

Stochastic unit commitment in hydro-thermal power production planning

N. Gröwe-Kuska and W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

Abstract

We present a mixed-integer multistage stochastic programming model for the short term unit commitment of a hydro-thermal power system under uncertainty in load, inflow to reservoirs, and prices for fuel and delivery contracts. The model is implemented for uncertain load and tested on realistic data from a German power utility. Load scenario trees are generated by a procedure consisting of two steps: (i) Simulation of load scenarios using an explicit representation of the load distribution and (ii) construction of a tree out of these scenarios. The dimension of the corresponding mixed-integer programs ranges up to 200,000 binary and 350,000 continuous variables. The model is solved by a Lagrangian-based decomposition strategy exploiting the loose coupling structure. Solving the Lagrangian dual by a proximal bundle method leads to a successive decomposition into single unit subproblems, which are solved by specific algorithms. Finally, Lagrangian heuristics are used to construct nearly optimal first stage decisions.

Key Words: Stochastic programming, unit commitment, Lagrangian relaxation, bundle methods, scenario tree generation

1991 MSC: 90C15, 90C90, 90C11

1 Introduction

Economic needs and the ongoing liberalization of European electricity markets stimulate the interest of power utilities in developing models and optimization techniques for the generation and trading of electric power under uncertainty. Utilities participating in deregulated markets observe increasing uncertainty in load (i.e., demand for electric power) and prices for fuel and electricity on spot and contract markets. The mismatched power between actual and predicted demand may be supplied by the power system or by trading activities. The competitive environment forces the utilities to rate alternatives within a few minutes.

In this chapter, we describe a mathematical model for optimal short term operation and trading of a hydro-thermal based electric utility, which is usually called *unit commitment problem* because of the important role of the commitment or on/off decisions. Furthermore, we present a methodology for modelling the stochastic data process in

form of a scenario tree and report on a Lagrangian-based decomposition strategy for solving the optimization model. We also provide some numerical experience obtained from test runs on realistic data from the German utility Vereinigte Energiewerke AG (VEAG). In fact, the optimization model has emerged from a collaboration with engineers of VEAG. For our tests we use a configuration of the VEAG system consisting of 25 (coal-fired, gas-burning) thermal units and 7 pumped storage hydro units. Its total capacity is about 13,000 megawatts (MW), including a hydro capacity of 1,700 MW; the peak loads of the system are about 8,600 MW. In contrast to other hydro-thermal based utilities the amount of installed pumped storage capacity enables the inclusion of pumped storage plants into the optimization. It is an additional feature of the VEAG system that, for a weekly planning period, inflows to reservoirs are negligible.

In the literature, there is a growing number of contributions to stochastic power system optimization with emphasis on modelling aspects and solution methods. Stochastic models including commitment decisions are considered, for example, in [1, 2, 6, 8, 9, 10, 19, 24, 26, 27, 28, 34, 35, 36, 37]. While the recent papers [1, 19, 26, 37] propose variants of Lagrangian relaxation methods for their solution and present implementations, the work in [2, 24, 34] is directed to a comparison of stochastic and deterministic power system modelling, further engineering aspects and industrial applications. The solution methods in [1, 19, 37] differ from each other by the nondifferentiable optimization methods, the subproblem solvers, and the Lagrangian heuristics employed as components of the master algorithm. Modelling issues of stochastic data processes in power systems are addressed in [16, 18, 20, 28]. We also refer to the state-of-the-art survey [12] on scenario (tree) generation and to references therein.

The chapter is organized as follows. In Section 2 we describe the stochastic unit commitment model and its particular features. Section 3 contains a brief description of the solution algorithm based on Lagrangian relaxation. Our strategy for generating scenario trees for the electrical load process is presented in Section 4. It consists of two parts: simulation of load scenarios using a statistical model for the electrical load process and a method for constructing scenario trees out of simulation scenarios. In Section 5 we report on the performance of the Lagrangian relaxation algorithm and of the scenario tree generation technique.

2 Stochastic Power System Modeling

We consider a power generation system comprising thermal and hydro units and contracts for delivery and purchase, and address the unit commitment problem in short term operation planning. This problem concerns the scheduling of start-up/shut-down decisions and of operation levels for all power units and contracts, respectively, such that the operation costs over the time horizon are minimal. Although being short term, the uncertainty of important system parameters like electrical load, streamflows in hydro units, prices for fuel or electricity is a major modelling issue.

Let the planning horizon be discretized into T uniform subintervals and suppose there are sets \mathcal{I} and \mathcal{J} of thermal and hydro units, respectively. The decision variable of thermal unit $i \in \mathcal{I}$ is (u_i, p_i) where the components of u_i are binary variables taking the values 1 if the unit is on at some time period and 0 if off. The components of p_i are

the corresponding operation levels. As usual, contracts for delivery and purchase are regarded as special thermal units. The decision variable of hydro unit $j \in \mathcal{J}$ is (v_j, w_j) , where the components of v_j and w_j are the generation and pumping levels over time, respectively.

To formulate a unit commitment model that incorporates fluctuations of uncertain system parameters, we use a probabilistic description of uncertainty. Let

$$\xi = \{ \xi_t := (d_t, r_t, \gamma_t, a_t, b_t, c_t) \}_{t=1}^T$$

be a discrete-time stochastic process on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where ξ_1 is deterministic, d_t , r_t and γ_t represent the load, the spinning reserve and the hydro inflows in period t , while a_t , b_t and c_t collect the cost coefficients.

The scheduling decisions for period t are made *after* having learnt the realization of the stochastic data for that period. Denote by $\mathcal{F}_t \subseteq \mathcal{F}$ the σ -field generated by $\{\xi_\tau\}_{\tau=1}^t$, i.e., the events observable until period t . Since the information on ξ_1 is complete, $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., ξ_1 is deterministic. By assuming $\mathcal{F}_T = \mathcal{F}$ we require full information to be available at the end of the planning horizon. The sequence of scheduling decisions $\{(u_t, p_t, v_t, w_t)\}_{t=1}^T$ also forms a stochastic process on $(\Omega, \mathcal{F}, \mathcal{P})$, which is assumed to be adapted to the filtration of σ -fields, i.e., *nonanticipative*. Nonanticipativity means that the decisions (u_t, p_t, v_t, w_t) may only depend on the data observable until period t , or equivalently, that (u_t, p_t, v_t, w_t) is \mathcal{F}_t -measurable.

We now assume that the data process $\{\xi_t\}_{t=1}^T$ has a *discrete* probability distribution, i.e., its support consists of a finite number of *scenarios* (or realizations). Then there exist finite subsets \mathcal{E}_t of the σ -algebra \mathcal{F}_t , $t = 1, \dots, T$, such that \mathcal{E}_t is a partition of Ω and that the smallest σ -algebra containing \mathcal{E}_t is just \mathcal{F}_t . Using conditional expectations w.r.t. \mathcal{F}_t , the nonanticipativity conditions may be formulated as linear equality constraints. As $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, every element of \mathcal{E}_t can be represented as the union of certain elements of \mathcal{E}_{t+1} . Since the numbers of elements of \mathcal{E}_t , i.e., $|\mathcal{E}_t|$, and of scenarios of ξ_t coincide, the relations between the elements of \mathcal{E}_t and of \mathcal{E}_{t+1} for $t = 1, \dots, T-1$ may be represented in the form of a tree, called *scenario tree*. Let $\mathcal{N} = \{1, \dots, |\mathcal{N}|\}$ denote the set of nodes of the tree. The *root* node $n = 1$ stands for period $t = 1$. Every other node n has a unique *predecessor* node n_- . Let $\text{path}(n)$ be the set $\{1, \dots, n_-, n\}$ of nodes

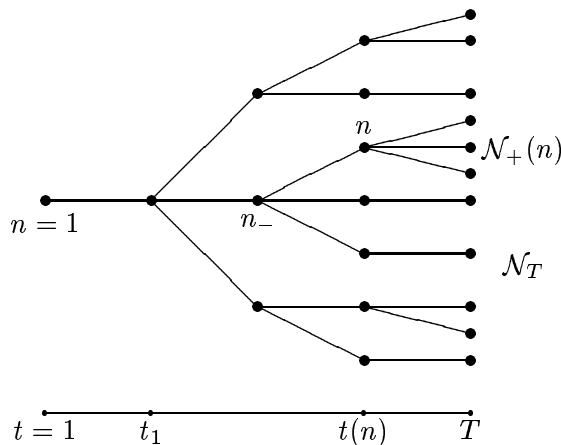


Figure 1: Scenario tree with $t_1 = 2$, $T = 5$, $|\mathcal{N}| = 23$ and 11 leaves

from the root to node n and $t(n) := |\text{path}(n)|$ denote the time period associated with node n . Then the nodes in $\mathcal{N}_t := \{n : t(n) = t\}$ correspond to the realizations of ξ_t , $t = 1, \dots, T$. Nodes n belonging to the set \mathcal{N}_T are called *leaves*. A scenario corresponds to a path from the root to some leaf, i.e., to $\text{path}(n)$ for some $n \in \mathcal{N}_T$. Furthermore, let $\mathcal{N}_+(n)$ denote the set of *successors* to node n . Hence, $\mathcal{N}_T = \{n : \mathcal{N}_+(n) = \emptyset\}$. With the given scenario probabilities $\{\pi_n\}_{n \in \mathcal{N}_T}$, we associate a probability π_n to each node n by the recursion $\pi_n := \sum_{n_+ \in \mathcal{N}_+(n)} \pi_{n_+}$, $n \in \mathcal{N}$. Clearly, $\sum_{n \in \mathcal{N}_t} \pi_n = 1$ holds for each $t = 1, \dots, T$. We denote by $\mathcal{N}_{\text{first}} := \cup_{t=1}^{t_1} \mathcal{N}_t$ the set of *first-stage* nodes, where t_1 is the maximal period such that the data process $\{\xi_t\}_{t=1}^{t_1}$ is deterministic, i.e., the sets \mathcal{N}_t , $t = 1, \dots, t_1$, are singletons. We use the following notation for the sequence of predecessors of any node $n \in \mathcal{N}$: $n_0 := n$, $n_{-1} := n_-$ if $n > 1$, $n_{-(\kappa+1)} := (n_{-\kappa})_-$ if $t(\kappa) > 1$.

Now, we use the notations $\{\xi^n = (d^n, r^n, \gamma^n, a^n, b^n, c^n)\}_{n \in \mathcal{N}}$ and $\{(u^n, p^n, v^n, w^n)\}_{n \in \mathcal{N}}$ for the scenario trees representing the stochastic data process ξ and the (stochastic) decision process (u, p, v, w) , respectively. The decisions (u^n, p^n, v^n, w^n) assigned to nodes n in \mathcal{N}_t are the realizations of the stochastic decisions (u_t, p_t, v_t, w_t) for $t = 1, \dots, T$. For the commitment decisions u_i we set $u_i^{\text{path}(n)} := (u_i^\nu)_{\nu \in \text{path}(n)}$. To handle initial values of the u_i we introduce a starting time $t_{\text{ini}} \leq 0$, set $n_\kappa := \kappa - t(n)$ for $\kappa = t(n) + t_{\text{ini}}, \dots, t(n)$, and assume that initial values $u_i^{1_\kappa}$ for $\kappa = t_{\text{ini}}, \dots, 0$, $i \in \mathcal{I}$ are given. The *fuel costs* for operating the thermal unit i at node n are of the form

$$C_i^n(p_i^n, u_i^n) := \max_{l=1, \dots, \bar{l}} \{a_{il}^n p_i^n + b_{il}^n u_i^n\}$$

with coefficients a_{il}^n and b_{il}^n such that $C_i^n(\cdot, 1)$ is convex and increasing on \mathbb{R}_+ . The *start-up costs* of unit i at node n depend on its downtime; it may vary from a maximum cold-start value to a much smaller value when the unit is still relatively close to its operating temperature. This is modelled by

$$S_i^n(u_i^{\text{path}(n)}) := \max_{\tau=0, \dots, t_i^c} c_{i\tau}^n \left(u_i^n - \sum_{\kappa=1}^{\tau} u_i^{n-\kappa} \right),$$

where $0 < c_{i0}^n < \dots < c_{it_i^c}^n$ are cost coefficients, t_i^c is the cool-down time and $c_{it_i^c}^n$ the maximum cold-start costs of unit $i \in \mathcal{I}$. Since the operating costs of hydro units are negligible in short term planning, the expected total system costs are given by the sum of fuel and start-up costs of all thermal units

$$\sum_{n \in \mathcal{N}} \pi_n \sum_{i \in \mathcal{I}} \left[C_i^n(p_i^n, u_i^n) + S_i^n(u_i^{\text{path}(n)}) \right]. \quad (1)$$

The operation of all thermal units is described by certain operating ranges and minimum up/down-time requirements, namely, by the inequalities

$$p_{it(n)}^{\min} u_i^n \leq p_i^n \leq p_{it(n)}^{\max} u_i^n, \quad u_i^n \in \{0, 1\}, \quad n \in \mathcal{N}, \quad i \in \mathcal{I}, \quad (2a)$$

$$u_i^{n-\kappa} - u_i^{n-(\kappa+1)} \leq u_i^n, \quad \kappa = 1, \dots, \bar{t}_i - 1, \quad n \in \mathcal{N}, \quad i \in \mathcal{I}, \quad (2b)$$

$$u_i^{n-(\kappa+1)} - u_i^{n-\kappa} \leq 1 - u_i^n, \quad \kappa = 1, \dots, \underline{t}_i - 1, \quad n \in \mathcal{N}, \quad i \in \mathcal{I}, \quad (2c)$$

where p_{it}^{\min} and p_{it}^{\max} are the minimum and maximum capacities of unit i at period t , and (2b), (2c) mean that unit i must remain on (off) for at least \bar{t}_i (and \underline{t}_i , respectively)

periods if it is switched on (off). The operating ranges and dynamics of hydro units are described by the constraints

$$0 \leq v_j^n \leq v_{jt(n)}^{\max}, 0 \leq w_j^n \leq w_{jt(n)}^{\max}, 0 \leq l_j^n \leq l_{jt(n)}^{\max}, n \in \mathcal{N}, j \in \mathcal{J}, \quad (3a)$$

$$l_j^n = l_j^{n-} - v_j^n + \eta_j w_j^n + \gamma_j^n, \quad n \in \mathcal{N}, j \in \mathcal{J}, \quad (3b)$$

$$l_j^0 = l_j^{\text{in}}, \quad l_j^n = l_j^{\text{end}}, \quad n \in \mathcal{N}_T, j \in \mathcal{J}, \quad (3c)$$

where v_{jt}^{\max} and w_{jt}^{\max} are the maximum capacities for the generation and pumping of hydro unit $j \in \mathcal{J}$ at period t , γ_j^n is the water inflow to reservoir j at node n and l_j^n is the reservoir *storage volume* at the end of period $t(n)$, with upper bound $l_{jt(n)}^{\max}$. By η_j we denote the *pumping efficiency* and by l_j^{in} and l_j^{end} the initial and final volumes, respectively, of unit j . In our model, we disregard spill and head variation effects. Furthermore, we prefer to prescribe final storage volumes of all hydro units instead of introducing a water value function depending on $\{l_j^n\}_{j \in \mathcal{J}, n \in \mathcal{N}_T}$ and of maximizing this water value as part of the objective function (1). The basic system requirements are the load and the spinning reserve constraint at each node n , i.e.,

$$\sum_{i \in \mathcal{I}} p_i^n + \sum_{j \in \mathcal{J}} (v_j^n - w_j^n) \geq d^n, \quad n \in \mathcal{N}, \quad (4a)$$

$$\sum_{i \in \mathcal{I}} (u_i^n p_{it(n)}^{\max} - p_i^n) \geq r^n, \quad n \in \mathcal{N}, \quad (4b)$$

where the constraint (4b) means that the total committed capacity at each node n should exceed d^n by a certain amount r^n , e.g., by a fraction of d^n . Figure 2 shows

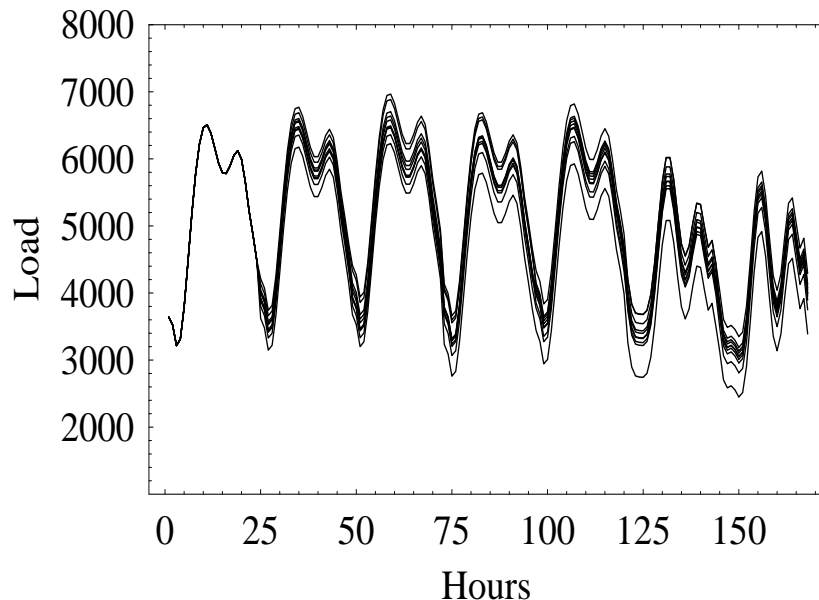


Figure 2: Typical weekly load scenarios

a collection of weekly load scenarios. They exhibit typical daily cycles, morning and evening peaks, and night and weekend off-peaks. Constraint (4a) says that the total generation has to follow the load scenario curves.

Now, the *stochastic unit commitment problem* consists in minimizing the expected costs (1) such that the system constraints (2), (3) and (4) are satisfied. This model forms a large scale linear mixed-integer program involving $|\mathcal{I}||\mathcal{N}|$ binary and $(|\mathcal{I}| + 2|\mathcal{J}|)|\mathcal{N}|$ continuous decision variables, and $(2 + |\mathcal{J}|)|\mathcal{N}| + |\mathcal{J}||\mathcal{N}_T|$ (in)equality constraints (without taking into account the bounds, the constraints of type (2b)–(2c), and the objective function). As usual, we call the model (1)–(4) with fixed binary decisions $\{u^n\}_{n \in \mathcal{N}}$ the *stochastic economic dispatch problem*. We notice that the model (1)–(4) is almost separable with respect to the unit index sets \mathcal{I} and \mathcal{J} and only loosely coupled by the $2|\mathcal{N}|$ constraints (4). This observation becomes particularly important when recalling that, for a typical weekly time horizon with hourly time steps (i.e., $T = 168$), scenario trees comprise more than $|\mathcal{N}| = 10^3$ nodes and that for mid-size generation systems one has $|\mathcal{I}| + |\mathcal{J}| \geq 30$.

3 Lagrangian Relaxation

Due to the enormous size of the stochastic unit commitment model in the previous section, the use of standard software for mixed-integer linear programs is appropriate for smaller models only. In general, one has to resort to *decomposition* approaches. While the algorithmic realization of *primal* decomposition methods leads to serious obstacles that are impossible to overcome by existing methods (cf. [5]), strategies based on dualizing the model (1)–(4) appear to be more promising.

Two general *dual* decomposition schemes have been elaborated so far: *scenario* and *nodal* decomposition (see [7, 31]). The first scheme is based on dualizing the nonanticipativity constraints and the second one on the dualization of the dynamic constraints. Due to the loose coupling structure of the model, a third scheme comes into play, which is called *geographical* or *component* decomposition in [11, 31]. It is based on assigning Lagrange multipliers to the coupling constraints and on minimizing the corresponding Lagrangian function. As a result, the dual problem decomposes into a finite number of (much) smaller stochastic subproblems. The application of all these dualization schemes is justified in convex situations (cf. [30] and [31, Section 2.5]).

When comparing dual decomposition approaches for stochastic integer programming models, two arguments appear to be important: the size of the corresponding duality gaps and the complexity of the dual and of the subproblems. Recent results on a comparison of duality gaps for scenario, nodal and geographic decomposition indicate that the geographic decomposition has the potential to lead to the smallest duality gaps (see [11]). In addition, in the case of the stochastic unit commitment model its subproblems have specific structures that allow for the use of efficient solution algorithms.

In the following, we give a brief description of the geographical decomposition or *Lagrangian relaxation* approach and of a Lagrangian-based algorithm for solving (1)–(4). For a more detailed presentation we refer to [19]. Let $x := (u, p, v, w)$ denote the decision and $\lambda = (\lambda_1, \lambda_2) := \{(\lambda_1^n, \lambda_2^n)\}_{n \in \mathcal{N}} \in \Lambda := \mathbb{R}_+^{|\mathcal{N}|} \times \mathbb{R}_+^{|\mathcal{N}|}$ the Lagrange multiplier in scenario-tree form to be associated with the coupling constraints (4). Then the *Lagrangian* function takes the form

$$L(x; \lambda) := \sum_{n \in \mathcal{N}} \pi_n \left\{ \sum_{i \in \mathcal{I}} [C_i^n(p_i^n, u_i^n) + S_i^n(u_i^{\text{path}(n)})] \right\} \quad (5)$$

$$+ \lambda_1^n \left[d^n - \sum_{i \in \mathcal{I}} p_i^n - \sum_{j \in \mathcal{J}} (v_j^n - w_j^n) \right] + \lambda_2^n \left[r^n - \sum_{i \in \mathcal{I}} (u_i^n p_{it(n)}^{\max} - p_i^n) \right] \Bigg\},$$

and the *dual function* and the *dual problem* read

$$D(\lambda) := \min \{L(x; \lambda) : x \text{ satisfies the constraints (2)–(3)}\} \quad (6)$$

$$\max \{D(\lambda) : \lambda \in \Lambda\}. \quad (7)$$

Due to the compactness of the constraint sets given by (2) and (3) for (u, p) and (v, w) , respectively, there exists a Lagrangian solution $x(\lambda)$, i.e., a solution of the minimization problem defining $D(\lambda)$ in (6) for every $\lambda \in \Lambda$. Under the assumptions made on the fuel costs, the dual function D is concave polyhedral. Hence, the dual (7) is solvable if the primal problem (1)–(4) is feasible. The dual function

$$D(\lambda) = \sum_{i \in \mathcal{I}} D_i(\lambda) + \sum_{j \in \mathcal{J}} \hat{D}_j(\lambda_1) + \sum_{n \in \mathcal{N}} \pi_n (\lambda_1^n d^n + \lambda_2^n r^n) \quad (8)$$

decomposes into the *thermal subproblems*

$$D_i(\lambda) = \min \left\{ \sum_{n \in \mathcal{N}} \pi_n \left[\min_{p_i^n} \{C_i^n(p_i^n, u_i^n) - (\lambda_1^n - \lambda_2^n) p_i^n\} \right. \right. \\ \left. \left. - \lambda_2^n u_i^n p_{it(n)}^{\max} + S_i^n(u_i^{\text{path}(n)}) \right] : u_i \text{ satisfies (2)} \right\}, \quad (9)$$

and the *hydro subproblems*

$$\hat{D}_j(\lambda_1) = \min \left\{ \sum_{n \in \mathcal{N}} \pi_n \lambda_1^n (w_j^n - v_j^n) : (v_j, w_j) \text{ satisfies (3)} \right\}. \quad (10)$$

Both subproblems represent multistage stochastic programming models for the operation of one single unit. While the thermal subproblem (9) represents a combinatorial multistage program involving stochastic costs, the hydro subproblem (10) is a linear multistage model with stochastic costs and stochastic right-hand sides. The thermal problem (9) was solved by (stochastic) dynamic programming. Incorporating the thermal state space into the scenario tree leads to a backward tree recursion for the cost-to-go of all states. This yields the optimal cost-to-go and, by forward tracing the tree, the optimal scheduling decisions $\{(u_i^n(\lambda), p_i^n(\lambda))\}_{n \in \mathcal{N}}$ (see [19, Section 3.6] and [25]). For solving the hydro subproblems (10) a specialized descent method has been developed. It generates a finite sequence of feasible hydro decisions, where the decision at some step differs from the preceding one only at the nodes of a certain subtree of \mathcal{N} . Such a subtree exists for each nonoptimal feasible hydro decision and, hence, the algorithm terminates with an optimal solution $\{(v_j^n(\lambda), w_j^n(\lambda))\}_{n \in \mathcal{N}}$ (see [19, Section 3.5] and [25, 26] for details and numerical results).

The dual problem (7) serves as the *master* program. Its iterative solution by a subgradient bundle method leads to a successive decomposition of the primal model (1)–(4). A subgradient of D at λ is given by

$$g_D(\lambda) = \left\{ \left(d^n - \sum_{i \in \mathcal{I}} p_i^n(\lambda) - \sum_{j \in \mathcal{J}} (v_j^n(\lambda) - w_j^n(\lambda)), r^n - \sum_{i \in \mathcal{I}} (u_i^n(\lambda) p_{it(n)}^{\max} - p_i^n(\lambda)) \right) \right\}_{n \in \mathcal{N}}.$$

In particular, the *proximal bundle method* [15, 22, 23] is used for solving the dual. Starting from an arbitrary point $\lambda^1 = \bar{\lambda}^1 \in \Lambda$, this method generates a sequence $\{\lambda^k\}_{k \in \mathcal{N}}$ in Λ that converges to some dual solution, and trial points $\bar{\lambda}^k$ for evaluating the solutions $x(\bar{\lambda}^k)$ of (6), the subgradients $g_D(\bar{\lambda}^k)$ of D and its linearizations

$$D^k(\cdot) := D(\bar{\lambda}^k) + \langle \cdot - \bar{\lambda}^k, g_D(\bar{\lambda}^k) \rangle \geq D(\cdot),$$

where $\langle \lambda, \mu \rangle := \sum_{n \in \mathcal{N}} \pi_n (\lambda_1^n \mu_1^n + \lambda_2^n \mu_2^n)$ is the dual pairing on Λ . Iteration k uses the polyhedral model $D_k(\cdot) := \min_{l \in N^k} D^l(\cdot)$ with $k \in N^k \subset \{1, \dots, k\}$ for finding the next trial point $\bar{\lambda}^{k+1}$ as a solution of the quadratic subproblem

$$\max \left\{ D_k(\lambda) - \frac{1}{2} \rho_k |\lambda - \lambda^k|^2 : \lambda \in \Lambda \right\}, \quad (11)$$

where the proximity weight $\rho_k > 0$ and the penalty term $|\cdot|^2 := \langle \cdot, \cdot \rangle$ should keep $\bar{\lambda}^{k+1}$ close to the prox-center λ^k . An ascent step to $\lambda^{k+1} = \bar{\lambda}^{k+1}$ occurs if $D(\bar{\lambda}^{k+1}) \geq D(\lambda^k) + \kappa \delta_k$, where $\kappa \in (0, 1)$ is a fixed Armijo-like parameter and $\delta_k := D_k(\bar{\lambda}^{k+1}) - D(\lambda^k) \geq 0$ is the predicted ascent (if $\delta_k = 0$, then λ^k is a solution and the method may stop). Otherwise, a null step $\lambda^{k+1} = \lambda^k$ improves the next model D_{k+1} with the new linearization D^{k+1} . The stopping criterion $\delta_k \leq \text{opt_tol}(1 + D(\lambda^k))$, the choices of the weights ρ_k and of the index set N^{k+1} , in particular its upper bound NGRAD, are discussed in [15, 22] (see also [19, Section 3.4]).

The optimal value $D(\lambda^*)$ of (7) resulting from the bundle method provides a lower bound for the optimal costs of the model (1)–(4). In general, however, the “dual optimal” scheduling decisions $x(\lambda^*) = (u(\lambda^*), p(\lambda^*), v(\lambda^*), w(\lambda^*))$ violate the load and reserve constraints (4) such that a low-cost primal feasible solution has to be determined by a *Lagrangian heuristic*. Two Lagrangian heuristics (see [19, Section 3.7], [25]) that determine nearly optimal first stage decisions $\{(u^n, p^n, v^n, w^n)\}_{n \in \mathcal{N}_{\text{first}}}$ starting from the optimal multiplier λ^* and the Lagrangian solution $x(\lambda^*)$ have been developed. The first heuristic LH1 is based on a combination of a water rescheduling procedure and a known thermal heuristic [39] applied to a (deterministic) unit commitment model where the stochastic quantities ξ , λ^* and $l(\lambda^*)$ are replaced by their mean values. Clearly, this heuristic provides a nearly optimal decision at nodes $n \in \mathcal{N}_{\text{first}}$ only. The second heuristic LH2 starts by finding some $\varepsilon > 0$ such that $x(\lambda^* + \varepsilon \underline{1})$ ($\underline{1}$ being the element in Λ with unit components) is feasible. Taking $u(\lambda^* + \varepsilon \underline{1})$ as a starting point then, a finite sequence of binary decisions is constructed such that their components are decreasing. This is done by selecting a node $n \in \mathcal{N}$ where the available reserve capacity $\sum_{i=1}^I (u_i^n p_{it(n)}^{\max} - p_i^n) - r^n$ is maximal, and switching some unit i off at node n and at some predecessor and successor nodes, where the unit i and the neighbouring nodes of n are detected by stochastic dynamic programming. Next, the corresponding stochastic economic dispatch problem is reformulated as a hydro problem with piecewise linear costs and solved by a modification of the descent method mentioned earlier (see [19, Section 3.5], [25, 26]). This procedure, which generates a sequence of scheduling decisions at all nodes, is continued until infeasibility is detected during economic dispatch and terminates with a nearly optimal solution at each node in \mathcal{N} .

The whole Lagrangian relaxation approach is based on the same, but *stochastic*, ingredients as in the classical deterministic unit commitment situation [17, 32]: a solver for the nondifferentiable dual, subproblem solvers, and some Lagrangian heuristic. The interaction of these ingredients is illustrated in Figure 3.

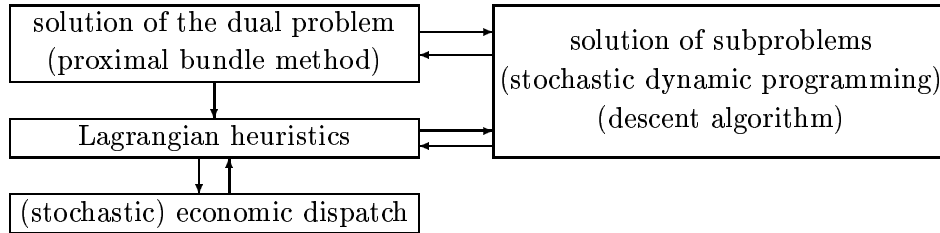


Figure 3: Scheme of the Lagrangian relaxation algorithm

4 Load Scenario Trees

4.1 A Statistical Model for the Electrical Load

The identification of a statistical model for the electric load of the VEAG generation system is based on an hourly load profile for a period of three years (1098 days). A plot of the hourly load data is displayed in Figure 4. The historical load records show seasonal variations caused by meteorological factors like temperature, cloud cover etc. In the weekly and monthly load data there are recurring patterns of length 24 (one day) and of length 168 (one week). The periodic patterns complete themselves within the calendar year and are then repeated on a yearly basis. Interruptions of this regularity are caused by customs like public holidays or the start/end of the daylight saving time. Thus, in principle the electric load depends on the category of the day (Monday, ..., Sunday, public holiday, etc.) and on the season. Figure 4 highlights the periodic components of our historical data.

In a first step, days of a similar load pattern are identified using daily load records (24 load data of a day). To each such record we assign a day category (1 for a Monday record, ..., 7 for a Sunday, 8 for a public holiday following a working day, 9 for days between holidays and weekends, 10 for a public holiday following a weekend or a holiday). Clustering methods from [33] are applied to answer the question whether the records can be grouped or classified into useful or informative clusters. After eliminating seasonal effects of the load records, clustering and ANOVA-tests lead to a classification of the load records into 8 categories (see Table 1).

The statistical modelling of the load process exploits the decomposition of the load process into a *daily mean load process* and a *mean-corrected load series* which are

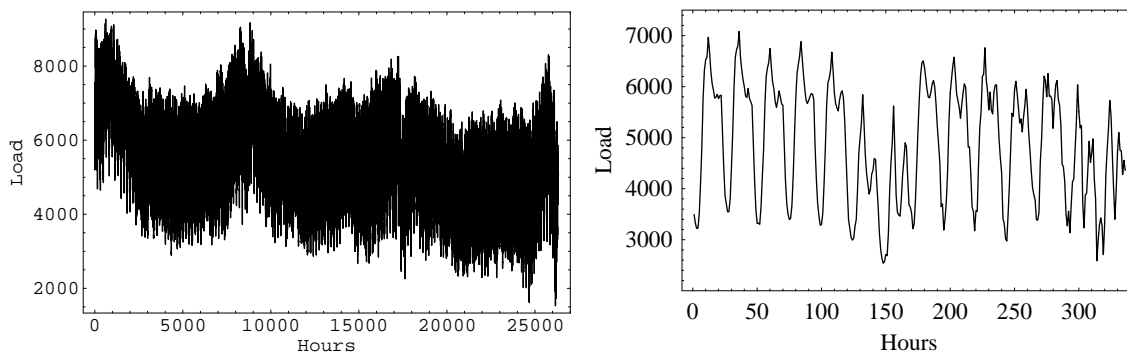


Figure 4: Hourly load data – three years (left) and two weeks (right)

Category	
1	Monday or working day after a public holiday
2	working day (Tuesday, Wednesday, Thursday)
3	Friday or working day before a public holiday
4	Saturday
5	Sunday
6	public holiday not following days of the categories 2,3
7	public holiday following days of the categories 2,3
8	working day between days of the categories 4-7

Table 1: Categories of daily load records

treated separately. Let $x_{j\tau}$ be the observed load at time period $\tau = 1, \dots, 24$ of day $j \in J := \{1, \dots, 1098\}$ (i.e., record j of the data base), $\bar{d}_j := \frac{1}{24} \sum_{\tau=1}^{24} x_{j\tau}$ the mean load of day j and $\text{cat}(j)$ the category of day j according to Table 1. Then the historical load records are decomposed according to

$$x_{j\tau} = d_{j\tau} + \bar{d}_j \quad (\tau = 1, \dots, 24; j \in J), \quad (12)$$

where $d_{j\tau}$, $\tau = 1, \dots, 24$, is the mean corrected load record of day $j \in J$. The daily mean load series versus the day number is plotted in Figure 5. Figure 6 displays the

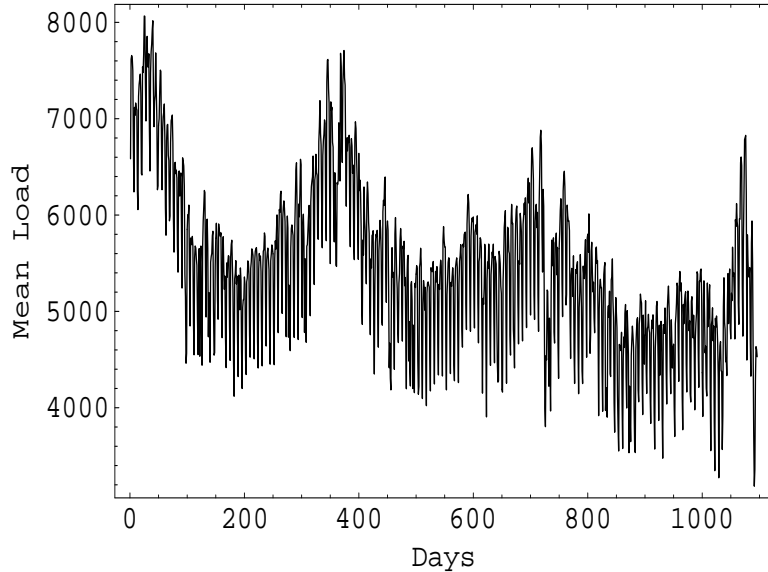


Figure 5: Daily mean load versus the day number

mean-corrected load records $(d_{j\tau})_{\text{cat}(j)=k}$ for days of category $k = 2$ and $k = 5$. The mean load depends on the category of the day and on the season. Further, there is an interaction between the mean load and meteorological factors like temperature, cloud cover etc. The meteorological impact on the daily mean demand could not be modelled because of missing meteorological parameters.

To select an appropriate class of (time series) models for the daily mean load series $\{\bar{d}_j\}_{j \in J}$ with $J \subset \mathbf{Z}$, $\{\bar{d}_j\}_{j \in J}$ is considered as a part of a realization of the stochastic

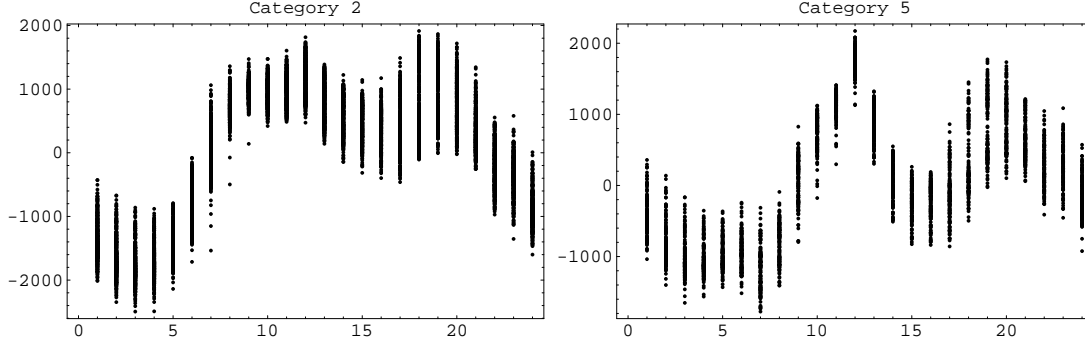


Figure 6: Mean-corrected load records for days of category 2 (left) and 5 (right)

mean load process $\{\bar{d}_j\}_{j \in \mathbf{Z}}$. Data analysis methods [3] are used to detect any *seasonal* (periodic) or *trend* (nonconstant mean) components, outlying observations or sharp changes in behaviour. Then suitable transformations are applied to the data to obtain a new stationary series (*residuals*) with zero mean and unit variance. The trend and seasonal components may be removed by estimating these components and subtracting them from the data. Another transformation is called *differencing*; it replaces the original process by differences of the process at t and at $t - s$ for some lag $s \in \mathbb{N}$ and eliminates a seasonal component of period s . The mean load series $\{\bar{d}_j\}_{j \in \mathbf{J}}$ clearly contains a recurring pattern with the seasonal period of 365 (one year). There are further periodic components of length 7 (one week) and change points due to the start/end of the daylight saving time. Irregularities of the weekly patterns have been removed from the time series by replacing outlying observations by the value of the nearest day of the same category.

Many approaches for fitting a time series to the deseasonalized data rely on classical linear models. *Autoregressive moving average* (ARMA) models are characterized by finite-order linear difference equations with constant coefficients. A real stochastic process $\{X_t\}_{t \in \mathbf{Z}}$ is called *ARMA*(p, q) if it is stationary, i.e., if $\mathbb{E}[X_t^2] < \infty$, $\mathbb{E}[X_t]$ is constant and $\mathbb{E}[X_r X_s] = \mathbb{E}[X_{r+t} X_{s+t}]$, $\forall r, s, t \in \mathbf{Z}$, and

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad \forall t \in \mathbf{Z}, \quad (13)$$

where ϕ_k , $k = 1, \dots, p$, and θ_l , $l = 1, \dots, q$, are real coefficients and $\{Z_t\}_{t \in \mathbf{Z}}$ is the *white noise* process $\text{WN}(0, \sigma^2)$ with zero mean and variance σ^2 , i.e., $\mathbb{E}[Z_t] = 0$, $\mathbb{E}[Z_t^2] = \sigma^2$, $\forall t \in \mathbf{Z}$, and $\mathbb{E}[Z_r Z_t] = 0$ if $r \neq t$. Using the *backward shift operator* B defined by $B^\ell X_t := X_{t-\ell}$ for $t, \ell \in \mathbf{Z}$, the ARMA equations (13) can be rewritten as

$$\phi(B)X_t = \theta(B)Z_t, \quad \forall t \in \mathbf{Z}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

where ϕ and θ denote the polynomials $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. An ARMA(p, q) process $\{X_t\}_{t \in \mathbf{Z}}$ is said to be *causal* if there exists a real sequence $\{\psi_\ell\}$ such that $\sum_{\ell=0}^{\infty} \psi_\ell < \infty$ and $X_t = \sum_{\ell=0}^{\infty} \psi_\ell Z_{t-\ell}$, $\forall t \in \mathbf{Z}$.

Seasonal autoregressive integrated moving average (SARIMA) models are defined as follows. The process $\{X_t\}_{t \in \mathbf{Z}}$ is said to be a *SARIMA*(p, d, q) \times (P, D, Q) $_S$ process with period S if the differenced process $Y_t := (1 - B)^d (1 - B^S)^D X_t$ is the causal ARMA process

$$\phi(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^S)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

where $\phi(z) = 1 - \dots - \phi_p z^p$, $\Phi(z) = 1 - \dots - \Phi_P z^P$, $\theta(z) = 1 + \dots + \theta_q z^q$ and $\Theta(z) = 1 + \dots + \Theta_Q z^Q$. Hence, the model for $\{X_t\}_{t \in \mathbf{Z}}$ reads $\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D X_t = \theta(B)\Theta(B^S)Z_t$.

To identify a suitable SARIMA model for a given time series, the differencing orders d , D , the model orders p , P , q , Q , and the length S of the seasonal component must be identified. They can be discovered by inspecting the empirical autocorrelation function, the empirical counterpart of the *autocorrelation function* $\mathbb{E}[X_\ell X_0]$, $\ell \in \mathbf{Z}$; see, e.g., [3]. The model coefficients $\{\phi_\ell\}_{\ell=1}^p$, $\{\Phi_\ell\}_{\ell=1}^P$, $\{\theta_\ell\}_{\ell=1}^q$, $\{\Theta_\ell\}_{\ell=1}^Q$, and the white noise variance σ^2 can be estimated via parameter estimation procedures for ARMA processes. The maximum likelihood method produces the most efficient estimates in the special case of Gaussian time series. Initial values for the model coefficients can be obtained by the Hannan-Rissanen algorithm (cf. [3, Section 5]), which solves the problem of order selection and parameter estimation for ARMA processes simultaneously.

In case of the daily mean load process, stationary residuals were obtained after three differencing operations (two lag-364 differencing operations followed by one lag-1 differencing). The residuals were treated as part of a realization of the stochastic process

$$\{Y_j := \bar{\mathbf{d}}_j - \bar{\mathbf{d}}_{j-1} - 2\bar{\mathbf{d}}_{j-364} + 2\bar{\mathbf{d}}_{j-365} + \bar{\mathbf{d}}_{j-728} - \bar{\mathbf{d}}_{j-729}\}.$$

For $\{Y_j\}$ the Hannan-Rissanen algorithm from the *Mathematica Time Series Pack* [38] selected an ARMA(1,1) model that served as an initial model for the maximum likelihood method. The maximum likelihood estimates for the model coefficients and random noise process led to the time series model

$$Y_j - \hat{\phi}_1 Y_{j-1} = Z_j + \hat{\theta}_1 Z_{j-1}, \quad j \in \mathbf{Z}, \quad \text{where}$$

$$\hat{\phi}_1 = 0.357756, \quad \hat{\theta}_1 = -0.639978, \quad \{Z_j\} \sim N(0, 15533.88), \quad j \in \mathbf{Z}.$$

Accordingly, for the daily mean load process $\{\bar{\mathbf{d}}_j\}_{j \in \mathbf{Z}}$ we obtain the SARIMA(1, 1, 1) \times (0, 2, 0)₃₆₄ model

$$(1-B)(1-B^{364})^2(1-\hat{\phi}_1 B)\bar{\mathbf{d}}_j = (1+\hat{\theta}_1 B)Z_j, \quad (14)$$

which can be converted into the general ARMA(730, 1) model

$$\begin{aligned} \bar{\mathbf{d}}_j &- (1+\hat{\phi}_1)\bar{\mathbf{d}}_{j-1} + \hat{\phi}_1\bar{\mathbf{d}}_{j-2} - 2\bar{\mathbf{d}}_{j-364} + 2(1+\hat{\phi}_1)\bar{\mathbf{d}}_{j-365} - 2\hat{\phi}_1\bar{\mathbf{d}}_{j-366} \\ &+ \bar{\mathbf{d}}_{j-728} + (\hat{\phi}_1 - 1)\bar{\mathbf{d}}_{j-729} + \hat{\phi}_1\bar{\mathbf{d}}_{j-730} = Z_j + \hat{\theta}_1 Z_{j-1}, \quad j \in \mathbf{Z}. \end{aligned} \quad (15)$$

For modelling the time dependence of the mean-corrected load records corresponding to days of the same category k , $k = 1, \dots, 8$, polynomial-based linear regression models of the form

$$\hat{d}_{k\tau} = \sum_{l=0}^{m_k} \beta_{kl} \tau^l + \epsilon_{km_k} \quad (\tau = 1, \dots, 24) \quad (16)$$

have been fitted, where the *error term* ϵ_{km_k} is normally distributed with zero mean and variance $\sigma_{m_k}^2$. The degree m_k of the polynomials will be fixed later. For model fitting, regression diagnostics, and forecasting we used the statistical package *S-PLUS* [33].

The statistical model for the load is obtained by combining the models for the daily mean load and the mean-corrected load records according to (12). The regression

models for the mean-corrected load records that correspond to different day categories are included into (12) by using *day category variables* D_{jk} defined by

$$D_{jk} := \begin{cases} 1, & \text{cat}(j) = k, \\ 0, & \text{otherwise,} \end{cases} \quad (j \in \mathbf{J}; k = 1, \dots, 8).$$

With these definitions (12) may be rewritten as

$$x_{j\tau} = \sum_{k=1}^8 D_{jk} \hat{d}_{k\tau} + \bar{d}_j \quad (j \in \mathbf{J}; \tau = 1, \dots, 24). \quad (17)$$

The different time scales for the historical load records and the load process can be synchronized by an index transformation:

$$d_t = \sum_{k=1}^8 D_{\lfloor \frac{t}{24} \rfloor, k} \hat{d}_{k, r(t/24)} + \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor}, \quad t \in \mathbf{Z}, \quad (18)$$

where $\lfloor \frac{t}{24} \rfloor$ denotes the lower integer part of $\frac{t}{24}$ and $r(t/24)$ the remainder of t upon division by 24. Finally, the *statistical model of the load* is obtained by inserting (15) and (16) into (18), i.e., it reads

$$\begin{aligned} d_t = & \sum_{k=1}^8 D_{\lfloor \frac{t}{24} \rfloor, k} \sum_{l=0}^{m_k} \beta_{kl} (r(t/24))^l \\ & + (1 + \hat{\phi}_1) \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 1} - \hat{\phi}_1 \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 2} + 2\bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 364} - 2(1 + \hat{\phi}_1) \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 365} \\ & + 2\hat{\phi}_1 \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 366} - \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 728} - (\hat{\phi}_1 - 1) \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 729} - \hat{\phi}_1 \bar{\mathbf{d}}_{\lfloor \frac{t}{24} \rfloor - 730} \\ & + \sum_{k=1}^8 D_{\lfloor \frac{t}{24} \rfloor, k} \epsilon_{km_k} + Z_{\lfloor \frac{t}{24} \rfloor} + \hat{\theta}_1 Z_{\lfloor \frac{t}{24} \rfloor - 1} \quad (t \in \mathbf{Z}). \end{aligned} \quad (19)$$

To select the degrees m_k of the regression polynomials we measured the squared distance between (19) and the historical load data for the third year. The best fit was obtained for $m_k = 10$, $k = 1, \dots, 8$.

The stochastic model (19) for the electrical load is used to simulate a number of load scenarios for the time horizon $\{1, \dots, T\}$ by employing a (pseudo) random number generator for the independent normal random variables ϵ_{km_k} , $k = 1, \dots, 8$, and Z_j , $j \in \mathbf{J}$. In this way, S load scenarios $\{d^i\}_{i=1}^S$ are generated which have identical probabilities $p_i = \frac{1}{S}$, $i = 1, \dots, S$, and coincide for $t = 1, \dots, t_1$. Hence, the nodes at $t = 1, \dots, t_1$ are the first stage nodes of a specific scenario tree forming a *fan of individual scenarios*. Clearly, this tree could be used as input of the optimization algorithm described in Section 3. We notice, however, that such a tree contains a relatively large number of nodes, namely, $|\mathcal{N}| = t_1 + (T - t_1)S$.

4.2 Construction of Scenario Trees

Next we describe a general methodology that successively reduces the number of nodes of a fan $\xi = \{\xi^i\}_{i=1}^S$ of individual scenarios by modifying the tree structure and by

bundling similar scenarios. This methodology is based on a successive scenario reduction technique developed in [13, 21] (see also the forthcoming work [14]). The idea is to compare the probability distance of original and reduced trees, and to delete scenarios if the reduced tree is still close enough to the original one. The probability distance has to be chosen such that the underlying stochastic program behaves stable with respect to this distance when changing the probability distribution. Here, stability means that the optimal costs and solution sets behave continuously with respect to such changes. The interested reader is referred to the discussion in [13].

In the context of stochastic power scheduling models, we use the Kantorovich distance D_K of (multivariate) probability distributions (cf. [29, Section 5]). For discrete probability distributions with finitely many scenarios the distance D_K is just the optimal value of a linear transportation problem. Let P denote the probability distribution of ξ with scenarios ξ^i and probabilities p_i , $i = 1, \dots, S$, and Q that with scenarios $\tilde{\xi}^j$ and probabilities q_j , $j = 1, \dots, \tilde{S}$. Then it holds

$$D_K(P, Q) = \inf \left\{ \sum_{i=1}^S \sum_{j=1}^{\tilde{S}} \eta_{ij} c_T(\xi^i, \tilde{\xi}^j) : \eta_{ij} \geq 0, \sum_{i=1}^S \eta_{ij} = q_j, \sum_{j=1}^{\tilde{S}} \eta_{ij} = p_i, \forall i, \forall j \right\}, \quad (20)$$

where c_t is defined by $c_t(\xi^i, \tilde{\xi}^j) := \sum_{\tau=1}^t |\xi_\tau^i - \tilde{\xi}_\tau^j|$ for each $t = 1, \dots, T$.

Now, let Q be the probability distribution of a reduced tree of ξ , i.e., the support of Q consists of scenarios ξ^j for $j \in \{1, \dots, S\} \setminus J$ and J denotes some index set of deleted scenarios. For fixed $J \subset \{1, \dots, S\}$, the scenario tree Q_* based on the scenarios $\{\xi^j\}_{j \notin J}$ having minimal D_K -distance to P may be computed explicitly ([13, Theorem 3.1]). The minimal distance is

$$D_K(P, Q_*) = \sum_{i \in J} \pi_i \min_{j \notin J} c_T(\xi^i, \xi^j) \quad (21)$$

and the probability q_j^* of scenario ξ^j , $j \notin J$, of Q_* is given by the rule

$$q_j^* := p_j + \sum_{i \in J(j)} p_i, \quad J(j) := \{i \in J : j = j(i)\}, \quad j(i) \in \arg \min_{j \notin J} c_T(\xi^i, \xi^j), \quad \forall i \in J. \quad (22)$$

This means that the scenario ξ^j of the reduced tree represents the bundle $\{\xi^i\}_{i \in J(j)}$ of original scenarios and its probability is given by formula (22). Since the solution sets of $\min_{j \notin J} c_T(\xi^i, \xi^j)$ are non-unique in general, the optimally reduced tree is not uniquely determined.

Our approach for constructing a scenario tree ξ_{app} that approximates the original fan ξ of individual scenarios consists in a successive reduction procedure by applying the above reduction argument recursively to (sub)trees on the time horizons $\{1, \dots, t\}$ for $t = T, \dots, t_1$. More precisely, given some tolerance $\varepsilon > 0$ and constant $\alpha > 1$, an index set J_t is determined in the $(T - t + 1)$ -th step such that $|J_{t_1}| = S - 1$ and

$$\sum_{i \in J_t} \pi_i \min_{j \notin J_t} c_t(\xi^i, \xi^j) \leq \frac{\varepsilon(\alpha - 1)}{\alpha^{T-t+1}} \quad (t = T, \dots, t_1) \quad (23)$$

by the simultaneous backward reduction algorithm ([21, Algorithm 2.2]). While J_t is the index set of deleted scenarios in the $(T - t + 1)$ -th step, the index set of remaining

scenarios is denoted by I_t , $t = t_1, \dots, T+1$, i.e., it holds that $I_t \cup J_t = I_{t+1}$, $t = t_1, \dots, T$, $|I_{t_1}| = 1$ and $I_{T+1} = \{1, \dots, S\}$.

The approximate scenario tree ξ_{app} with sets \mathcal{N}_t of nodes at time period t is then defined by setting $|\mathcal{N}_t| := |I_t|$ and $\{\xi_{\text{app}}^n\}_{n \in \mathcal{N}_t} := \{\xi_t^j\}_{j \in I_t}$ for every $t = t_1, \dots, T$. According to redistribution rule (22) the probability π_t^j of ξ_t^j is recursively given by $\pi_{T+1}^i := p_i$, $i = 1, \dots, S$, and

$$\pi_t^j := \pi_{t+1}^j + \sum_{i \in J(t,j)} \pi_{t+1}^i, \quad J(t, j) = \{i \in J_t : j = j(t, i)\}, \quad j(t, i) \in \arg \min_{j \notin J_t} c_t(\xi^i, \xi^j),$$

for every $j \in I_t$ and $t = T, \dots, t_1$. Hence, the approximate tree ξ_{app} exhibits the following structure. It holds that $\xi^n = \xi_{\text{app}}^n$ for every $n \in \mathcal{N}_{\text{first}}$ and $|\mathcal{N}_T| = |I_T|$. The cardinality of $\mathcal{N}_+(n)$ is equal to $|\{j\} \cup J(t, j)|$ if the node $n \in \mathcal{N}_t$ corresponds to the index $j \in I_t$. This means that the index sets $\{J(t, j)\}_{j \in I_t}$ characterize the branching degree of ξ_{app} at period t .

5 Numerical Results

The Lagrangian relaxation algorithm was implemented in C++ except for the proximal bundle method, for which the Fortran package NOA 3.0 [23] was used as a callable library. For numerical tests we considered the hydro-thermal power system of VEAG (with $T = 168$, $|\mathcal{I}| = 25$ and $|\mathcal{J}| = 7$) under uncertain load (i.e., the remaining data were deterministic). The test runs were performed on an HP 9000 (780/J280) computer with 180 MHz frequency and 768 MByte main memory under HP-UX 10.20.

For testing the performance of the optimization algorithm a bunch of load scenario trees was randomly generated. The upper part of Table 2 contains test results of the Lagrangian relaxation algorithm based on the heuristic LH1 with the parameters $\text{opt_tol} = 10^{-3}$ and $\text{NGRAD} = 50$ for NOA 3.0. In particular, it provides computing times and gaps for different numbers of scenarios (S) and four randomly generated scenario trees, each having a different number of nodes (N). The gap refers to the relative difference

$$\frac{1}{D_*} \left(\sum_{t=1}^T \sum_{i \in \mathcal{I}} [C_{it}(p_{it}, u_{it}) + S_{it}(u_i)] - D_* \right)$$

of the costs of the scheduling decision (u, p, v, w) and the optimal value D_* of the dual. We note that, in general, this gap does not provide a quality measure for the approximate first stage solution (it may even become nonpositive). When reading the computing times in Table 2, it is worth recalling that $N = 4000$ and $N = 8000$ correspond to 100,000 and 200,000 binary variables in the model (1)–(4), respectively.

The lower part of Table 2 reports computing times and gaps for the Lagrangian relaxation algorithm based on LH2. Here, the parameters for NOA 3.0 are $\text{opt_tol} = 10^{-5}$ and $\text{NGRAD} = 200$, and the gap refers to the following bound of the relative duality gap

$$\frac{1}{D_*} \left(\sum_{n \in \mathcal{N}} \pi_n \sum_{i \in \mathcal{I}} [C_i^n(p_i^n, u_i^n) + S_i^n(u_i^{\text{path}(n)})] - D_* \right).$$

This bound provides an accuracy certificate for the approximate primal-feasible solution $\{(u^n, p^n, v^n, w^n)\}_{n \in \mathcal{N}}$. Figure 7 displays the final output of the Lagrangian relaxation

Lagrangian relaxation algorithm based on heuristic LH1						
S	N	time[s]	gap[%]	N	time[s]	gap[%]
20	1982	89	0.15	1627	94	0.10
20	1651	68	0.37	1805	85	0.07
50	4530	475	0.18	4060	274	0.10
50	4041	313	0.10	4457	288	0.43
100	9230	1183	0.11	9224	1072	0.13
100	7727	930	0.09	8867	1234	0.30

Lagrangian relaxation algorithm based on LH2				
S	N	NOA time[s]	total time[s]	gap[%]
1	168	10	16	0.20
5	542	65	101	0.19
10	983	128	230	0.71
17	1786	278	733	0.45
21	2098	351	531	0.39
27	2208	380	8349	0.73
32	2173	359	3337	0.66
39	3848	874	4092	0.82

Table 2: Test results for randomly generated load trees

algorithm based on LH2 and on a load scenario tree with 17 scenarios and 1786 nodes. While the “deterministic” Lagrangian heuristics LH1 requires only short computing times, this becomes quite different for the “stochastic” heuristics LH2. Table 2 gives more insight into the (total) computing times of different test runs. Higher computing times are always due to lots of economic dispatch runs required by LH2. It is worth mentioning here that LH2 is quite sensitive to the accuracy of the dual solution, i.e., to the optimality tolerance of the proximal bundle method. The advantage of using LH1 consists in low running times even for mid-size scenario trees, while its drawbacks are that only first-stage solutions are provided with no accuracy bounds. The advantage of LH2 is that it produces a “stochastic” solution together with a guaranteed accuracy

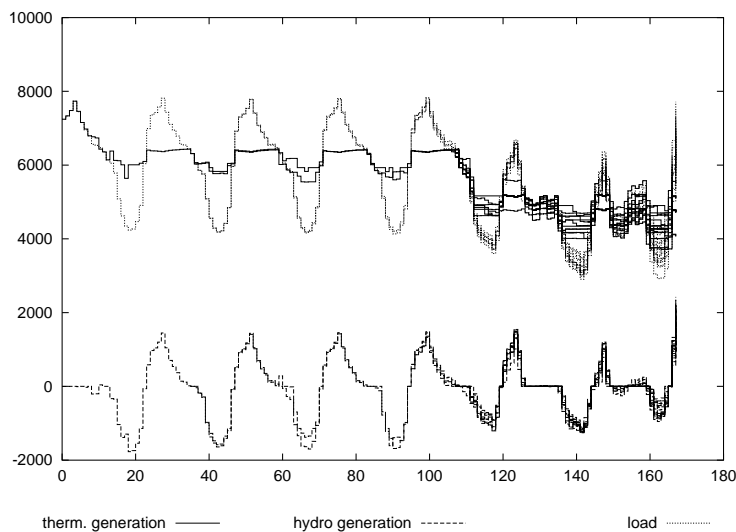


Figure 7: Weekly optimal stochastic solution with 17 scenarios

ε_{rel}	S	N	Variables		Constraints	Nonzeros	time[s]
			binary	continuous			
0.6	1	168	4200	7728	16975	44695	7.83
0.1	67	515	12875	23690	52484	137459	17.09
0.05	81	901	22525	41446	91568	240233	37.82
0.01	94	2660	66500	122360	269318	708218	150.14
0.005	96	3811	95275	175306	385583	1014398	291.65
0.001	100	9247	231175	425362	934647	2460402	1176.38

Table 3: Test results for solving the stochastic dual based on a reduced load scenario tree of relative tolerance ε_{rel}

bound, but at the expense of higher computing times even for scenario trees of smaller size. For further information the interested reader is referred to [25].

Another test combined the Lagrangian relaxation algorithm with the load scenario tree generation technique of Section 4.2. First, we used the statistical model to generate $S = 100$ load scenarios d^i with identical probabilities $p_i = 0.01$ for an hourly discretized time horizon of one week in summer. The accuracy of the load model for the summer season allowed to choose the first day of the optimization horizon as the first stage period, i.e., $t_1 := 24$. Hence, the scenario values for the first-stage periods $t = 1, \dots, 24$ coincide with the load prediction for this period. Given an appropriate number of starting load values, the prediction can be computed from (19) by ignoring the realizations of the random noise process $\{Z_t\}$ and of the error terms ε_{km_k} . To compute the remaining $T - t_1 = 144$ values of a single scenario from (19) we simulated realizations of the random noise process and of the error terms using the random number generators contained in the RANLIBC library [4]. Table 3 reports the computing times for solving the stochastic dual (7) based on different reduced load scenario trees, each having a different numbers of scenarios (S) and of nodes (N). The trees are constructed by the algorithm in Section 4.2 for $\alpha := 2$ and different relative tolerances $\varepsilon_{\text{rel}} := \frac{\varepsilon}{\varepsilon_{\text{max}}}$, where ε_{max} is the best possible Kantorovich distance D_K of the probability distribution $P = 0.01 \sum_{i=1}^{100} \delta_{d^i}$ to one of its scenarios endowed with unit mass. Figure 8 provides an impression of the improved accuracy of the dual optimum and of the scenario tree structure for different relative tolerances.

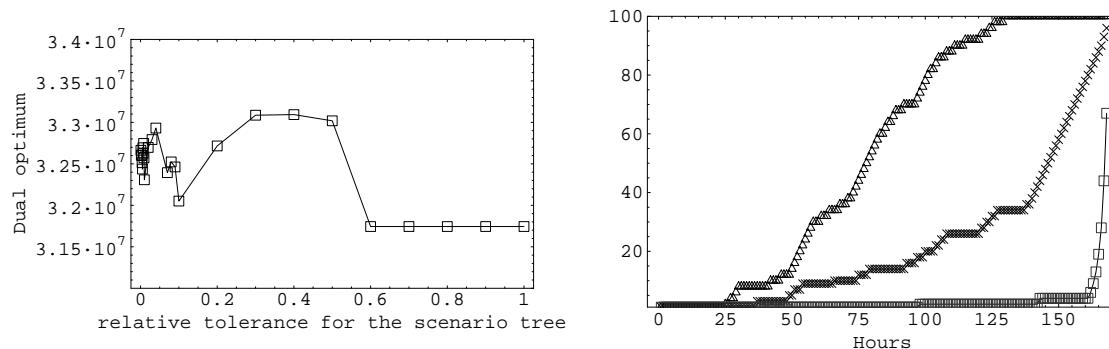


Figure 8: Dual optimum (left) and $|I_t|$ ($t = 1, \dots, T$) for scenario trees with relative tolerance $\varepsilon_{\text{rel}} = 0.001$ (Δ), 0.005 (\times), 0.01 (\square) (right)

Acknowledgement:

This research was supported by the Schwerpunktprogramm *Echtzeit-Optimierung grosser Systeme* of the Deutsche Forschungsgemeinschaft and by the BMBF-Programm *Neue mathematische Verfahren in Industrie und Dienstleistungen*. The authors are grateful to their colleagues D. Dentcheva, M.P. Nowak and I. Wegner (formerly Humboldt-University Berlin) and to G. Schwarzbach, J. Thomas and J. Krause (VEAG Vereinigte Energiewerke AG, Berlin) for the outstanding cooperation over many years. Further thanks are due to K.C. Kiwiel (Polish Academy of Sciences, Warsaw) for his invaluable contributions to this project during his visits to Berlin as a Fellow of the Alexander von Humboldt Foundation and for the permission to use his NOA 3.0 package, and to J. Dupačová (Charles University Prague) for her collaboration in modelling load profiles.

References

- [1] Bacaud, L., Lemaréchal, C., Renaud, A., Sagastizábal, C.: Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners. *Computational Optimization and Applications* 20 (2001), 227–244.
- [2] Bogensperger, J.H.: Wochenplanung in Stromhandel und Erzeugung. In: *Optimierung in der Energieversorgung III*, VDI-Berichte 1508, VDI-Verlag, Düsseldorf 1999, 183–189.
- [3] Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*, Springer-Verlag, New York 1996.
- [4] Brown, B., Lovato, J.: RANLIBC Library of C routines for random number generation. Department of Biomathematics, The University of Texas, Houston, 1989, available from the STATLIB UTexas Archive (<http://lib.stat.cmu.edu/general/Utexas/>).
- [5] Carøe, C.C.: Decomposition in stochastic integer programming. Ph.D. thesis, Institute of Mathematical Sciences, University of Copenhagen, 1998.
- [6] Carøe, C.C., Schultz, R.: 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 (<http://www.zib.de/pub/pw/index.en.html>).
- [7] Carøe, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. *Operations Research Letters* 24 (1999), 37–45.
- [8] Carpentier, P., Cohen, G., Culioli, J.-C., Renaud, A.: Stochastic optimization of unit commitment: a new decomposition framework. *IEEE Transactions on Power Systems* 11 (1996), 1067–1073.
- [9] Conejo, A.J., Arroyo, J.M., Jiménez Redondo, N., Prieto, F.J.: Lagrangian relaxation applications to electric power operations and planning problems. In: *Modern Optimization Techniques in Power Systems* (Song, Y.H. ed.). Kluwer, Dordrecht 1999, 173–203.
- [10] Dentcheva, D., Römisch, W.: Optimal power generation under uncertainty via stochastic programming. *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer, Berlin 1998, 22–56.

- [11] Dentcheva, D., Römisch, W.: Duality gap estimation for nonconvex dynamic stochastic optimization models. Manuscript, 2002.
- [12] Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. *Annals of Operations Research* 100 (2000), 25–53.
- [13] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics. Preprint 00-9, Institut für Mathematik, Humboldt-Universität Berlin, 2000 and revised version submitted to *Mathematical Programming*.
- [14] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Constructing scenario trees from simulation scenarios. Manuscript, 2002.
- [15] Feltenmark, S., Kiwiel, K.C.: Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization* 10 (2000), 697–721.
- [16] Fleten, S.-E., Wallace, S.W., Ziemba, W.T.: Hedging electricity portfolios via stochastic programming. In: *Decision Making under Uncertainty: Energy and Power* (Greengard, C., Ruszczyński, A. eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag, New York 2002, 71-94 (to appear).
- [17] Gollmer, R., Nowak, M.P., Römisch, W., Schultz, R.: Unit commitment in power generation – A basic model and some extensions. *Annals Operations Research* 96 (2000), 167–189.
- [18] Gröwe-Kuska, N., Heitsch, H., Römisch, W.: Modellierung stochastischer Datenprozesse für Optimierungsmodelle der Energiewirtschaft. In: *IT-Lösungen für die Energiewirtschaft in liberalisierten Märkten*, VDI-Berichte 1647, VDI-Verlag, Düsseldorf, 2001, 69–78.
- [19] Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In: *Decision Making under Uncertainty: Energy and Power* (Greengard, C., Ruszczyński, A. eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag, New York 2002, 39-70 (to appear).
- [20] Gröwe-Kuska, N., Nowak, M.P., Wegner, I.: Modeling of uncertainty for the real-time management of power systems. In: *Online Optimization of Large Scale Systems* (Grötschel, M., Krumke, S.O., Rambau, J. eds.), Springer-Verlag, Berlin 2001, 621–645.
- [21] Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Preprint 01-8, Institute of Mathematics, Humboldt-University Berlin, 2001.
- [22] Kiwiel, K.C.: Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming* 46 (1990), 105–122.
- [23] Kiwiel, K.C.: 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.

- [24] Krasenbrink, B.: Integrierte Jahresplanung von Elektrizitätserzeugung und -handel, *Aachener Beiträge zur Energieversorgung* (Haubrich, H.-J. ed.), Band 81, Aachen 2002.
- [25] Nowak, M.P.: Stochastic Lagrangian Relaxation in Power Scheduling of a Hydro-Thermal System under Uncertainty. Ph.D. thesis, Institute of Mathematics, Humboldt-University Berlin, 2000 (<http://dochost.rz.hu-berlin.de/dissertationen/>).
- [26] Nowak, M.P., Römisch, W.: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Annals Operations Research* 100 (2000), 251–272.
- [27] Nürnberg, R., Römisch, W.: A two-stage planning model for power scheduling in a hydro-thermal system under uncertainty, Preprint 00-11, Institute of Mathematics, Humboldt-University Berlin, 2000 and submitted.
- [28] Philpott, A.B., Craddock, M., Waterer, H.: Hydro-electric unit commitment subject to uncertain demand. *European Journal Operational Research* 125 (2000), 410–424.
- [29] Rachev, S.T.: *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester, 1991.
- [30] Rockafellar, R.T., Wets, R.J-B: The optimal recourse problem in discrete time: L^1 -multipliers for inequality constraints. *SIAM Journal Control Optimization* 16 (1978), 16–36.
- [31] Römisch, W., Schultz, R.: Multistage stochastic integer programs: An introduction. In: *Online Optimization of Large Scale Systems* (Grötschel, M., Krumke, S.O, Rambau, J. eds.), Springer-Verlag, Berlin 2001, 579–598.
- [32] Sheble, G.B., Fahd, G.N.: Unit commitment literature synopsis. *IEEE Transactions on Power Systems* 9 (1994), 128–135.
- [33] S-PLUS User's Guide Version 4.5. Data Analysis Products Division, MathSoft, Seattle, 1998.
- [34] Stern, B.: Kraftwerkseinsatz und Stromhandel unter Berücksichtigung von Planungsunsicherheiten. *Aachener Beiträge zur Energieversorgung* (Haubrich, H.-J. ed.), Band 78, Aachen 2001.
- [35] Takriti, S., Birge, J.R., Long, E.: A stochastic model for the unit commitment problem. *IEEE Transactions Power Systems* 11 (1996), 1497–1508.
- [36] Takriti, S., Birge, J.R.: Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs. *Operations Research* 48 (2000), 91–98.
- [37] Takriti, S., Krasenbrink, B., Wu, L.S.-Y.: Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. *Operations Research* 48 (2000), 268–280.
- [38] Wolfram Research: *Mathematica Time Series Pack: Reference and User's Guide*. Champaign, IL, 1995.
- [39] Zhuang, F., Galiana, F.D.: Towards a more rigorous and practical unit commitment by Lagrangian relaxation. *IEEE Transactions Power Systems* 3 (1988), 763–773.