i,t-1}, 0\}, \tag{15b}$$

where $\boldsymbol{c}_{it}$ are stochastic startup cost coefficients and $\mathbf{u}_i := (\mathbf{u}_{it})_{t=0}^{T}$, $\mathbf{u}_{i0} = u_{i0}$ $\mathbb{P}$—a.s., $i = 1, \ldots, I$.

In consistency with common two-stage stochastic programming the objective function corresponds to the total operating costs of the generation system in the first stage plus the expected costs in the second stage, i.e.,

$$\sum_{i=1}^{I}\sum_{t=1}^{T}[C_{it}(p_{it}, u_{it}) + S_{it}(u_i)] + \mathbb{E}\sum_{i=1}^{I}\sum_{t=1}^{T}[\boldsymbol{C}_{it}(\mathbf{p}_{it}, \mathbf{u}_{it}) + \boldsymbol{S}_{it}(\mathbf{u}_i)]. \tag{16}$$

Then the stochastic power production planning model consists in minimizing the objective function over all deterministic decisions (u, p, w, s) and all stochastic decisions $(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w}) \in L^\infty(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{2T(I+J)})$ satisfying the constraints (1)–(3), (9)–(12). The model represents a two-stage stochastic mixed-integer program with relatively complete recourse involving $2(I + J)T$ deterministic and $2(I + J)T$ stochastic decision variables.

Remark 1. Note that the original model ($\mathcal{I}_2 = \emptyset$) decomposes into independent single scenario problems if we do not enforce the compatibility constraints (10), or any constraint similar to (10). Note furthermore, that if we had introduced

$$u_{it} - u_{i,t-1} \le \mathbf{u}_{i\sigma}, \quad \sigma = t+1, \ldots, \min\{t + \sigma_i - 1, T\}, \tag{17a}$$

$$u_{i,t-1} - u_{it} \le 1 - \mathbf{u}_{i\tau}, \quad \tau = t+1, \ldots, \min\{t + \tau_i - 1, T\}, \tag{17b}$$

$$\mathbf{u}_{it} - \mathbf{u}_{i,t-1} \le u_{i\sigma}, \quad \sigma = t+1, \ldots, \min\{t + \sigma_i - 1, T\}, \tag{17c}$$

$$\mathbf{u}_{i,t-1} - \mathbf{u}_{it} \le 1 - u_{i\tau}, \quad \tau = t+1, \ldots, \min\{t + \tau_i - 1, T\}, \tag{17d}$$

$$\forall\omega \in A_{it}, \quad A_{it} \in \mathcal{A}, \quad \mathbb{P}(A_{it}) \ge \pi_{it}, \quad i = 1, \ldots, I, \quad t = 1, \ldots, T-1$$

instead of (10), that is, if we had simply enforced the minimum up- and down-times across the stages, then this would have implied that the scheduling decisions in the two stages are almost identical. In particular, this model would have not allowed for recourse actions that include switching-on or -off decisions. Insofar it corresponds to the case $\mathcal{I}_2 = \mathcal{I}$.

3. Stochastic Lagrangian relaxation

The stochastic program elaborated in the previous section is almost separable with respect to the units since only the constraints (11) and (12) couple different units. This structure allows us to apply a stochastic version of the Lagrangian relaxation by associating stochastic Lagrange multipliers with the coupling constraints. We restate the relevant inequalities in order to illustrate some modifications:

$$\sum_{i=1}^{I} p_{it} + \sum_{j=1}^{J}(s_{jt} - w_{jt}) \ge \mathbb{E}(\boldsymbol{d}_t), \quad t = 1, \ldots, T, \tag{18a}$$

$$\sum_{i=1}^{I} \mathbf{p}_{it} + \sum_{j=1}^{J}(\mathbf{s}_{jt} - \mathbf{w}_{jt}) \ge \boldsymbol{d}_t, \quad t = 1, \ldots, T, \quad \mathbb{P}\text{—a.s.}, \tag{18b}$$

$$\sum_{i=1}^{I} u_{it} p_{it}^{\max} + \sum_{j=1}^{J} (s_{jt} - w_{jt}) \geq \mathbb{E}(\boldsymbol{d}_t + \boldsymbol{r}_t), \quad t = 1, \ldots, T, \tag{18c}$$

$$\sum_{i=1}^{I} \mathbf{u}_{it} p_{it}^{\max} + \sum_{j=1}^{J} (\mathbf{s}_{jt} - \mathbf{w}_{jt}) \geq \boldsymbol{d}_t + \boldsymbol{r}_t, \quad t = 1, \ldots, T, \ \mathbb{P}\text{—a.s.} \tag{18d}$$

This formulation of the reserve constraints is equivalent to the one in Section 2 in the sense that—due to the monotonicity of the objective function—a solution that satisfies (18) can be improved to a better solution that satisfies (11), (12). More importantly, it will enable a stochastic Lagrangian heuristics (see Section 3.4) similar to the deterministic heuristics described in Gollmer et al. (1999) and Zhuang and Galiana (1988).

We relax the coupling constraints (18) by introducing Lagrange multipliers $\boldsymbol{\lambda} := (\lambda^1, \lambda^2, \boldsymbol{\lambda}^3, \boldsymbol{\lambda}^4)$, where $\lambda^1, \lambda^2 \in \mathbb{R}_+^T$ and $\boldsymbol{\lambda}^3, \boldsymbol{\lambda}^4 \in L^1(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R}_+^T)$. For convex two-stage stochastic programs this approach is justified by the general duality theory developed in Rockafellar and Wets (1976). Hence, suppose for the moment that the integrality constraints in (1a) and (9a) are relaxed to $u_{it}, \mathbf{u}_{it} \in [0, 1]$ so that the program in Section 2 becomes convex. Then, setting $x := (u, p, s, w)$ and $\mathbf{x} := (\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w})$, the *Lagrangian* takes the form:

$$\begin{aligned}
L(x, \mathbf{x}; \boldsymbol{\lambda}) := \sum_{t=1}^{T} \Bigg\{ & \sum_{i=1}^{I} [C_{it}(p_{it}, u_{it}) + S_{it}(u_i)] + \mathbb{E} \sum_{i=1}^{I} [\boldsymbol{C}_{it}(\mathbf{p}_{it}, \mathbf{u}_{it}) + \boldsymbol{S}_{it}(\mathbf{u}_i)] \\
& + \lambda_t^1 \left[\mathbb{E}(\boldsymbol{d}_t) - \sum_{i=1}^{I} p_{it} - \sum_{j=1}^{J} (s_{jt} - w_{jt}) \right] \\
& + \lambda_t^2 \left[\mathbb{E}(\boldsymbol{d}_t + \boldsymbol{r}_t) - \sum_{i=1}^{I} u_{it} p_{it}^{\max} - \sum_{j=1}^{J} (s_{jt} - w_{jt}) \right] \\
& + \mathbb{E}\left(\boldsymbol{\lambda}_t^3 \left[\boldsymbol{d}_t - \sum_{i=1}^{I} \mathbf{p}_{it} - \sum_{j=1}^{J} (\mathbf{s}_{jt} - \mathbf{w}_{jt}) \right] \right) \\
& + \mathbb{E}\left(\boldsymbol{\lambda}_t^4 \left[\boldsymbol{d}_t + \boldsymbol{r}_t - \sum_{i=1}^{I} \mathbf{u}_{it} p_{it}^{\max} - \sum_{j=1}^{J} (\mathbf{s}_{jt} - \mathbf{w}_{jt}) \right] \right) \Bigg\}.
\end{aligned} \tag{19}$$

Hence, with the *dual function*

$$D(\boldsymbol{\lambda}) := \min_{(x, \mathbf{x})} L(x, \mathbf{x}; \boldsymbol{\lambda}) \text{ s.t. constraints (1)–(3), (9)–(10)} \tag{20}$$

the *dual problem* reads

$$\max\{D(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \mathbb{R}_+^{2T} \times L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}_+^{2T})\}. \tag{21}$$

This means in particular that the stochastic multiplier processes $\boldsymbol{\lambda}^3$ and $\boldsymbol{\lambda}^4$ are nonnegative $\mathbb{P}$-almost surely. Due to the presence of integrality constraints in (1a) and (9a) in the general case, the primal problem is nonconvex. That is why the optimal value of the dual problem (21) provides only a lower bound for the primal optimal costs.

The minimization in (20) decomposes into stochastic single unit subproblems. Specifically the dual function

$$\begin{aligned} D(\boldsymbol{\lambda}) &= \sum_{i=1}^{I} D_i(\boldsymbol{\lambda}) + \sum_{j=1}^{J} \hat{D}_j(\boldsymbol{\lambda}) \\ &\quad + \sum_{t=1}^{T} \left[\lambda_t^1 \mathbb{E}(\boldsymbol{d}_t) + \lambda_t^2 \mathbb{E}(\boldsymbol{d}_t + \boldsymbol{r}_t) + \mathbb{E}\big(\boldsymbol{\lambda}_t^3 \boldsymbol{d}_t + \boldsymbol{\lambda}_t^4 (\boldsymbol{d}_t + \boldsymbol{r}_t)\big) \right] \end{aligned} \tag{22}$$

may be evaluated by solving the thermal subproblems $D_i(\boldsymbol{\lambda})$:

$$\begin{aligned} \min_{(u_i, \mathbf{u}_i)} \quad & \sum_{t=1}^{T} \Bigg[\min_{p_{it}} \left\{ C_{it}(p_{it}, u_{it}) - \lambda_t^1 p_{it} \right\} - \lambda_t^2 u_{it} p_{it}^{\max} + S_{it}(u_i) \\ & \quad + \mathbb{E} \left\{ \min_{\mathbf{p}_{it}} \left\{ \boldsymbol{C}_{it}(\mathbf{p}_{it}, \mathbf{u}_{it}) - \boldsymbol{\lambda}_t^3 \mathbf{p}_{it} \right\} - \boldsymbol{\lambda}_t^4 \mathbf{u}_{it} p_{it}^{\max} + \boldsymbol{S}_{it}(\mathbf{u}_i) \right\} \Bigg] \\ \text{s.t.} \quad & \text{(1a), (3), (9a), (9b), (10)} \end{aligned} \tag{23}$$

and the hydro subproblems $\hat{D}_j(\boldsymbol{\lambda})$:

$$\begin{aligned} \min_{(w_j; s_j, \mathbf{w}_j, \mathbf{s}_j)} \quad & \sum_{t=1}^{T} \left[(\lambda_t^1 + \lambda_t^2)(w_{jt} - s_{jt}) + \mathbb{E}\big(\boldsymbol{\lambda}_t^3 + \boldsymbol{\lambda}_t^4\big)(\mathbf{w}_{jt} - \mathbf{s}_{jt}) \right] \\ \text{s.t.} \quad & \text{(1b)–(2), (9c)–(9d).} \end{aligned} \tag{24}$$

The two kinds of subproblems represent two-stage stochastic programming models for the operation of a single unit. While the thermal subproblem (23) is a mixed-integer two-stage stochastic program that reduces to a combinatorial two-stage stochastic problem, the hydro subproblem (24) is a linear two-stage model with stochastic costs and stochastic right-hand sides. It is worth noting that problem (23) simplifies essentially for the case that $i \in \mathcal{I}_2$ since then the compensation program does not contain any binary decisions.

Extending Lagrangian relaxation approaches for deterministic power management models, our method for solving the model in Section 2 consists of the following ingredients:

(a) Generating scenarios $\hat{\xi}_n, n = 1, \ldots, N$, for the stochastic process $\boldsymbol{\xi}$ and replacing it with this discrete approximation;
(b) Solving the dual problem (21) by a proximal bundle method using function and subgradient information (note that (21) has dimension $2T(N+1)$);
(c) Solving the single unit subproblems with stochastic dynamic programming for (23) and a special descent algorithm for (24);
(d) Applying a Lagrangian heuristics for determining a primal feasible solution;
(e) Employing an economic dispatch for determining a nearly optimal solution.

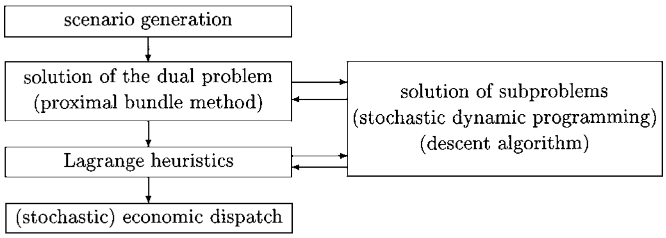

Figure 2. Structure of the solution method.

The interaction of these components is illustrated in figure 2. In the remaining part of this section we provide a description of the components (b)–(e). The scenario generation will be covered briefly in Section 4.

3.1. *Proximal bundle method*

We consider the maximization of the dual concave function D on the set $\mathbb{R}_+^L$, where $L := 2T(N+1)$. Function values $D(\boldsymbol{\lambda})$ are evaluated according to (22) and a corresponding subgradient $g(\boldsymbol{\lambda}) \in \partial D(\boldsymbol{\lambda})$ is given by $(g_0(\boldsymbol{\lambda}), h_0(\boldsymbol{\lambda}), g_1(\boldsymbol{\lambda}), \ldots, g_N(\boldsymbol{\lambda}), h_1(\boldsymbol{\lambda}), \ldots, h_N(\boldsymbol{\lambda}))$, where $g_n(\boldsymbol{\lambda})$ for $n = 1, \ldots, N$ is equal to the realization of the stochastic process

$$\left\{ \boldsymbol{d}_t - \sum_{i=1}^{I} \mathbf{p}_{it}(\boldsymbol{\lambda}) - \sum_{j=1}^{J} (\mathbf{s}_{jt}(\boldsymbol{\lambda}) - \mathbf{w}_{jt}(\boldsymbol{\lambda})) \right\}_{t=1}^{T}$$

in scenario $\hat{\xi}_n$ and $h_n(\boldsymbol{\lambda})$ for $n = 1, \ldots, N$ is equal to

$$\left\{ \boldsymbol{d}_t + \boldsymbol{r}_t - \sum_{i=1}^{I} \mathbf{u}_{it}(\boldsymbol{\lambda}) p_{it}^{\max} - \sum_{j=1}^{J} (\mathbf{s}_{jt}(\boldsymbol{\lambda}) - \mathbf{w}_{jt}(\boldsymbol{\lambda})) \right\}_{t=1}^{T}$$

in scenario $\hat{\xi}_n$. Similarly, $g_0(\boldsymbol{\lambda})$ and $h_0(\boldsymbol{\lambda})$ correspond to the violations of the deterministic first-stage load and reserve constraints in $x(\boldsymbol{\lambda})$. Here $x(\boldsymbol{\lambda})$ and $\mathbf{x}(\boldsymbol{\lambda}) = (\mathbf{u}(\boldsymbol{\lambda}), \mathbf{p}(\boldsymbol{\lambda}), \mathbf{s}(\boldsymbol{\lambda}), \mathbf{w}(\boldsymbol{\lambda}))$ are Lagrangian solutions, i.e., they belong to $\arg\min_{(x,\mathbf{x})} L(x, \mathbf{x}; \boldsymbol{\lambda})$. We assume that the set of dual maximizers is nonempty, which is guaranteed if the primal problem is feasible. Hence the proximal bundle method (Kiwiel, 1990) may be used for solving the dual problem. This method generates a sequence $\{\boldsymbol{\lambda}_c^k\}_{k=1}^{\infty} \subset \mathbb{R}_+^L$ converging to some maximizer $\boldsymbol{\lambda}^*$, and trial points $\boldsymbol{\lambda}^k \in \mathbb{R}_+^L$ for evaluating subgradients $g^k := g(\boldsymbol{\lambda}^k)$ starting with an arbitrary point $\boldsymbol{\lambda}_c^1 - \boldsymbol{\lambda}^1 \in \mathbb{R}_+^L$. Iteration k uses the linearizations.

$$\tilde{D}^l(\cdot) := D(\boldsymbol{\lambda}^l) + \langle \cdot - \boldsymbol{\lambda}^l, g^l \rangle \geq D(\cdot)$$

of D and its polyhedral upper approximation

$$\tilde{D}_k(\cdot) := \min_{l \in L^k} \tilde{D}^l(\cdot) \quad \text{with} \quad k \in L^k \subset \{1, \ldots, k\}, \tag{25}$$

for finding the next trial point

$$\boldsymbol{\lambda}^{k+1} \in \arg\max\left\{\tilde{D}_k(\boldsymbol{\lambda}) - \frac{1}{2}u_k\left\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_c^k\right\|^2 : \boldsymbol{\lambda} \in \mathbb{R}_+^L\right\}, \tag{26}$$

where the proximity weight $u_k > 0$ and the penalty term $\|\cdot\|^2$ should keep $\boldsymbol{\lambda}^{k+1}$ close to the prox-center $\boldsymbol{\lambda}_c^k$. An ascent step to $\boldsymbol{\lambda}_c^{k+1} = \boldsymbol{\lambda}^{k+1}$ occurs if $\boldsymbol{\lambda}^{k+1}$ is significantly better than $\boldsymbol{\lambda}_c^k$ as measured by

$$D(\boldsymbol{\lambda}^{k+1}) \geq D\big(\boldsymbol{\lambda}_c^k\big) + \kappa\delta_k,$$

where $\kappa \in (0, 1)$ is a fixed parameter and $\delta_k = \tilde{D}_k(\boldsymbol{\lambda}^{k+1}) - D(\boldsymbol{\lambda}_c^k) \geq 0$ is the predicted ascent. If $\delta_k = 0$, then $\boldsymbol{\lambda}_c^k$ is optimal and the method may stop. If a significant improvement of the objective value, on the other hand, cannot be achieved, a null step $\boldsymbol{\lambda}_c^{k+1} = \boldsymbol{\lambda}_c^k$ takes place. This will improve the next polyhedral function $\tilde{D}_{k+1}$. Strategies for updating u_k and choosing L^{k+1} are discussed in Kiwiel (1990). For the choice of L^{k+1} both subgradient selection and subgradient aggregation can be employed. In the first case, since D is polyhedral, finite convergence is guaranteed. But since subgradient selection may require too much storage, alternatively one may use subgradient aggregation, in which groups of past linearizations are replaced by their convex combinations so that at most NGRAD ≥ 2 linearizations are stored. Here finite convergence need not occur, but $\boldsymbol{\lambda}_c^k \to \boldsymbol{\lambda}^*$ and $\delta_k \to 0$ so that for any optimality tolerance opt.tol > 0 the method eventually meets the stopping criterion $\delta_k \geq$ opt.tol $(1 + |D(\boldsymbol{\lambda}_c^k|)$.

3.2. Stochastic dynamic programming

In order to solve the thermal subproblem (23) for unit i by dynamic programming we consider a graph of nodes, where each node q corresponds to the recent history of unit i. In particular, any node q represents an $(N + 1)$ tuple of states that comprises the recent scheduling behaviour of unit i in the first stage and in all scenarios $\hat{\xi}_n, n = 1, \ldots, N$, of the second stage. Hence, the minimum up/down-time constraints and the compatibility constraints can be expressed as feasible transitions in the state graph. In the scope of deterministic unit commitment similar graph representations are well known (see e.g. Gollmer et al., 1999; Takriti et al., 2000; Zhuang and Galiana, 1988). A part of such a (deterministic) transition graph is shown in figure 3, where we chose minimum up- and down-times of 2 and 3 hours, respectively, in order to constrain the complexity of the figure. In our case of two-stage stochastic programming the dynamic programming graph looks far more complicated, though. We present a corresponding part of it in figure 3 for the instance $N = 1$. The figure shows possible state transitions for an arbitrary time t, where the arrows refer to feasible transitions. Here each node represents a pair of states $(\tau, \bar{\tau})$, where τ denotes the state of the unit in the first stage and is represented in the figure by a capital number, while $\bar{\tau}$ refers to the state in the only scenario of the second stage and is shown by a small number. The dashed lines indicate the border between on-line and off-line states in the respective stages. Apart from the minimum up/down-time constraints the feasible state transitions now also incorporate the compatibility constraints. We illustrate this fact with the help of node

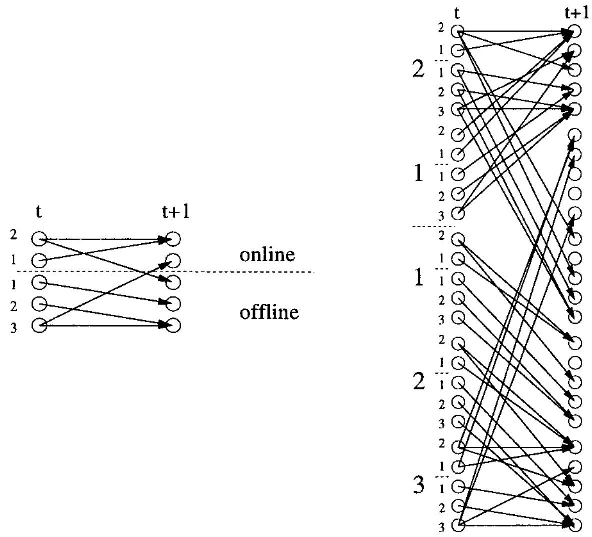

Figure 3. Transition graph for 2 time periods in the deterministic case (left) and in the stochastic case with $N = 1$ (right).

$(1^{\text{on}}, 2^{\text{on}})$ (s. figure 3). The only feasible state transition leads to node $(2^{\text{on}}, 2^{\text{on}})$. When only the minimum up/down-time constraints in both stages were taken into account, however, there would be another feasible state transition to node $(2^{\text{on}}, 1^{\text{off}})$. But this would correspond to a switching off of the unit in the second stage while the unit satisfies the minimum up-time constraint in the first stage. This would clearly violate the compatibility constraint (10a) and is thus no feasible state transition.

Observe that the size of the graph grows exponentially with the number of scenarios N.

Let $\varphi_{it}(q)$ denote the node weight at time t for node q and $\vartheta_{it}(q, \tilde{q})$ the arc weight for the arc from node q to node $\tilde{q}$ at time t in the state transition graph. The node weights $\varphi_{it}(q)$ are equal to 0 for nodes q that represent off-line states in both the first and $\mathbb{P}$-almost surely in the second stage. In general the following holds for all nodes:

$$\varphi_{it}(q) = \min_{p_{it}^{\min} \le p \le p_{it}^{\max}} \left\{ C_{it}(p, 1) - \lambda_t^1 p - \lambda_t^2 p_{it}^{\max} \right\} \chi_{\{\hat{u}_i^q = 1\}} + \mathbb{E}\left[\min_{p_{it}^{\min} \le \mathbf{p} \le p_{it}^{\max}} \left\{ \mathbf{C}_{it}(\mathbf{p}, 1) - \boldsymbol{\lambda}_t^3 \mathbf{p} - \boldsymbol{\lambda}_t^4 p_{it}^{\max} \right\} \chi_{\{\hat{\mathbf{u}}_i^q = 1\}} \right], \tag{27}$$

where χ is the indicator function on the set $\mathcal{Q}$ of all nodes and $\hat{u}_i^q$ and $\hat{\mathbf{u}}_i^q$ indicate whether node q represents an on-line or off-line state in the first stage and second stage, respectively. The arc weights $\vartheta_{it}(q, \tilde{q})$ describe start-up costs for the thermal unit. They are independent of $\boldsymbol{\lambda}$ and are non-zero only for arcs leading from off-line states to on-line states. The cost-to-go functions are given by

$$\psi_{it}(q) := \varphi_{it}(q) + \min_{\tilde{q}} \{ \vartheta_{it}(q, \tilde{q}) + \psi_{i,t+1}(\tilde{q}) \}, \tag{28}$$

where $\psi_{iT}(q) := \varphi_{iT}(q)$. Solving the combinatorial problem (23) is now equivalent to finding a shortest path from the set of source nodes $\mathcal{Q}_1 \subset \mathcal{Q}$ (i.e., all nodes at time $t = 1$) to the set of destination nodes $\mathcal{Q}_T$ in the state transition graph.

The dynamic programming algorithm works as follows. First, the cost-to-go functions are computed for all nodes q via the backward recursion (28). Then the optimal decisions are obtained from the shortest path by a forward computation starting at $q_1 \in \arg\min\{\psi_{i1}(q) : q \in \mathcal{Q}_1\}$. It is worth mentioning that the algorithm described above gets computationally expensive with a growing number of scenarios. However, since the thermal subproblems have to be solved many times in the course of the dual maximization, efficiency is of utmost importance. For the improvement of the algorithm the following fact is beneficial. In most of the cases during the dual maximization the optimal solution to (23) does not depend on whether we enforce the compatibility constraints or not. Furthermore, problem (23) decomposes into single scenario subproblems once we relax the constraints (10). These can be solved by a deterministic dynamic programming algorithm, which is far more efficient due to the relative simplicity of the involved dynamic programming graph (cf. figure 3). Hence, the algorithm for solving (23) incorporates the following steps. First, neglecting the compatibility constraints, $N + 1$ deterministic single scenario subproblems are solved. Note that solving the relaxed problem for the first-stage decisions is equivalent to the program for one of the single scenarios. If the compatibility constraints are not violated by the determined optimal solution, the algorithm terminates. Otherwise the compatibility constraints have to be enforced and stochastic dynamic programming is employed in order to obtain the optimal solution.

3.3. *Hydro subproblems*

Considering the hydro subproblem (24) we observe that it can be decomposed into linear single scenario subproblems. These are solved with a specialized descent method that creates a finite sequence of hydro decisions with decreasing objective values. For our purposes we implemented a deterministic version of the descent algorithm that is described in Nowak (2000).

3.4. *Lagrangian heuristics*

When the bundle method delivers an optimal multiplier $\boldsymbol{\lambda}^*$, the optimal value $D(\boldsymbol{\lambda}^*)$ provides a lower bound for the optimal cost of the primal mode. In general, however, the "dual optimal" scheduling decisions $(x(\boldsymbol{\lambda}^*), \mathbf{x}(\boldsymbol{\lambda}^*))$ violate the load and reserve constraints (11) and (12). In the following, we describe Lagrangian heuristics that determine a nearly optimal primal decision starting from the optimal multiplier $\boldsymbol{\lambda}^*$. We will distinguish between the general case and the instance $\mathcal{I}_2 = \mathcal{I}$. Our first heuristics aims at treating all scenarios $\hat{\xi}_n$, $n = 1, \ldots, N$, separately. That is, we try to find a nearly optimal allocation for the decision variables in each of the scenarios independently of all the others. The procedure starts with the first stage (treating it like one of the scenarios) followed by the scenarios in decreasing order of their probabilities. We describe the procedure for an arbitrary scenario $\hat{\xi}_n$. The heuristics iteratively employs two steps that interact with each other. In the *first step* the hydro decisions $\mathbf{s}_j$ and $\mathbf{w}_j$ are rescheduled in order to meet the reserve constraints (18d):

$$\sum_{i=1}^{I} \mathbf{u}_{it} p_{it}^{\max} + \sum_{j=1}^{J} (\mathbf{s}_{jt} - \mathbf{w}_{jt}) \geq \boldsymbol{d}_t + \boldsymbol{r}_t$$

as best as possible. To this end our procedure reduces the value

$$\boldsymbol{d}_t + \boldsymbol{r}_t + \sum_{j=1}^{J} (\mathbf{w}_{jt} - \mathbf{s}_{jt}) \tag{29}$$

by modifying the hydro schedules at those times t where the constraint is violated and the value (29) is largest in a certain set of neighbouring time periods. This local rescheduling procedure is repeated several times (see also Gollmer et al., 1999). In a *second step* the hydro variables are fixed and, following Zhuang and Galiana (1988), we search for binary variables $\mathbf{u}_i$ that further reduce the violation of (18d). More precisely, we are looking for a unit i that causes lowest costs when being switched on at the period t^* where (18d) is violated most. To this end we consider all thermal units that are scheduled off-line at time t^*. With the aid of dynamic programming—in addition with single must-on constraints—we calculate the minimal increase $\Delta\boldsymbol{\lambda}_{t^*}$ that is necessary to switch on one of the respective units at time t^*. Here we try to restrict ourselves to those units for which we need not emply the stochastic version of the dynamic programming algorithm (cf. Section 3.2). In this way we save computation time without running the risk of loosing near optimality, since scheduling decisions that put restrictions on the first or second stage, respectively, are likely to produce extra costs in the objective function. Having increased $\boldsymbol{\lambda}_{t^*}$ by the computed amount and having solved the thermal subproblems (23) for the new $\boldsymbol{\lambda}$, the procedure returns to the first step. This is repeated until the reserve constraint is satisfied in all time periods. Since the described technique does not distinguish between identical units that appear quite often in real-life power systems, the startup costs of such units are slightly modified.

Our strategy for the case of $\mathcal{I}_2 = \mathcal{I}$ differs in some respects from the general approach. As indicated above we treat the two stages and all scenarios at the same time. This strategy is made necessary by the fact that all thermal units exhibit the same scheduling behaviour in the two stages, i.e., we have $\mathbf{u}_i = u_i$, $\mathbb{P}$—a.s., $i \in \mathcal{I}$. Again we iteratively make use of two steps. These are a water rescheduling procedure like before, followed by a selective switching on of a thermal unit. However, this time the regarded period t^*, in which a unit is going to be switched on, will be chosen differently. Particularly we will consider the mean value of the reserve constraint violation, i.e., we define

$$\begin{aligned} v(t) := \mathbb{E}[\boldsymbol{d}_t - \boldsymbol{r}_t] + \sum_{j=1}^{J} (w_{jt} - s_{jt}) - \sum_{i=1}^{I} u_{it} p_{it}^{\max} \\ + \mathbb{E}\left[\boldsymbol{d}_t + \boldsymbol{r}_t + \sum_{j=1}^{J} (\mathbf{w}_{jt} - \mathbf{s}_{jt}) - \sum_{i=1}^{I} \mathbf{u}_{it} p_{it}^{\max} \right] \end{aligned}$$

and take $t^* \in \arg\max\{v(t) : t = 1, \ldots, T\}$. Then we calculate the necessary increase $\Delta\boldsymbol{\lambda}_{t^*}$ of the Lagrange multiplier in order to switch on one thermal unit at time t^*. As in the general case the two steps will be employed repeatedly until the reserve constraint is satisfied in the time periods.

3.5. *Economic dispatch*

The Lagrangian heuristics terminates with a binary schedule $(u_i, \mathbf{u}_i)$ for each thermal unit $i \in \mathcal{I}$, such that a primal feasible solution $(x, \mathbf{x})$ with these binary components exists. Now the objective value can still be improved by changing the not yet optimal values of the continuous variables. This task amounts to an economic dispatch problem. Since the binary decisions are kept fixed, this problem can be decomposed into single scenario subproblems. In particular, we can determine cost-optimal schedules for each scenario independently of all the others. To this end we employ a deterministic version of the economic dispatch algorithm presented in Nowak (2000).

4. Numerical results

The stochastic Lagrangian relaxation algorithm was implemented in C++ except for the proximal bundle method, for which the FORTRAN-package NOA 3.0 (Kiwiel, 1993) was used as a callable library. For testing the implementation a number of load scenarios was simulated from the following SARIMA $(7, 0, 9) \times (0, 1, 0)_{168}$ time series model for the load process (see Gröwe-Kuska et al., 2002):

$$\begin{aligned}\boldsymbol{d}_t &= \hat{\phi}_1\boldsymbol{d}_{t-1} + \cdots + \hat{\phi}_7\boldsymbol{d}_{t-7} - \boldsymbol{d}_{t-168} - \hat{\phi}_1\boldsymbol{d}_{t-169} - \cdots - \hat{\phi}_7\boldsymbol{d}_{t-175}\\ &\quad + \boldsymbol{Z}_t + \hat{\theta}_1\boldsymbol{Z}_{t-1} + \cdots + \hat{\theta}_9\boldsymbol{Z}_{t-9}, \quad t \in \mathbb{Z},\end{aligned}$$

where the model coefficients $\hat{\phi}_i$ and $\hat{\theta}_j$ are given and $\boldsymbol{Z}_t$, $t \in \mathbb{Z}$, are independent, identically normal distributed random variables with given mean and standard deviation.

Furthermore, the stochastic prices have been simulated by a discretized geometric Brownian motion. More precisely, we have simulated a solution to the stochastic differential equation

$$d\boldsymbol{X}_t = \boldsymbol{X}_t(\sigma\, d\boldsymbol{W}_t + m\, dt), \tag{30}$$

where $\boldsymbol{W}$ is a standard Brownian motion and σ, m are the volatility and drift of $\boldsymbol{X}$, respectively. A solution to (30) with initial value $\boldsymbol{X}_0$ is given by

$$\boldsymbol{X}_t = \boldsymbol{X}_0 \exp\left(\sigma\boldsymbol{W}_t + \left(m - \frac{1}{2}\sigma^2\right)t\right)$$

and can be easily sampled from. For our purposes we let $\boldsymbol{X}_0$ be normally distributed with $\mathrm{N}(1, \sigma^2)$ independent of $\boldsymbol{W}$, and $m > \frac{1}{2}\sigma^2$. All our computational results have been obtained with a choice of the price volatility of $\sigma = 0.01$ and with $m = 5.2 \times 10^{-5}$. An example of

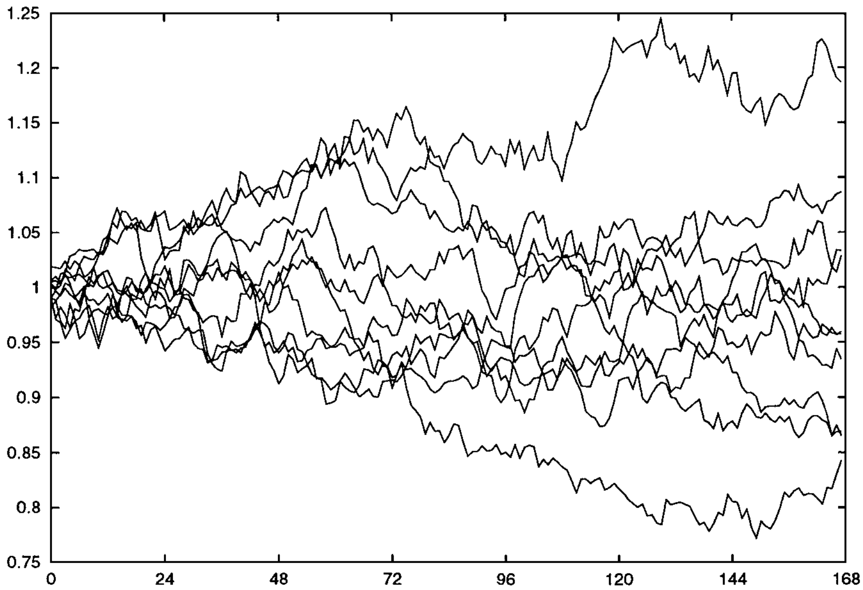

Figure 4. 10 trajectories from the discretized geometric Brownian motion.

generated price scenarios is shown in figure 4. In order to describe the cost functions $\boldsymbol{C}_{it}$ and $\boldsymbol{S}_{it}$ (cf. Section 2), we finally set

$$\boldsymbol{a}_{ilt} = \bar{a}_{ilt}\boldsymbol{X}_t, \quad \boldsymbol{b}_{ilt} = \bar{b}_{ilt}\boldsymbol{X}_t, \quad \boldsymbol{c}_{it} = \bar{c}_{it}\boldsymbol{X}_t,$$

where $\bar{a}_{ilt}, \bar{b}_{ilt}, \bar{c}_{it}$ are characteristic mean values of the price coefficients.

Test runs have been carried out for the weekly production planning (i.e., $T = 168$) of a configuration of the VEAG system comprising 25 thermal units and 7 pumped storage plants and for a number of scenarios ranging from 5 to 100. The dimensions of the corresponding primal optimization problems are shown in Table 1. Furthermore, the compatibility

Table 1. Dimension of the primal optimization problem.

	Variables			
Scenarios	Binary	Continuous	Constraints	Nonzeros
1	8400	13104	26882	39314
5	25200	39312	80646	117942
10	46200	72072	147851	216227
20	88200	137592	282261	412797
50	214200	334152	685491	1002507
100	424200	661752	1357541	1985357

Table 2. Computing times and gaps.

		Opt.tol: 10^{-3}			Opt.tol: 10^{-4}			
		Time (min)			Time (min)			
Scenarios	$\mathcal{I}_2$	NOA	All	Gap (%)	NOA	All	Gap (%)	obj.val ($\times 10^8$)
5	$\emptyset$	0:16	1:00	0.61	0:41	1:45	0.21	1.15511
5	$\emptyset$	0:18	0:35	0.25	0:45	1:18	0.17	1.49352
10	$\emptyset$	0:57	2:50	0.45	1:55	5:08	0.30	1.15563
10	$\emptyset$	0:44	2:04	0.31	1:32	3:53	0.13	1.42486
5	$\mathcal{I}$	0:10	0:28	0.90	0:26	0:43	0.70	1.50801
10	$\mathcal{I}$	0:27	0:59	1.25	1:14	1:43	0.87	1.44497
50	$\mathcal{I}$	9:00	12:04	1.83	12:36	16:07	1.36	1.41429
100	$\mathcal{I}$	30:17	35:31	1.99	35:37	42:48	1.70	1.43047

(NOA 3.0: NGRAD = 20).

constraints have been enforced $\mathbb{P}-$almost surely, i.e., we chose $\pi_{it} = 1, t = 1, \ldots, T, i = 1, \ldots, I$. The test runs have been performed on an HP 9000 (780/J280) Compute-Server with 180 MHz frequency and 768 MByte main memory under HP-UX 10.20. Table 2 shows computing times and gaps for different choices of the optimality tolerance for the proximal bundle method. The results show that a smaller optimality tolerance leads to smaller gaps at the expense of higher computing times. Here the gap refers to the relative difference of the cost for the scheduling decision $(x, \mathbf{x})$ and the optimal value $D(\boldsymbol{\lambda}^*)$ of the dual problem. Figure 5 provides a sample output of the algorithm for the general case. The performance of the algorithm in this situation (i.e., $\mathcal{I}_2 = \emptyset$) is closely related to the efficiency of the thermal subproblem solver. In particular, it depends on how often the stochastic dynamic programming algorithm is used during the dual maximization. In fact, the complexity of the involved dynamic memory structures increases very fast (cf. Section 3.2 and figure 3 therein), so that problem instances with more than 10 scenarios cannot be handled so far. However, computing times can be improved if one takes into account a typical feature of real-life power generation systems. Often those systems comprise so-called *base load units*, which due to their specifications are usually scheduled on-line over the whole time horizon. Thus it makes sense to include them a priori in the set $\mathcal{I}_2$. For the considered VEAG-owned generation system we have identified six base load units. Our computational experience shows that this approach has almost no effect on the objective value, while the computing times improve. Details can be seen in Table 3. It is worth mentioning that the computing times increase with growing values for the price volatility σ. This is due to the fact that the respective scenarios become less and less compatible, since they favour different schedulings for the thermal units, which then leads to potential violations of the compatibility constraints.

Furthermore, we studied the relation between our original model (i.e., the general case of $\mathcal{I}_2 = \emptyset$) and the case that $\mathcal{I}_2 = \mathcal{I}$. To this end we created test instances that were successively solved for the two cases. Typical solutions for the general case exhibit both

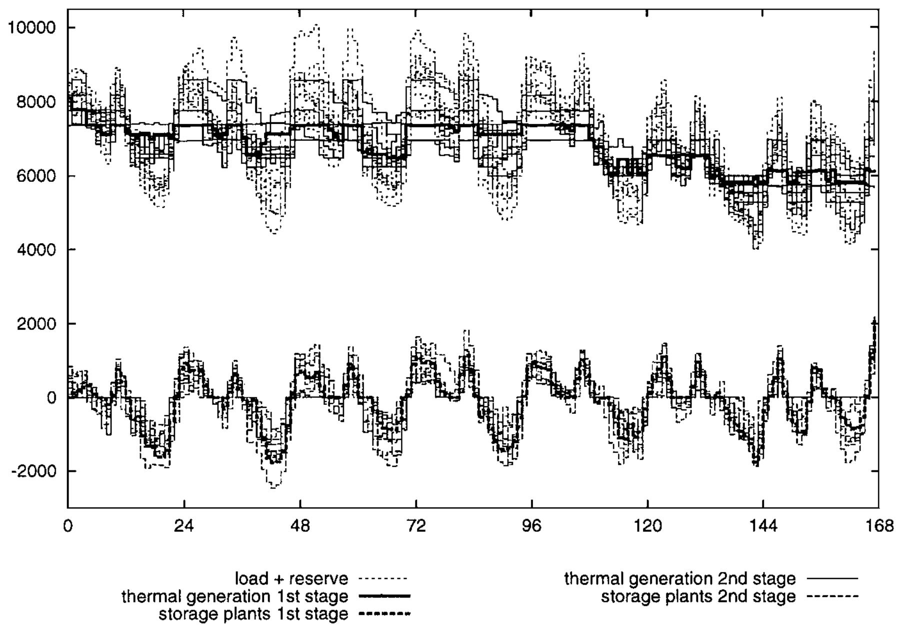

Figure 5. Solution for 10 scenarios and $\mathcal{I}_2 = \emptyset$.

switching on and off decisions at the transition from the first to the second stage. On the other hand, this scheduling behaviour is prevented for any unit $i \in \mathcal{I}_2$. Thus, a solution for the general case usually yields a better objective function value than the solution to the corresponding problem, where the index set $\mathcal{I}_2$ consists of all thermal units. Computational examples are shown in Table 4. Figure 6 gives a sample output for the instance that $\mathcal{I}_2 = \mathcal{I}$. It shows that there is little variance in the thermal output of the whole system compared to the general case, as can be seen in figure 5 for example.

Remark 2. We would like to note that in place of (16) one can also consider a weighted average of the costs in the two respective stages. More precisely, one could introduce a

Table 3. Effect of including base load units in $\mathcal{I}_2$.

	w/o base units		w base units		
Scenarios	Time	obj.val [$\times 10^8$]	Time	obj.val [$\times 10^8$]	Change (%)
5	1:23	1.38973	1:14	1.39177	0.147
5	1:21	1.34035	1:11	1.34138	0.077
10	4:47	1.42597	3:50	1.42589	−0.006
10	4:21	1.38813	3:55	1.38695	−0.085
10	4:18	1.38833	4:02	1.38863	0.022

(NOA 3.0: opt.tol = 10^{-4}, NGRAD = 20).

Table 4. Comparison between $\mathcal{I}_2 = \mathcal{I}$ and $\mathcal{I}_2 = \emptyset$.

	Objective value [$\times 10^8$]		
Scenarios	$\mathcal{I}_2 = \mathcal{I}$	$\mathcal{I}_2 = \emptyset$	Improvement (%)
5	1.16399	1.15511	0.76
5	1.50801	1.49352	0.96
10	1.17607	1.16349	1.07
10	1.44497	1.42486	1.39

(NOA 3.0: opt.tol = 10^{-4}, NGRAD = 20).

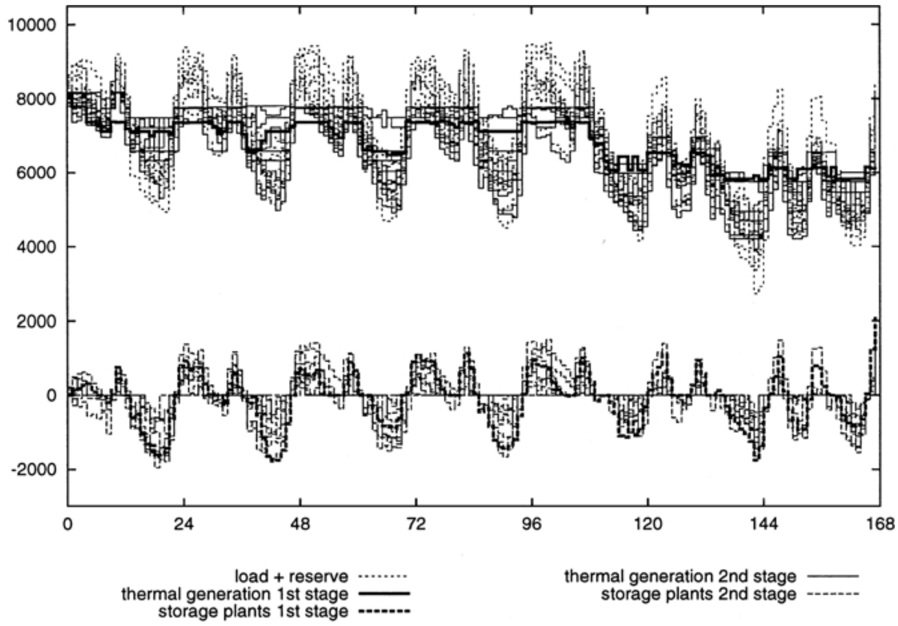

Figure 6. Solution for 10 scenarios and $\mathcal{I}_2 = I$.

weight coefficient $\theta \in [0, 1]$ and consider the convex combination

$$(1-\theta)\sum_{i=1}^{I}\sum_{t=1}^{T}[C_{it}(p_{it}, u_{it}) + S_{it}(u_i)] + \theta\mathbb{E}\sum_{i=1}^{I}\sum_{t=1}^{T}[\boldsymbol{C}_{it}(\mathbf{p}_{it}, \mathbf{u}_{it}) + \boldsymbol{S}_{it}(\mathbf{u}_i)] \tag{31}$$

as the objective function. The changes necessary to adapt the algorithm to (31) are minimal. The results presented in this section would then correspond to the case $\theta = \frac{1}{2}$.

Acknowledgments

The authors are grateful to G. Schwarzbach and J. Thomas (VEAG Vereinigte Energiewerke AG) for providing us with the data used in our test runs and to K. C. Kiwiel (Polish Academy of Sciences, Warsaw) for helpful comments and the permission to use the NOA 3.0 package. We extend our gratitude to Matthias P. Nowak (SINTEF, Trondheim) for valuable discussions and for making available parts of his software and to Isabel Wegner (Humboldt-University Berlin) for providing the electrical load scenarios for our test runs. Moreover, suggestions of the referees are gratefully acknowledged. This research was supported by the BMBF project *Langfristige Einsatzplanung in Kraftwerkssystemen* (SC7DU1).

References

S. Ahmed, A. J. King, and G. Parija, "A multi-stage stochastic integer programming approach for capacity expansion under uncertainty," *Journal of Global Optimization*, 2002 (to appear).

D. P. Bertsekas, G. S. Lauer, N. R. Sandell Jr., and T. A. Posbergh, "Optimal short-term scheduling of large-scale power systems," *IEEE Transactions on Automatic Control* vol. AC-28, pp. 1–11, 1983.

J. R. Birge, "Stochastic programming computations and applications," *INFORMS Journal on Computing* vol. 9, pp. 111–133, 1997.

C. C. Carøe and R. Schultz, "A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system," DFG-Schwerpunktprogramm *Echtzeit-Optimierung großer Systeme*, Preprint 98-13, 1998.

C. C. Carøe and R. Schultz, "Dual decomposition in stochastic integer programming," *Operations Research Letters* vol. 24, pp. 37–45, 1999.

P. Carpentier, G. Cohen, J.-C. Culioli, and A. Renaud, "Stochastic optimization of unit commitment: A new decomposition framework," *IEEE Transactions on Power Systems* vol. 11, pp. 1067–1073, 1996.

D. Dentcheva and W. Romisch, "Optimal power generation under uncertainty via stochastic programming," in *Stochastic Programming Methods and Technical Applications*, K. Marti and P. Kall, eds., Lecture Notes in Economics and Mathematical Systems, vol. 458, Springer-Verlag: Berlin, pp. 22–56, 1998.

S. Feltenmark and K. C. Kiwiel, "Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems," *SIAM Journal on Optimization* vol. 10, pp. 697–721, 2000.

R. Gollmer, A. Möller, M. P. Nowak, W. Römisch, and R. Schultz, "Primal and dual methods for unit commitment in a hydro-thermal power system," in *Proceedings 13th Power Systems Computation Conference*, Trondheim, Norway, vol. 2, pp 724–730, 1999.

N. Gröwe, W. Römisch, and R. Schultz, "A simple recourse model for power dispatch under uncertain demand," *Annals of Operations Research* vol. 59, pp. 135–164, 1995.

N. Gröwe-Kuska, K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner, "Power management in a hydro-thermal system under uncertainty by Langrangian relaxation," in *Decision Making under Uncertainty: Energy and Power*, C. Greengard and A. Rusz Czyúski, eds., IMA Volumes and its Applications, vol. 128, Springer-Verlag: New York, pp. 39–70, 2002.

K. C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable optimization," *Mathematical Programming* vol. 46, pp. 105–122, 1990.

K. C. Kiwiel, "User's Guide for NOA 2.0/3.0: A Fortran package for convex nondifferentiable optimization," Polish Academy of Sciences, Systems Research Institute, Warsaw, Poland, 1993/94.

C. Lemaréchal, "Lagrangian decomposition and nonsmooth optimization," in *Nonsmooth Optimization Methods and Applications*, F. Gianessi, ed., Gordon and Breach, pp. 201–216, 1992.

C. Lemaréchal and A. Renaud, "A geometric study of duality gaps, with applications," *Mathematical Programming* vol. 90, pp. 399–427, 2001.

M. P. Nowak, "Stochastic Lagrangian relaxation in power scheduling of a hydrothermal system under uncertainty," Ph.D. Thesis, Institut für Mathematik, Humboldt-Univ. Berlin, 2000.

M. P. Nowak, R. Nürnberg, W. Römisch, R. Schultz, and M. Westphalen, "Stochastic programming for power production and trading, in *Mathematics-Key Technology for the Future*, W. Jäger and H.-J. Krebs, eds., Springer, 2002 (to appear).

R. T. Rockafellar and R.J.-B. Wets, "Stochastic convex programming: Basic duality," *Pacific Journal of Mathematics* vol. 62, pp. 173–195, 1976.

A. Ruszczyński, "Decomposition methods in stochastic programming," *Mathematical Programming* B vol. 79, pp. 333–353, 1997.

G. B. Sheble and G. N. Fahd, "Unit commitment literature synopsis," *IEEE Transactions on Power Systems* vol. 9, pp. 128–135, 1994.

J. M. Swaminathan, "Tool capacity planning for semiconductor fabrication facilities under demand uncertainty," *European Journal of Operational Research* vol. 120, pp. 545–558, 2000.

S. Takriti and J. R. Birge, "Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs," *Operations Research* vol. 48, pp. 91–98, 2000.

S. Takriti, B. Krasenbrink, and L.S.-Y. Wu, "Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem," *Operations Research* vol. 48, pp. 268–280, 2000.

F. Zhuang and F. D. Galiana, "Towards a more rigorous and practical unit commitment by Lagrangian relaxation," *IEEE Transactions on Power Systems* vol. 3, pp. 763–773, 1988.