A simple recourse model for power dispatch under uncertain demand*

Nicole Gröwe[a], Werner Römisch[b] and Rüdiger Schultz[a]

[a]*Konrad-Zuse-Zentrum für Informationstechnik Berlin, Heilbronner Straße 10, D-10711 Berlin, Germany*

[b]*Humboldt-Universität Berlin, Institut für Mathematik, Unter den Linden 6, D-10099 Berlin, Germany*

Optimal power dispatch under uncertainty of power demand is tackled via a stochastic programming model with simple recourse. The decision variables correspond to generation policies of a system comprising thermal units, pumped storage plants and energy contracts. The paper is a case study to test the kernel estimation method in the context of stochastic programming. Kernel estimates are used to approximate the unknown probability distribution of power demand. General stability results from stochastic programming yield the asymptotic stability of optimal solutions. Kernel estimates lead to favourable numerical properties of the recourse model (no numerical integration, the optimization problem is smooth convex and of moderate dimension). Test runs based on real-life data are reported. We compute the value of the stochastic solution for different problem instances and compare the stochastic programming solution with deterministic solutions involving adjusted demand portions.

Keywords: Power dispatch under uncertainty, stochastic programming, asymptotic stability.

The dispatch of electric power is one of the most challenging contemporary planning problems. Mathematical models for optimal power dispatch are usually characterized by the combination of several difficulties such as a very large number of (often also discrete) variables, unavoidable non-linearities (mainly when including the transmission network into the model) and last but not least uncertainty of problem data. The latter typically occurs with future demand of electric power but is also encountered for the output of generating units or the reliability of the transmission network (contingencies). A comprehensive power dispatch model reflecting both the economical and technological reality almost completely and with sufficient accuracy is still beyond the contemporary algorithmic and computational abilities (cf. [32] and, for a recent account, [14]).

*This research is supported by the Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung" of the Deutsche Forschungsgemeinschaft.

The present paper considers electrical power dispatch from the viewpoint of stochastic programming.

We concentrate on short-term cost-optimal planning of electricity production with a fixed configuration of generating units in the presence of uncertainty about the power demand. Our model is a simplified one in that we exclude unit commitment (start-ups and shut-downs of units) and network questions (transmission losses, phase synchronization etc.). We impose a very simple compensation scheme (simple recourse) for random deviations between here-and-now scheduling decisions and demand realizations. Due to the simple recourse scheme, compensation actions are separated and only the one-dimensional marginals of the demand distribution come into play. Dependencies between demand values in the different time steps are thus neglected. Our model was developed in cooperation with the power company serving the eastern part of Germany. Generating units comprise thermal (coal fired) power stations and pumped (hydro) storage plants. The latter differ from traditional hydro power stations by their dual mode of operation: in addition to the generation there is a pumping mode, i.e. water that was used for power generation can be pumped upward and then be used for generation again. Therefore, constraints interconnecting all the different time intervals are mandatory for each pumped storage plant. For thermal units such constraints are comparatively rare: ramping constraints and fuel quotas can be mentioned in this respect. For our energy system ramping constraints turned out to be nonbinding for almost all units, and fuel quotas, if at all, occurred only very rarely. Let us also mention that for pumped storage plants in East Germany inflow to and outflow from the reservoirs are negligibly small. Altogether there are 26 thermal units and 5 pumped storage plants. Peak loads reached 11000 MW, during the night the power demand decreased to around 8000 MW. The 5 pumped storage plants together have a working capacity of 7580 MWh with a maximal output of 1600 MW. Pumping efficiencies range from 50–73%. Due to technological reasons (formerly almost no, in the meantime only some first interconnections between East Germany and the West European power nets exist) there is only very limited exchange with external producers or customers. Therefore, only a very simple exchange contract was included into our model.

Basic features of the above model were already presented in [16]. The present paper goes beyond [16] by attuning the kernel estimation method, performing numerical experiments with real life data and discussing benefits of our stochastic programming power dispatch model. Moreover, the underlying theory is developed towards improved convergence rates for estimates of optimal solutions.

We place accent on utilizing modelling techniques, theory and algorithms from two-stage stochastic programming to handle the randomness of power demand. The models that were in operation at the power company did not include the randomness. Fixed demand values based on statistical estimates were in use. Compensating deviations between scheduled power and demand realizations was not part of the optimization.

As a first step towards a more comprehensive power dispatch model we set up a stochastic program with simple recourse about which we report here. The model offers here-and-now decisions to the dispatcher to minimize the sum of generation costs for the schedule fixed before the realization of the demand plus expected costs for compensating deviations occurring after demand realization. For the conversion of statistical information into an estimate for the probability distribution of power demand we propose (as in [16]) the use of non-parametric kernel estimators (which is non-standard in the stochastic programming literature). Of course, there exists the alternative to derive discrete probability distributions from the raw data (empirical distributions, scenarios) which enables to use adapted solution methods for large-scale structured optimization problems. With discrete probability distributions objective function values of the stochastic program can be computed by summation. Multidimensional integration is then avoided. The simple recourse point of view adapted in the present paper leads to separability of second-stage costs and, hence, only one-dimensional integrals occur which can be calculated analytically. Therefore numerical integration is avoided at the cost of having to confine to a simple (i.e. separable) compensation scheme. Our techniques lead to nonlinear (smooth convex) models without increased dimension (in contrast to the usually large-scale models arising from discrete probability distributions). Numerical treatment of the model becomes possible by using standard nonlinear optimization software (e.g. MINOS, [21]).

Before starting computations based on kernel estimators it is necessary to analyze whether this type of approximation is in tune with the stochastic program. In other words, we have to check the stability behaviour of optimal solutions with growing sample size. Applying recent results from the stability theory for more general recourse models [29] we address this issue in our asymptotic analysis in section 3. Section 2 contains a model description. In sections 4 and 5 we discuss the numerical treatment and the test runs.

Our paper should be seen as a first case study to test the kernel estimation method at a simplified power dispatch problem. For more information on tackling decision problems under uncertainty via stochastic programs with recourse the reader is referred to [8, 11, 13, 18, 34]. Further stochastic programs arising in planning of electricity production are discussed in [8, 13, 17].

2. The model

Given a fixed configuration of generating units, the problem of optimal load dispatch reduces to distributing a load (electric power demand) among the units such that the total generation costs are minimal, while additional operational constraints are met. In our model, these operational constraints only concern the generating units themselves and not the electrical network connecting producers and consumers. The reason for such a simplification is that network constraints would lead to a highly non-linear problem ([2]). Therefore, the degrees of freedom in the optimization are commonly reduced by fitting the output bounds of the units to the capabilities of the network or by including an adjusted portion of power demand to compensate transmission losses.

In our model, the generation system comprises thermal units, pumped storage plants and energy contracts. The generation process is up to one (or a few) day(s) with a discretization into hourly (or half-hourly) time intervals. Let N_1 and N_2 denote the numbers of thermal units and pumped storage plants. Let T be the number of time intervals in the discretization. By y_t^i ($i = 1,\dots,N_1$; $t = 1,\dots,T$) we denote the unknown levels of production in the thermal units. For the pumped storage plants we have a generation and a pumping (i.e. power consumption) mode whose unknown levels are s_t^j and w_t^j ($j = 1,\dots,N_2$; $t = 1,\dots,T$), respectively. (In the generation mode the pumped storage plant produces electricity and in the pumping mode it acts as a consumer of electricity.) Furthermore, there are N_3 energy contracts with external companies whose unknown levels are z_t^k ($k = 1,\dots,N_3$; $t = 1,\dots,T$).

Generation costs of the thermal units are fuel costs which we model as strictly convex quadratic functions of power generation levels with no interdependencies between different units. In the literature, this is quite common [6, 15]. Sometimes, also piecewise linear convex functions are met [6]. The energy contracts are modelled independent of each other here and we assume linear costs functions. The pumped storage plants do not cause generation costs. They indirectly contribute to our cost function via the generation costs in the thermal units for the electrical energy that is needed to pump water upward. Altogether, we have the following cost function for the generating units

$$y^T Hy + h^T y + g^T z, \tag{2.1}$$

where $y \in \mathbb{R}^{N_1 T}$, $z \in \mathbb{R}^{N_3 T}$, $h \in \mathbb{R}^{N_1 T}$, $g \in \mathbb{R}^{N_3 T}$, and H is a positive semidefinite diagonal $N_1 T \times N_1 T$ matrix. (Diagonal entries in H and the corresponding components in h are zero if the respective unit is switched off in the respective time interval.)

For each time interval $t = 1,\dots,T$ the total generation amounts to

$$\sum_{i \in I_t} y_t^i + \sum_{j=1}^{N_2} (s_t^j - w_t^j) + \sum_{k=1}^{N_3} z_t^k, \tag{2.2}$$

where $I_t \subset \{1,\dots,N_1\}$ is the index set of units which are on-line in the time interval t. Note that the pumping levels w_t^j enter with a negative sign, since pumping energy has to be made available by the system itself and not via external sources. Introducing the notation $x = (y, s, w, z)^T \in \mathbb{R}^m$ ($m = T(N_1 + 2N_2 + N_3)$) we express the total generation as the T-dimensional vector Ax where A is the $T \times m$ matrix with entries determined by the coefficients in (2.2).

The differences that occur between the planned generation and the observed demand are compensated at costs q_t^+ and q_t^- for under- and overdispatching at time interval t ($t = 1,\dots,T$), respectively. The power demand is now considered as a T-dimensional random vector with distribution μ on $\mathbb{R}^T$. By F_t ($t = 1,\dots,T$) we denote its one-dimensional marginal distribution functions. The expected costs to adjust the planned generation x to the actual demand then read

$$\sum_{t=1}^{T} \int_{\mathbb{R}} \tilde{Q}_t(\tau - [Ax]_t) \mathrm{d}F_t(\tau), \tag{2.3}$$

where

$$\tilde{Q}_t(\tau) = \begin{cases} q_t^+ \tau & \tau \geq 0, \\ -q_t^- \tau & \tau < 0, \end{cases}$$

and $[Ax]_t$ denotes the t-th component of Ax ($t = 1, \ldots, T$). Of course, in a general power generation system random deviations of the demand can be compensated in different ways, for instance by adjusting the output of on-line units (or contracts) or by committing further units (or contracts). This leads to complicated functions for the expected compensation costs and is referred to as complete recourse in two-stage stochastic programming. In this context, the cost function (2.3) corresponds to what is called simple recourse. Of course, the simple recourse approach suppresses important information on various interconnections both with respect to operational constraints and stochastic dependencies between demand values of different time intervals (cf. also the general remarks on the purpose of our model made in the introduction).

Denoting by $g(x)$ the expression in (2.1) we end up with the following cost function for our stochastic program with simple recourse

$$g(x) + \sum_{t=1}^{T} \int_{\mathbb{R}} \tilde{Q}_t(\tau - [Ax]_t) \, \mathrm{d}F_t(\tau). \tag{2.4}$$

The constraints are given by the relations

$$\underline{a}_1 \leq y \leq \bar{a}_1, \quad 0 \leq s \leq \bar{a}_2, \quad 0 \leq w \leq \bar{a}_3, \quad \underline{a}_4 \leq z \leq \bar{a}_4, \tag{2.5}$$

$$S_j^{\text{in}} - S_j^{\max} \leq \sum_{t=1}^{\bar{t}} (s_t^j - \eta_j w_t^j) \leq S_j^{\text{in}}, \quad j = 1, \ldots, N_2, \quad \bar{t} = 1, \ldots, T, \tag{2.6}$$

$$\sum_{t=1}^{T} (s_t^j - \eta_j w_t^j) = b_{1j}, \quad j = 1, \ldots, N_2, \quad \sum_{t=1}^{T} z_t^k = b_{2k}, \quad k = 1, \ldots, N_3, \tag{2.7}$$

$$\sum_{t \in T_i} y_t^i \leq \bar{c}_i, \quad i = 1, \ldots, N_1. \tag{2.8}$$

The box constraints (2.5) model the limitations for the power output. Thermal units, pumped storage plants and energy contracts, clearly, have output limitations per hour which may vary in time. The inequalities (2.6) express the balance between generation and pumping (measured in energy) in the pumped storage plants: S_j^{in} and $S_j^{\max}$ denote the initial and maximal stocks in energy in the upper dam. In each pumped storage plant the maximal stock in water of the upper dam equals that of the lower dam. Additional in- or outflows are negligibly small. Therefore, pumped storage plants operate with a fixed amount of water that is moved up and down during the time period $\{1,\ldots,T\}$. Excessive hydro generation always has to be compensated by proper pumping (i.e. consumption of electricity) to ensure the availability of the pumped storage plant in later time intervals. The pumping efficiency η_j is then put as the quotient of the energy gained when sending the full content of the upper dam down and the energy needed to pump the whole content of the lower dam upward. The inequalities (2.6) reflect the operational limitations for the pumped storage plants at time step $\bar{t}$. Of course the latter only depend on the initial water stocks and the "history" up to time step $\bar{t}$. (Recall that no additional in- or outflows occur.) In (2.7) balances over time for pumped storage plants and energy contracts are modeled. For a pumped storage plant, for instance, a typical constraint of this type is caused by claiming that at the end of the optimization period there is still a sufficient amount of water in the upper dam. Of course, rules for external contracts are often much more involved than the simple balances in (2.7). For the application mentioned in the introduction, however, these were appropriately accurate. By (2.8) it is possible to model fuel quotas for the thermal units: $T_i \subset \{1,\ldots,T\}$ is a subset of (consecutive) time intervals and $\bar{c}_i$ reflect the maximal outputs for the time period T_i. Of course, a constraint of the type (2.8) is not mandatory for each of the thermal units.

Let us mention that there is no principal difficulty to model further operational constraints via linear inequalities: intermediate and final water levels in the dams of the pumped storage plants, ramping constraints to avoid fluctuating production levels in time that are operationally infeasible due to response limitations of the production plants. With a time discretization into hourly intervals the latter turned out irrelevant for most of the generation units. Our computational results indicate critical fluctuations for 4 units only (cf. table 5 in the appendix, unit no. 7 timesteps 8/9 and 17/18, unit no. 12 timesteps 8/9, unit no. 13 timesteps 7/8, unit no. 14 timesteps 8/9 and 23/24).

The relations (2.5)–(2.8) determine a polyhedron $C \subset \mathbb{R}^m$. Finally we have the following stochastic program with simple recourse

$$\min\left\{ g(x) + \sum_{t=1}^{T} \int_{R} \tilde{Q}_t(\tau - [Ax]_t)\mathrm{d}F_t(\tau) : x \in C \right\}. \tag{2.9}$$

Let us remark that the inclusion of the more complex modes of compensation indicated after (2.3) would here have led us to a stochastic program with complete and mixed-integer recourse, respectively. At the moment, in particular integer recourse models fail to be computationally tractable.

A peculiarity of the above model is that it circumvents Boolean variables to avoid simultaneous generation and pumping in the pumped storage plants. Of course, there are points $x \in C$ for which both $s_{t_o}^{j_o}$, $w_{t_o}^{j_o}$ are non-zero for some $j_o \in \{1,\ldots,N_2\}$,

$t_o \in \{1,\ldots,T\}$. However, such points cannot be optimal as can be seen as follows: Let x be as above and form a new point $\bar{x}$ by replacing $s_{t_o}^{j_o}$ by $s_{t_o}^{j_o} - \eta_{j_o} w_{t_o}^{j_o}$ and $w_{t_o}^{j_o}$ by zero (in case $s_{t_o}^{j_o} - \eta_{j_o} w_{t_o}^{j_o} \geq 0$) or $s_{t_o}^{j_o}$ by zero and $w_{t_o}^{j_o}$ by $-(1/\eta_{j_o})\, s_{t_o}^{j_o} + w_{t_o}^{j_o}$ (in case $s_{t_o}^{j_o} - \eta_{j_o} w_{t_o}^{j_o} < 0$), and leave the remaining components fixed. It is easy to see that $\bar{x}$ fulfils (2.5)–(2.8). Furthermore, $[Ax]_{t_o} < [A\bar{x}]_{t_o}$. Now construct $\tilde{x}$ from $\bar{x}$ by decreasing a suitable number of outputs $\bar{y}_{t_o}^i$ $(i \in \{1,\ldots,N_1\})$ such that $A\tilde{x} = Ax$. Provided that the $y_{t_o}^i$ $(i = 1,\ldots,N_1)$ are not too close to their lower bounds (which is no restriction in practice) we have that $\tilde{x} \in C$. However, due to the strict monotonicity of g with respect to $y_{t_o}^i$ we have $g(\tilde{x}) < g(x)$, by $A\tilde{x} = Ax$ there is no change in (2.3) and x cannot be optimal.

3. Estimation and asymptotic analysis of the model

In this section we present a nonparametric estimation procedure for the unknown marginal distribution functions F_t $(t = 1,\ldots,T)$ of the power demand and derive asymptotic properties of the estimates of the solution sets to problem (2.9). This excursion to mathematical theory is necessary, since it is not a priori clear whether estimating the unknown probability distributions in (2.9) produces estimates (of optimal solutions) that asymptotically converge to solutions of (2.9).

The stochastic program with simple recourse (2.9) can be rewritten in the following form (cf. e.g. [18]):

$$\min\{g(x) + Q(\chi) : Ax = \chi, x \in C\}, \tag{3.1}$$

where

$$\begin{aligned} Q(\chi) &:= \int_{\mathbb{R}^T} \min\{q^+ v^+ + q^- v^- : v^+ - v^- = \zeta - \chi, v^+, v^- \in \mathbb{R}_+^T\}\mu(\mathrm{d}\zeta) \\ &= \sum_{t=1}^{T} \int_{\mathbb{R}} \tilde{Q}_t(\tau - \chi_t)\mathrm{d}F_t(\tau), \end{aligned} \tag{3.2}$$

and μ denotes the (multivariate) probability distribution of the power demand (on $\mathbb{R}^T$), whose marginal distribution functions are $F_t(t = 1,\ldots,T)$. $\tilde{Q}_t$ is defined as in (2.3). q^+ and q^- denote the vectors of compensation costs introduced in section 2. Under the basic assumption that

$$q_t^+ + q_t^- \geq 0 \text{ and } \int_{\mathbb{R}} |\tau|\mathrm{d}F_t(\tau) < +\infty \quad (t = 1,\ldots,T), \tag{3.3}$$

(3.1), (3.2) is a convex program having linear constraints (cf. [18]).

Let $\xi_{1t}, \xi_{2t},\ldots,\xi_{nt},\ldots$ be an independent sample from the distribution function F_t $(t = 1,\ldots,T)$ on some probability space $(\Omega, \mathcal{A}, P)$ $k : \mathbb{R} \to \mathbb{R}$ a function having the property $\int_{\mathbb{R}} k(\tau)\,\mathrm{d}\tau = 1$ ("kernel"), and (b_n) be a sequence of positive numbers tending to zero ("smoothing parameters"). Then we consider the kernel estimates

$$\hat{F}_t^{(n)}(u) := \frac{1}{nb_n} \sum_{i=1}^{n} \int_{-\infty}^{u} k\left(\frac{\tau - \xi_{it}}{b_n}\right) \mathrm{d}\tau \quad (u \in \mathbb{R};\ n \in \mathbb{N}) \tag{3.4}$$

for F_t $(t = 1,\ldots,T)$, $\hat{F}_t^{(n)}$ may be interpreted as a smoothed empirical distribution function to F_t. The advantage of kernel estimators for our purposes lies in the fact that (2.9) becomes a *smooth* convex program if the unknown F_t are replaced by $\hat{F}_t^{(n)}$. For more information and background on kernel-type estimators we refer to [24, 35, 16, 5] and the literature cited therein.

To derive our asymptotic results when the sample size n tends to infinity, we still need some notations. Let $C_b^s = C_b^s(\mathbb{R})$ denote the class of s-times continuously differentiable functions on $\mathbb{R}$ such that their sth derivative is bounded on $\mathbb{R}$. A kernel k is called a class s kernel for some $s \in \mathbb{N}$ if

$$\int_{\mathbb{R}} \tau^i k(\tau)\,\mathrm{d}\tau = 0, \quad i = 1,\ldots,s-1, \quad \int_{\mathbb{R}} |\tau|^s k(\tau)\,\mathrm{d}\tau < +\infty.$$

The following kernels, which, in fact, are both class 2 kernels, will be used in this paper:

(i) triangular kernel : $$k(\tau) := \begin{cases} 1 - |\tau|, & |\tau| \le 1, \\ 0, & \text{otherwise;} \end{cases}$$

(ii) Epanechnikov kernel : $$k(\tau) := \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{\tau^2}{5}\right), & |\tau| \le \sqrt{5} \\ 0, & \text{otherwise.} \end{cases}$$

For a detailed discussion of class s kernels the reader is referred to [10]. For the purpose of this paper we need the following result, providing convergence rates for the uniform distance

$$\|\hat{F}_n - F\|_\infty := \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|,$$

where $\hat{F}_n$ is a kernel type estimator (3.4) for a sufficiently smooth distribution function F. Its proof is taken from [16] and is included for the convenience of the reader.

PROPOSITION 3.1

Let $s \in \mathbb{N}$, assume that $F \in C_b^s$ and k is a class s kernel. Then

$$\|\hat{F}_n - F\|_\infty \le C_s b_n^s + \|F_n - F\|_\infty, \quad \text{for all } n \in \mathbb{N}, \tag{3.5}$$

where $C_s := (1/s!)\,\|F^{(s)}\|_\infty \int_{\mathbb{R}} |x|^s k(x)\,\mathrm{d}x$ and F_n denotes the empirical distribution function to F.

If, in addition, (b_n) is chosen such that $\limsup_{n\to\infty} b_n^s n^{1/2} < \infty$ there holds that

$$\text{(i)} \quad \limsup_{n\to\infty} \left(\frac{2n}{\log\log n}\right)^{1/2} \|\hat{F}_n - F\|_\infty \le 1, \quad P\text{-almost surely, and}$$

$$\text{(ii)} \quad \limsup_{n\to\infty} n^{1/2} E(\|\hat{F}_n - F\|_\infty) < \infty.$$

Here $E(\cdot)$ denotes the expected value with respect to P.

Proof

From lemma 2.3 in [35] we have the estimate

$$\|\hat{F}_n - F\|_\infty \le \|F_n - F\|_\infty + \sup_{x\in\mathbb{R}} |E\,\hat{F}_n(x) - F(x)|.$$

Since k is a class s kernel, we obtain by Taylor's expansion for each $x \in \mathbb{R}$,

$$E\,\hat{F}_n(x) - F(x) = \int_{\mathbb{R}} [F(x - tb_n) - F(x)]k(t)\mathrm{d}t$$
$$= \int_{\mathbb{R}} (-tb_n)^s \frac{1}{s!} F^{(s)}(x - \Theta t b_n)\, k(t)\mathrm{d}t,$$

with some $\Theta \in (0,1)$ depending possibly on x, t and n. This leads to the estimate (3.5). (i) then follows from (3.5) and the Smirnov–Chung law of iterated logarithm for the empirical distribution functions (cf. e.g. [30]). For (ii) we use the following known result for empirical distribution functions:

$$E[\|F_n - F\|_\infty] = O(n^{-1/2}). \qquad \square$$

Remark 3.2

For class 2 kernels with compact support (e.g. for both the triangular and Epanechnikov kernel) the following convergence result is an immediate consequence of corollary 1 in chapter 23.2 of [31]: If $b_n = Cn^{-\alpha}$, for all $n \in \mathbb{N}$, where $\alpha \in (1/4, 1)$ and C is a positive constant, we have

$$\sqrt{n}\,\|\hat{F}_n - F_n\|_\infty \to 0, \quad P\text{-almost surely.}$$

Let ψ denote the set of optimal solutions to (2.9) (or (3.1), respectively) and $\hat{\psi}_n$ the corresponding solution set if F_t is replaced by $\hat{F}_t^{(n)} (t = 1,\ldots,T)$ in (2.9). The following result states two types of asymptotic properties of the Hausdorff distance $d_H(\psi, \hat{\psi}_n)$ as the sample size n tends to infinity. It extends the results of [16] by using a recent quantitative stability result for stochastic programs with recourse [29].

THEOREM 3.3

Let $q_t^+ + q_t^- > 0$, $\int_{\mathbb{R}} |\tau| \mathrm{d} F_t(\tau) < +\infty$ and $F_t \in C_b^s$ for some $s \in \mathbb{N}$ and all $t = 1,\dots,T$. Let k be a class s kernel and assume that (b_n) has the property $\limsup_{n\to\infty} b_n^s n^{1/2} < +\infty$.

Furthermore, assume that there exists an open neighbourhood U of the set $A(\psi) \subset \mathbb{R}^{\mathrm{T}}$ and a constant $r > 0$ such that

$$\prod_{t=1}^{T} F_t'(\tau_t) \geq r, \quad \text{for all } (\tau_1,\dots,\tau_T) \in U. \tag{3.6}$$

Then there exist constants $L, K > 0$ such that

(a) $\limsup_{n\to\infty} \left(\frac{2n}{\log\log n}\right)^{1/2} d_H(\psi, \hat{\psi}_n) \leq LT$ P-almost surely,

(b) $\limsup_{n\to\infty} n^{1/2} E[d_H(\psi, \hat{\psi}_n)] \leq K$, and

(c) $\liminf_{n\to\infty} P(n^{1/2} d_H(\psi, \hat{\psi}_n) < u) \geq 1 + 2T \sum_{j=1}^{\infty} (-1)^j \exp\left(-2j^2 \left(\frac{u}{LT}\right)^2\right)$

for all $u \geq 0$ if $\lim_{n\to\infty} b_n^s n^{1/2} = 0$.

Proof

Due to general measurability results for set-valued mappings (theorems 2.J and 2.K in [26]) the Hausdorff distance $d_H(\psi, \hat{\psi}_n)$ is an extended-real valued random variable (on $(\Omega, \mathcal{A}, P)$). For the following we introduce the notation

$$\eta_{n,t} := \sup_{\tau\in\mathbb{R}} | F_t(\tau) - \hat{F}_t^{(n)}(\tau)| \quad (n \in \mathbb{N};\ t = 1,\dots,T).$$

The assumptions on F_t, k and (b_n) imply (proposition 3.1) the asymptotic properties

$$\limsup_{n\to\infty} \left(\frac{2n}{\log\log n}\right)^{1/2} \eta_{n,t} \leq 1 \quad P\text{-almost surely, and} \tag{3.7}$$

$$\limsup_{n\to\infty} n^{1/2} E[\eta_{n,t}] < +\infty. \tag{3.8}$$

Let $\tilde{\Omega} \subseteq \Omega$ be such that $P(\tilde{\Omega}) = 1$ and (3.7) holds on $\tilde{\Omega}$. Corollary 2.13 in [29] then implies that there exists a constant $L > 0$ such that for each element of $\tilde{\Omega}$ the inequality

$$d_H(\psi, \hat{\psi}_n) \le L \sum_{t=1}^{T} \eta_{n,t} \quad \text{holds for sufficiently large } n \in \mathbb{N}. \tag{3.9}$$

Hence, (a) follows from (3.7) and (3.9).

To prove (b), let $\delta > 0$ be chosen such that for all elements in $\Omega_n := \{\omega \in \Omega : \sum_{t=1}^{T} \eta_{n,t}(\omega) < \delta\}$ (3.9) is valid. Then

$$\begin{aligned} E[d_H(\psi, \hat{\psi}_n)] &= \int_{\Omega_n} d_H(\psi, \hat{\psi}_n)\,\mathrm{d}P + \int_{\overline{\Omega}_n} d_H(\psi, \hat{\psi}_n)\,\mathrm{d}P \\ &\le L \sum_{t=1}^{T} E[\eta_{n,t}] + \operatorname{diam}(C) P\left(\sum_{t=1}^{T} \eta_{n,t} \ge \delta \right) \\ &\le \left(L + \frac{1}{\delta} \operatorname{diam}(C) \right) \sum_{t=1}^{T} E[\eta_{n,t}], \end{aligned}$$

where we used Chebyshev's inequality, and diam(C) denotes the diameter of the bounded set C. Assertion (b) now follows from (3.8).

To show (c) we first proceed in the same way as in the proof of proposition 3.3 in [29]. Assume that $\lim_{n \to \infty} b_n^s n^{1/2} = 0$, and let $u \ge 0$, $n \in \mathbb{N}$. We consider the following events in $\mathcal{A}$:

$$A_o := \left\{ L n^{1/2} \sum_{t=1}^{T} \eta_{n,t} < u \right\}, \quad A_t := \left\{ n^{1/2} \eta_{n,t} < \frac{u}{LT} \right\} \quad (t = 1, \ldots, T),$$

$$B_\delta := \left\{ \sum_{t=1}^{T} \eta_{n,t} < \delta \right\}, \quad \text{where } L > 0, \delta > 0 \text{ are chosen as before.}$$

Then we obtain the estimate:

$$\begin{aligned} P(n^{1/2} d_H(\psi, \hat{\psi}_n) < u) &\ge P(\{n^{1/2} d_H(\psi, \hat{\psi}_n) < u\} \cap B_\delta) \\ &\ge P(A_o \cap B_\delta) \ge P\left(\bigcap_{t=1}^{T} A_t \cap B_\delta \right) \ge P\left(\bigcap_{t=1}^{T} A_t \right) - P(\overline{B}_\delta) \\ &= 1 - P\left(\bigcup_{t=1}^{T} \overline{A}_t \right) - P(\overline{B}_\delta) \ge 1 + \sum_{t=1}^{T} (P(A_t) - 1) - P(\overline{B}_\delta) \\ &= 1 + \sum_{t=1}^{T} \left(P\left(n^{1/2} \eta_{n,t} < \frac{u}{LT} \right) - 1 \right) - P\left(\sum_{t=1}^{T} \eta_{n,t} \ge \delta \right). \end{aligned}$$

(3.7) implies, in particular, $\lim_{n\to\infty} P(\sum_{t=1}^{T} \eta_{n,t} \geq \delta) = 0$.
(3.5) yields the estimate

$$\eta_{n,t} \leq \hat{K} b_n^s + \sup_{\tau\in\mathbb{R}} |F_t(\tau) - F_t^{(n)}(\tau)|$$

for all $n \in \mathbb{N}$, $t = 1,\ldots,T$, and some constant $\hat{K} > 0$. Here $F_t^{(n)}$ denotes the empirical distribution function for F_t and sample size n. Hence, we can continue our estimate and obtain:

$$\liminf_{n\to\infty} P(n^{1/2} d_H(\psi, \hat{\psi}_n) < u) \geq 1 + \sum_{t=1}^{T}\left(\liminf_{n\to\infty} P\left(\hat{K} b_n^s n^{1/2} + \zeta_{n,t} < \frac{u}{LT}\right) - 1\right),$$

where

$$\zeta_{n,t} := n^{1/2} \sup_{\tau\in\mathbb{R}} |F_t(\tau) - F_t^{(n)}(\tau)| (n \in \mathbb{N}, t \in \{1,\ldots,T\}).$$

The Kolmogorov limit theorem (see e.g. [30]) then implies that the sequence $(\zeta_{n,t})$ converges in distribution to a random variable ζ which is independent of t and has the distribution function

$$F_\zeta(r) = 1 - 2\sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2 r^2).$$

Hence, the sequence $(\hat{K} b_n^s n^{1/2} + \zeta_{n,t})$ also converges in distribution to ζ and the Portmanteau Theorem implies

$$\liminf_{n\to\infty} P\left(\hat{K} b_n^s n^{1/2} + \zeta_{n,t} < \frac{u}{LT}\right) \geq F_\zeta\left(\frac{u}{LT}\right).$$

This argument completes the proof. □

Remark 3.4

For the class 2 kernels described above, theorem 3.3 says that, if all marginal distribution functions F_t, $t = 1,\ldots,T$, belong to C_b^2, then $b_n := const.\, n^{-\alpha}$ with $\alpha > 1/4$ is an appropriate choice and leads to "optimal" convergence rates for the optimal sets. For the validity of theorem 3.3 it is obviously not mandatory to select the same sequence of b_n for all $t = 1,\ldots,T$. In fact, in section 4 we will select sequences (b_n) that depend on t.

Remark 3.5

The proper selection of b_n is discussed in [1]: For a symmetric class 2 kernel whose support is a bounded interval, $\alpha = 1/3$ is an asymptotically optimal choice. Moreover, it is suggested in [1] to choose $b_n = 0.5\, \sigma_t n^{-1/3}$, where σ_t denotes the standard deviation of ξ_{1t}.

Remark 3.6

The convergence rate in part (b) of theorem 3.3 improves an earlier result, which was obtained in [16] using material from [28]. Part (c) of theorem 3.3 was inspired by the work in [12] on normalized convergence in stochastic programming. the following large deviation estimate is a particular consequence of theorem 3.3(c):

$$P(d_H(\psi, \hat{\psi}_n) \geq u) \leq 2T \exp\left(-2n\left(\frac{u}{LT}\right)^2\right)$$

for all $u \geq 0$ and sufficiently large $n \in \mathbb{N}$. (Consider only the first member of the alternating series on the right and pass over to the complementary event on the left.)

Remark 3.7

Condition (3.6) is the only critical assumption when applying theorem 3.3 to the model (2.9). It is natural to assume that the demand distributions $F_t(t = 1,\ldots,T)$ are strictly monotonically increasing on their supports. Hence, (3.6) is violated only in the degenerate situation where optimal tenders fall outside the interior of the support of μ. This, however, was never observed in our test runs.

Altogether, the asymptotic analysis of the present section gives a justification for replacing, in our model (2.9), the distributions F_t by estimates $\hat{F}_t^{(n)}$ provided that a (sufficiently) large sample for the random power demand is available.

4. Numerical treatment

The statistical information about the power demand that was accessible to us consisted of records of hourly load for a 3-year period. In a first step, seasonal influences were removed and the daily records were assigned to certain categories characterizing similar days (for instance, mid-week working day, working day before/after a Sunday/public holiday, Saturday, Sunday/public holiday etc.). Then, independent samples of electric power demand for the single time intervals can be read off the records after having categorized the day for which the power dispatch is to be carried out. Let us mention, in this context, that in [20] the same statistical material was used to derive load forecasts.

In the literature both normal [8] and non-normal [7] distributions are suggested for the one-dimensional marginal distribution functions $F_t(t = 1,\ldots,T)$. Our empirical data, however, did not give rise to the normality assumption (especially for day-time intervals). For an illustration please refer to figure 1 showing a kernel estimate for the density and the distribution function of power demand for a midweek working day and the time interval 1 p.m. to 2 p.m. In both cases the triangular kernel with sample size $n = 436$ and smoothing parameter $b_n = 50$ (cf. section 3) was used.

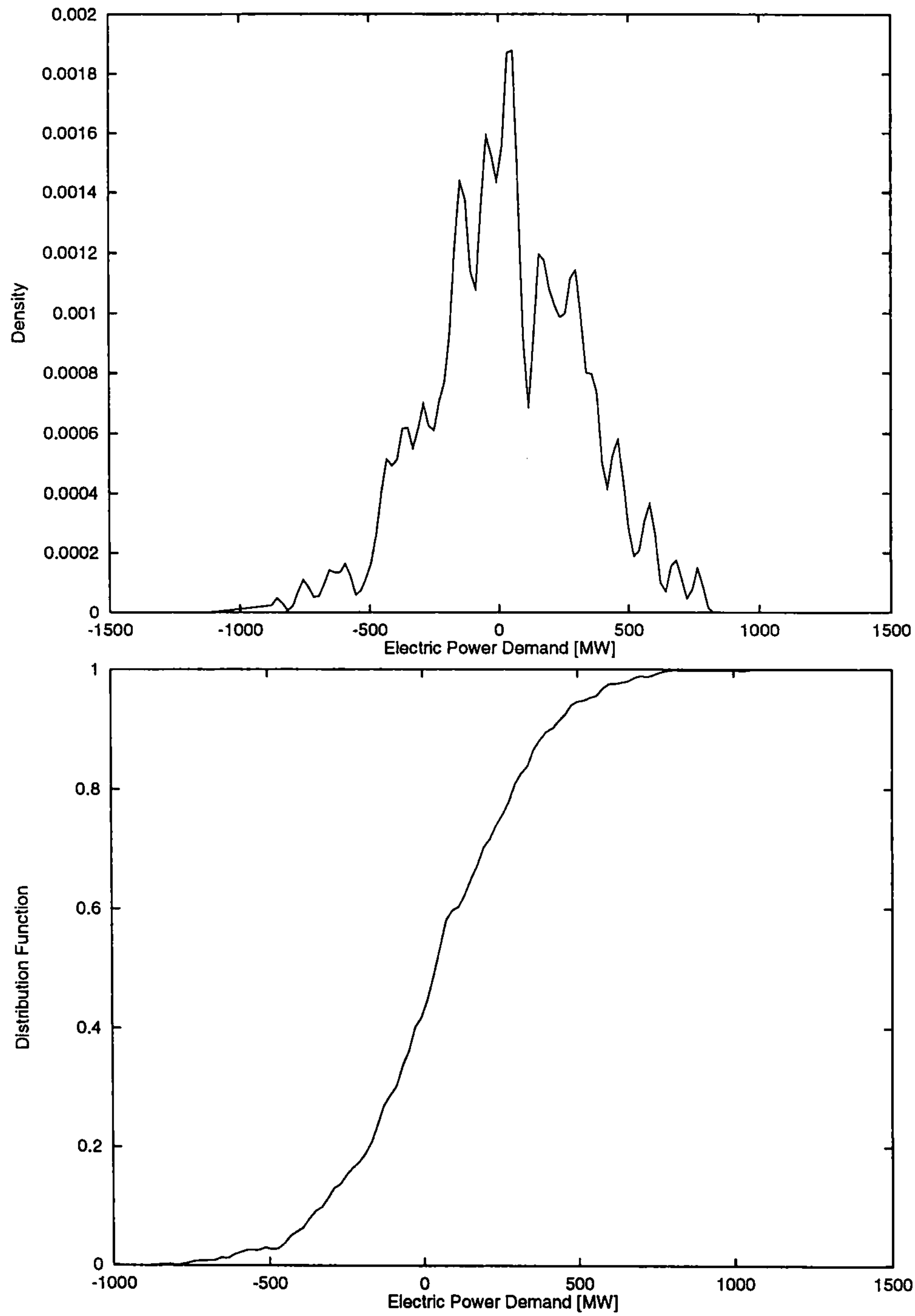

Figure 1. Kernel estimates for the density function (above) and the distribution function (below) for a mid-week working day and the time interval from 1 p.m. to 2 p.m.

An alternative to setting up continuous distributions for the random demand values would consist in estimating the unknown distributions by (discrete) empirical measures. The sample size then corresponds to the cardinality of the support of the empirical measure. Our quadratic two-stage stochastic program becomes a large scale quadratic program with structured constraint matrix. Each mass point of the probability measure induces a block in the constraint matrix. Therefore, when following the above idea, large sample sizes would yield huge models. Although there are efficient procedures for exploiting this structure by decomposition techniques (see e.g. [19,22,23,25,33]), we have preferred to convert the statistical information into a continuous probability distribution which makes the problem size independent on the size of the sample. For this advantage we have to pay by more complicated formulae for objective function values and gradients. Growing sample sizes only influence these formulae but not the size of the problem. Using kernel estimators we capture the whole available statistical information. Clearly, since our sample sizes are in magnitudes of several hundreds, empirical measures would have led to linear programs that are intractable even when using the most advanced decomposition techniques. To arrive at moderate problem sizes one possibility is to decrease the number of mass points, for instance by importance (re)sampling [9].

After having selected proper candidates for the marginal distribution functions F_t $(t = 1,\dots,T)$ there is, in principle, no difficulty to solve the stochastic program (2.9). Of course, we ensure that the assumptions in (3.3) are met such that (2.9) becomes a convex optimization problem with linear constraints. In contrast to stochastic programs with more complicated recourse we can benefit from the simplicity of the second stage and obtain explicit formulas for the function Q in (3.2) and its gradient (cf. [18]). It holds

$$Q(\chi) = \sum_{t=1}^{T} Q_t(\chi_t),$$

$$Q_t(\chi_t) = q_t^+(d_t - \chi_t) - (q_t^+ + q_t^-)\int_{-\infty}^{\chi_t} (\tau - \chi_t)\,\mathrm{d}F_t(\tau), \tag{4.1}$$

where $d_t = \int_{-\infty}^{\infty} \tau dF_t(\tau)$ and, provided that F_t is continuous,

$$\frac{\mathrm{d}}{\mathrm{d}\chi_t} Q_t(\chi_t) = -q_t^+ + (q_t^+ + q_t^-)\,F_t(\chi_t). \tag{4.2}$$

Now let $t \in \{1,\dots,T\}$ and $\xi_{1t},\dots,\xi_{nt}$ be an independent sample of the electric power demand for the time interval t. The sample is extracted from the raw statistical data as described at the beginning of this section. The asymptotic results derived in section 3 (theorem 3.3, remark 3.7) justify to replace in our calculation the unknown distribution function F_t by the kernel estimate $\hat{F}_t^{(n)}$ (cf. (3.4))

$$\hat{F}_t^{(n)}(u) := \frac{1}{nb_n}\sum_{i=1}^{n}\int_{-\infty}^{u} k\left(\frac{\tau-\xi_{it}}{b_n}\right)\mathrm{d}\tau \quad (u \in \mathbb{R};\ n \in \mathbb{N}). \tag{4.3}$$

For notational comfort we introduce

$$\mathcal{K}_1(u) = \int_{-\infty}^{u} k(\tau)\mathrm{d}\tau \quad \text{and} \quad \mathcal{K}_2(u) = \int_{-\infty}^{u} \tau k(\tau)\mathrm{d}\tau.$$

In (4.1), we have to compute an integral $\int_{-\infty}^{u}(\tau - u)\mathrm{d}\hat{F}_t^{(n)}(\tau)$. Using the representation of $\hat{F}_t^{(n)}$ we obtain

$$\begin{aligned}
&\int_{-\infty}^{u}(\tau-u)\mathrm{d}\hat{F}_t^{(n)}(\tau)\\
&\quad = \frac{1}{nb_n}\sum_{i=1}^{n}\left\{\int_{-\infty}^{u}\tau k((\tau-\xi_{it})b_n^{-1})\mathrm{d}\tau - u\int_{-\infty}^{u}k((\tau-\xi_{it})b_n^{-1})\mathrm{d}\tau\right\}\\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\left\{\int_{-\infty}^{(u-\xi_{it})b_n^{-1}}(\tau^* b_n + \xi_{it})k(\tau^*)\mathrm{d}\tau^* - u\int_{-\infty}^{(u-\xi_{it})b_n^{-1}}k(\tau^*)\mathrm{d}\tau^*\right\},
\end{aligned}$$

where we have used the transformation $\tau^* = (\tau - \xi_{it})b_n^{-1}$. Therefore,

$$\hat{Q}^{(n)}(\chi) = \sum_{t=1}^{T}\hat{Q}_t^{(n)}(\chi_t),$$

$$\begin{aligned}
\hat{Q}_t^{(n)}(\chi_t) = q_t^+(\hat{d}_t - \chi_t) - (q_t^+ + q_t^-)\frac{1}{n}\sum_{i=1}^{n}\{&(\xi_{it}-\chi_t)\mathcal{K}_1((\chi_t-\xi_{it})b_n^{-1})\\
&+ b_n\mathcal{K}_2((\chi_t-\xi_{it})b_n^{-1})\},
\end{aligned} \tag{4.4}$$

where $\hat{d}_t = (1/n)\sum_{i=1}^{n}\xi_{it}$.

According to (4.2) the term for the (estimated) partial derivative reads

$$\frac{\mathrm{d}}{\mathrm{d}\chi_t}\hat{Q}_t^{(n)}(\chi_t) = -q_t^+ + (q_t^+ + q_t^-)\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_1((\chi_t-\xi_{it})b_n^{-1}). \tag{4.5}$$

Let us now discuss some aspects of efficiently calculating the expressions in (4.4) and (4.5).

(i) The function $\hat{Q}_t^{(n)}$ and its derivative do not directly depend on the kernel function k, but on the integrals $\mathcal{K}_1$ and $\mathcal{K}_2$. To reduce computation time we avoid numerical integration by selecting only such kernels k for which $\mathcal{K}_1$ and $\mathcal{K}_2$ can be computed explicitly. For the triangular kernel (cf. section 3) this leads to the following formulas:

$$\mathcal{K}_1(u) = \begin{cases} 0 & u \le -1, \\ \frac{1}{2} + u + \frac{1}{2}u^2 & -1 < u \le 0, \\ \frac{1}{2} - u + \frac{1}{2}u^2 & 0 < u < 1, \\ 1 & u \ge 1; \end{cases}$$

$$\mathcal{K}_2(u) = \begin{cases} 0 & |u| \ge 1, \\ -\frac{1}{6} + \frac{1}{2}u^2 + \frac{1}{3}u^3 & -1 < u \le 0, \\ -\frac{1}{6} + \frac{1}{2}u^2 - \frac{1}{3}u^3 & 0 < u < 1. \end{cases}$$

(ii) To calculate one function value $\hat{Q}^{(n)}(\chi)$ the integrals $\mathcal{K}_1$ and $\mathcal{K}_2$ must be computed at nT points. Transforming the sample $(\xi_{1t},\ldots,\xi_{nt})$ $(t = 1,\ldots,T)$ into ordered samples $(\tilde{\xi}_{1t},\ldots,\tilde{\xi}_{nt})$ such that $\tilde{\xi}_{1t} \le \tilde{\xi}_{2t} \le \ldots \le \tilde{\xi}_{nt}$ can save computation time. Indeed, if k is symmetric (as e.g. the triangular and Epanechnikov kernels in section 3) and $k(\tau) = 0$ for $|\tau| \ge a$ (for some $a > 0$), then

$$\mathcal{K}_1((\chi_t - \tilde{\xi}_{it})b_n^{-1}) = \begin{cases} 0 & \text{if } \tilde{\xi}_{it} \le \chi_t - ab_n, \\ 1 & \text{if } \tilde{\xi}_{it} \ge \chi_t + ab_n \end{cases}$$

and

$$\mathcal{K}_2((\chi_t - \tilde{\xi}_{it})b_n^{-1}) = 0 \qquad \text{if } \tilde{\xi}_{it} \notin [\chi_t - ab_n, \chi_t + ab_n].$$

Hence, to calculate (4.4) and (4.5) we pass over to the ordered sample and evaluate $\mathcal{K}_1$ and $\mathcal{K}_2$ only at those points for which

$$\chi_t - ab_n < \tilde{\xi}_{it} < \chi_t + ab_n \quad (t = 1,\ldots,T).$$

(iii) According to remarks 3.4 and 3.5 we select $b_n = b_{n,t} = 0.5\,\hat{\sigma}_t n^{-1/3}$, where $\hat{\sigma}_t$ denotes the sample standard deviation of $\xi_{1t},\ldots,\xi_{nt}$.

A program system STOCHOPT for estimating unknown distribution functions via kernel estimators and solving stochastic programs with simple recourse has been developed. It uses MINOS 5.1 [21] for the non-linear programming part and, hence, benefits from the sparsity and linearity of the constraints.

Let us mention that our approach to solving (2.9) via kernel estimators also applies if the second-stage is more general such that (2.9) becomes a stochastic program with (non-linear) convex simple recourse (cf. e.g. [27] for the linear/quadratic case). Again, the function Q is separable (cf. (4.1)) and explicit formulae for the objective function are available [27].

5. Test runs

In [16], first test runs with STOCHOPT for a model with a comparatively small number of decision variables were carried out. In the present paper we report on test runs with STOCHOPT for the much larger model (2.9). The model was validated by solving its full sized version with a time horizon of one day for several instances based on real data reflecting the energy situation in the eastern part of Germany during the time period from 1986 to 1989. The instances refer to different day categories arising in the analysis of the demand curves (recall the discussion at the beginning of this section). Our test runs were directed to measuring the (economic) impact of solving the stochastic program (2.9) instead of running a purely deterministic model where the random demand is replaced by its expectation $d \in \mathbb{R}^T$, where $d_t = \int_{\mathbb{R}} \tau \mathrm{d} F_t(\tau)$ $(t = 1,\dots,T)$. The latter would lead to the quadratic program

$$\min\{\, g(x) : x \in C,\ Ax = d \,\}, \tag{5.1}$$

where the notation is as in section 2. Let x_d denote an optimal solution to (5.1). If we plan the electricity production of our generation system according to the policy x_d, our expected costs for adjusting this policy to the actual demand amount to $Q(Ax_d) = Q(d)$ (cf. (2.3), (3.2)). Altogether we end up with the total costs $g(x_d) + Q(Ax_d)$. Of course, from a numerical point of view it needs much less effort to run the above procedure compared to solving the stochastic program (2.9) (no examination of Q is needed to find the policy!). On the other hand, we cannot expect that the generation policy x_d is optimal for (2.9). Let x_{opt} denote an optimal solution to (2.9). In the literature ([3]) the difference

$$\mathrm{VSS} = g(x_d) + Q(Ax_d) - g(x_{\mathrm{opt}}) - Q(Ax_{\mathrm{opt}})$$

is called the value of the stochastic solution (cf. also [4] for a recent application in financial planning). It reflects the benefit of solving the stochastic program (2.9) versus resorting to the deterministic procedure explained above.

Our test runs were based on the following power generation system:

number of thermal units (N_1): 26,
number of on-line thermal units (card I_t, $t = 1,\dots,T$): 24,
number of pumped storage plants (all on-line, N_2): 5,
number of energy contracts (N_3): 1,
number of (hourly) time intervals (T): 24.

For the penalty costs q^+, q^- in (2.9) we elaborate two cases. First we assume that there are no extra costs for overdispatching (i.e. $q_1^- = \dots = q_T^- = 0$) and a penalty for underdispatching that is constant in time (i.e. $q_1^+ = \dots = q_T^+ = q^{++}$ with some parameter $q^{++} > 0$). On the one hand, this is inconsistent since there is no penalty for overdispatch. On the other hand, the above mentioned power company had some difficulties to quantify costs for overdispatching (e.g. costs for frequency errors), such that putting $q^{-} = 0$ had a certain relevance. To see how the model behaves in the more realistic situation where $q^{-} \neq 0$ we ran a second series of tests where the overdispatching costs were set to 80% of the costs for underdispatching.

Of course, there is no problem in running further test series. One only has to adapt q^{+} and q^{-} properly.

Table 1 compares the amounts of accumulated under- and overdispatching for different penalty levels. It is seen that overdispatching vanishes as soon as it is substantially penalized.

Tables 2 and 3 display the value of the stochastic solution for different penalty levels.

Let us add a few comments on the test runs. We confine ourselves to the case $q^{-} = 0$. When evaluating the value of the stochastic solution it is useful to know that the average generation costs of the thermal power stations involved in the test run

Table 1

Accumulated under- and overdispatching for different penalty levels.

q^{++}	q^{--}	Accumulated Overload	Accumulated Underload
10	0	0	20766
10	8	0	21310
20	0	0	20768
20	16	0	20898
30	0	0	18936
30	24	0	19068
40	0	0	6576
40	32	0	8229
50	0	0	3312
50	40	0	5289
60	0	126	1735
60	48	0	4128
70	0	333	878
70	56	0	3343
80	0	588	398
80	64	0	2908
90	0	955	169
90	72	0	2522
100	0	1401	12
100	80	0	2163

Table 2

Comparison of optimal costs (no penalty for overdispatching).

Penalty costs for underdispatching q^{++}	$g(x_{\text{opt}}) + Q(Ax_{\text{opt}})$ $(\times 10^6)$	$g(x_d) + Q(Ax_d)$ $(\times 10^6)$	VSS $(\times 10^3)$
10	8.15900	8.69617	537.170
20	8.37030	8.72000	349.700
30	8.57512	8.74383	168.710
40	8.70952	8.76766	58.140
50	8.76486	8.79149	26.630
60	8.80322	8.81531	12.090
70	8.83292	8.83914	6.220
80	8.85794	8.86297	5.030

Table 3

Comparison of optimal costs (including penalty for overdispatching).

Penalty costs for underdispatching q^{++}	overdispatching q^{--}	$g(x_{\text{opt}}) + Q(Ax_{\text{opt}})$ $(\times 10^6)$	$g(x_d) + Q(Ax_d)$ $(\times 10^6)$	VSS $(\times 10^3)$
10	8	8.16110	8.71524	554.140
20	16	8.37529	8.75812	382.830
30	24	8.58288	8.80102	218.140
40	32	8.72565	8.84390	118.250
50	40	8.80028	8.88679	86.510
60	48	8.86154	8.92964	68.100
70	56	8.91678	8.97257	55.790
80	64	8.96879	9.01546	46.670

approximately corresponded to $q^{++} = 40$. Moderate variations around $q^{++} = 40$ may occur due to changing market situations. Of course, penalties below 30 and above 50 are purely academic. However, our results for these values also indicate the validity of our model: if q^{++} is small, cheap compensation clearly outperforms production and the "savings" in the last column are utopic, if q^{++} is large, the model tends to avoid infeasibilities and compensation is reduced to a negligible amount. The stochastic program (2.9) yields a generation policy that minimizes the sum of generation costs plus expected future costs for compensating a possible underdispatch. As an alternative, engineers often avoid the consideration of future compensation costs and add adjusted demand portions (about 3 to 5%) to load predictions in certain time intervals (e.g. during times of peak load). To see how this compares to (2.9) we have solved the quadratic program (5.1) with a demand portion of 3% added to the expected demand d during times of peak load (time intervals 9 to 14). The optimal solution is displayed in table 6 in the appendix. The optimal costs compare as follows optimal value of (2.9) (with $q^{++} = 40$): 8.70952×10^6;

optimal value of (5.1) (with adjusted demand d): 8.84922×10^6.

This corresponds to savings of about 1.6% when using the stochastic program (2.9).

Let us now turn to a comparison of generation policies. Table 4 displays the aggregated outputs for thermal units and pumped storage plants under varying compensation costs q^{++} and for the expected-value deterministic model (5.1).

Table 4

Comparison of aggregated generation policies ($q^{--} = 0$).

Stochastic power dispatch model (2.9)			
	Thermal units	Pumped storage plants generation	Pumped storage plants pumping
q^{++}	$\sum_{t=1}^{T} \sum_{i \in I_t} y^i_{t,\text{opt}}$	$\sum_{t=1}^{T} \sum_{j=1}^{N_2} s^j_{t,\text{opt}}$	$\sum_{t=1}^{T} \sum_{j=1}^{N_2} w^j_{t,\text{opt}}$
10	2.186670E+05	3204	4461
20	2.186670E+05	3203	4462
30	2.204970E+05	3204	4461
40	2.328440E+05	3166	4410
50	2.360800E+05	3085	4301
60	2.377810E+05	3086	4300
70	2.388450E+05	3085	4299
80	2.395810E+05	3084	4299
Expected-value deterministic model (5.1)			
	Thermal units	Pumped storage plants generation	Pumped storage plants pumping
	$\sum_{t=1}^{T} \sum_{i \in I_t} y^i_{t,\text{opt}}$	$\sum_{t=1}^{T} \sum_{j=1}^{N_2} s^j_{t,\text{opt}}$	$\sum_{t=1}^{T} \sum_{j=1}^{N_2} w^j_{t,\text{opt}}$
	2.393920E+05	3087	4303

If q^{++} is small (≤ 30), i.e. if compensation prices are undervalued in relation to production prices in the thermal units, then the thermal units will work at their lower output bounds since it is advantageous to use the compensation instead. Moreover, pumped storage plants will be heavily used since filling the upper dam is comparatively cheap. For $q^{++} = 40$, compensation and production prices are in a rough equilibrium which is reflected by a balanced relation between the outputs of thermal units and pumped storage plants. For $q^{++} \geq 50$ compensation becomes more expensive than production. Therefore, thermal units will act in a way that deviations between production and demand are avoided. According to the ranking of their cost functions, more and more thermal units are driven to their upper bounds when q^{++} is increasing. The generation policies for the pumped storage plants are getting very close to the policy occurring in the expected-value deterministic model (5.1) where the equilibrium between total output and demand is mandatory (constraint $Ax = d$).

To give an impression on detailed generation policies we refer to the appendix (tables 7–9) where optimal solutions for $q^{++} = 30, 40, 50$ are displayed.

We end this section by studying the model's behaviour under changes of the demand distribution with accent on changes of the sample sizes used for computing the kernel estimators. To this end, we resampled from the estimated marginal distribution functions $\hat{F}_t^{(n)}$. For different fixed sample sizes we applied our kernel estimation procedure and solved the resulting stochastic program. Per sample size ten different samples were processed and the mean as well as the standard deviation of optimal values were computed. Figures 2 and 3 display the dependence of these quantities on the sample size. Both figures show significant instabilities for sample sizes below 100. On the other hand, for sample sizes above 300 the means stabilize and the standard deviations are acceptably small. This indicates that, if available, larger samples should be fully exploited for the estimation of the demand distribution.

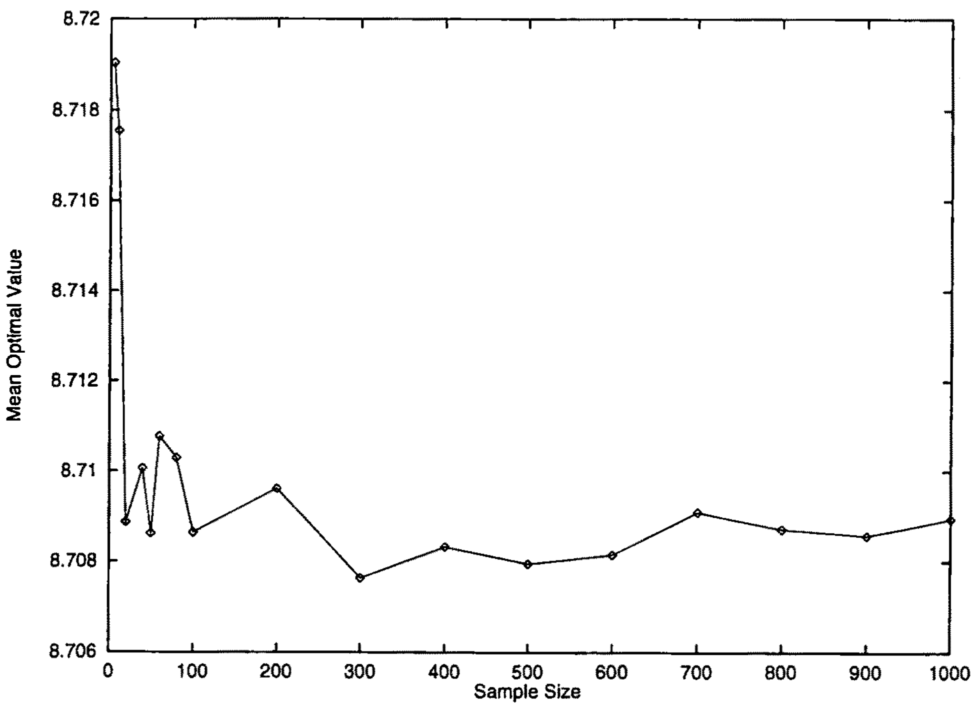

Figure 2. Mean optimal values.

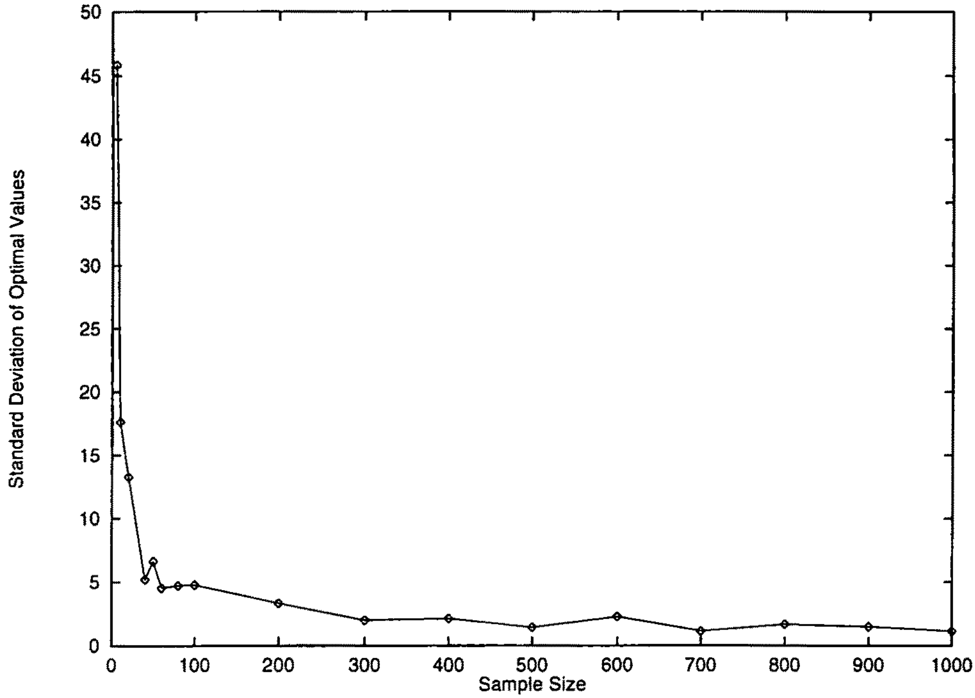

Figure 3. Standard deviation of optimal values.

Appendix

Table 5(a)

Stochastic power dispatch model ($q^{++} = 40$, $q^{--} = 0$), optimal outputs of the thermal units.

Timestep	Thermal unit no.											
T	1	2	3	4	5	6	7	8	9	10	11	12
1	285	350	310	485	230	390	850	850	224	80	350	700
2	285	350	310	485	230	410	850	850	224	80	350	700
3	285	350	310	485	230	440	850	850	224	80	350	700
4	285	350	310	485	230	460	850	850	224	80	350	700
5	285	350	310	485	230	510	850	850	224	80	350	700
6	285	350	310	485	230	540	850	850	·224	80	350	700
7	285	350	310	485	230	540	850	850	224	80	350	700
8	285	350	310	485	330	540	860	850	224	80	400	700
9	285	420	310	555	380	540	880	1000	224	80	400	932
10	285	420	310	555	380	540	910	1000	224	80	400	932
11	285	420	310	555	380	540	930	1000	224	80	400	932
12	285	420	310	555	380	540	980	1000	224	80	400	932
13	285	420	310	555	380	540	1010	1000	224	80	400	932
14	285	420	310	555	380	540	1010	1000	224	80	400	932
15	285	383	310	555	380	540	1010	1000	224	80	400	932

. . . continues

Table 5(a) (continued)

Timestep	Thermal unit no.											
T	1	2	3	4	5	6	7	8	9	10	11	12
16	285	350	310	513	380	540	1010	1000	224	80	400	807
17	285	350	310	485	380	540	1010	1000	224	80	400	784
18	285	350	310	485	380	540	1010	850	224	80	400	700
19	285	350	310	485	330	540	1010	850	224	80	400	700
20	285	350	310	485	330	540	1010	850	224	80	400	700
21	285	350	310	485	380	540	1010	971	224	80	400	700
22	285	350	310	485	330	540	1010	850	224	80	400	700
23	285	350	310	485	330	540	1010	850	224	80	400	700
24	285	350	310	485	330	540	1010	850	224	80	400	700

Table 5(b)

Stochastic power dispatch model ($q^{++} = 40$, $q^{--} = 0$), optimal outputs of the thermal units.

Timestep	Thermal unit no.											
T	13	14	15	16	17	18	19	20	21	22	23	24
1	810	700	385	720	270	93	161	142	108	21	13	265
2	810	700	385	660	270	93	161	142	108	21	13	265
3	810	700	385	660	270	93	161	142	108	21	13	265
4	810	700	385	660	270	93	161	142	108	21	13	300
5	810	700	385	660	270	93	161	142	108	21	13	300
6	810	700	385	660	270	93	161	142	108	21	13	300
7	810	700	395	660	270	93	161	142	108	21	13	300
8	987	700	395	770	297	93	161	142	108	21	13	305
9	1000	1000	395	770	297	93	161	142	108	21	13	305
10	1000	1000	395	770	297	93	161	142	108	21	13	305
11	1000	1000	395	770	297	93	161	142	108	21	13	305
12	1000	1000	395	770	297	93	161	142	108	21	13	305
13	1000	1000	395	780	297	93	161	142	108	21	13	305
14	1000	1000	395	770	297	93	161	142	108	21	13	305
15	1000	1000	395	770	297	93	161	142	108	21	13	300
16	1000	1000	395	770	297	93	161	142	108	21	13	300
17	1000	1000	395	770	297	93	161	142	108	21	13	300
18	1000	949	395	770	297	93	161	142	108	21	13	300
19	1000	825	395	770	297	93	161	142	108	21	13	300
20	1000	960	395	720	297	93	161	142	108	21	13	300
21	1000	1000	395	720	297	93	161	142	108	21	13	300
22	1000	1000	395	720	297	93	161	142	108	21	13	300
23	1000	1000	395	720	297	93	161	142	108	21	13	300
24	1000	779	395	720	297	93	161	142	108	21	13	300

Table 6

Expected-value deterministic model with reserve adjustment: optimal solution.

Time	Optimal tender	Thermal units	Pumped storage plants generation	Pumped storage plants pumping	Energy contract
t	$[Ax_{\mathrm{opt}}]_t$	$\sum_{i \in I_t} y^i_{t,\mathrm{opt}}$	$\sum_{j=1}^{N_2} s^j_{t,\mathrm{opt}}$	$\sum_{j=1}^{N_2} w^j_{t,\mathrm{opt}}$	$z^1_{t,\mathrm{opt}}$
1	9067	9017	0	0	50
2	8460	8799	0	389	50
3	8160	8829	0	719	50
4	8052	8884	0	882	50
5	8049	8947	0	948	50
6	8246	8964	0	768	50
7	8204	8987	0	833	50
8	9686	9636	0	0	50
9	11656	10877	729	0	50
10	11968	10907	1111	0	–50
11	11767	10927	890	0	–50
12	12089	10977	1062	0	50
13	12170	11017	1103	0	50
14	11768	11007	711	0	50
15	10872	10922	0	0	–50
16	10593	10643	0	0	–50
17	10582	10632	0	0	–50
18	10077	10586	0	459	–50
19	9961	10586	0	575	–50
20	9971	10536	0	515	–50
21	10285	10536	0	201	–50
22	10088	10536	0	398	–50
23	10047	10536	0	439	–50
24	9761	10536	0	725	–50
		Optimal value:			8.849223E+06

Table 7

Stochastic power dispatch model: optimal solution for $q^{++} = 30$, $q^{--} = 0$.

Time	Optimal tender	Thermal units	Pumped storage plants generation	Pumped storage plants pumping	Energy contract
t	$[Ax_{\text{opt}}]_t$	$\sum_{i \in I_t} y^i_{t,\text{opt}}$	$\sum_{j=1}^{N_2} s^j_{t,\text{opt}}$	$\sum_{j=1}^{N_2} w^j_{t,\text{opt}}$	$z^1_{t,\text{opt}}$
1	8900	8732	118	0	50
2	8417	8752	0	385	50
3	8121	8782	0	711	50
4	8018	8837	0	869	50
5	8011	8887	0	926	50
6	8209	8917	0	758	50
7	8165	8927	0	812	50
8	9279	9229	0	0	50
9	9909	9249	610	0	50
10	10667	9279	1438	0	−50
11	10287	9299	1038	0	−50
12	9299	9349	0	0	−50
13	9339	9389	0	0	−50
14	9329	9379	0	0	−50
15	9424	9374	0	0	50
16	9374	9374	0	0	0
17	9374	9374	0	0	0
18	9324	9374	0	0	−50
19	9324	9374	0	0	−50
20	9274	9324	0	0	−50
21	9324	9324	0	0	0
22	9274	9324	0	0	−50
23	9324	9324	0	0	0
24	9274	9324	0	0	−50
		Optimal value:			8.575120E+06

Table 8

Stochastic power dispatch model: optimal solution for $q^{++} = 40$, $q^{--} = 0$.

Time	Optimal tender	Thermal units	Pumped storage plants generation	Pumped storage plants pumping	Energy contract
t	$[Ax_{\text{opt}}]_t$	$\sum_{i \in I_t} y^i_{t,\text{opt}}$	$\sum_{j=1}^{N_2} s^j_{t,\text{opt}}$	$\sum_{j=1}^{N_2} w^j_{t,\text{opt}}$	$z^1_{t,\text{opt}}$
1	8923	8792	81	0	50
2	8425	8752	0	377	50
3	8129	8782	0	703	50
4	8026	8837	0	861	50
5	8019	8887	0	918	50
6	8217	8917	0	750	50
7	8176	8927	0	801	50
8	9456	9406	0	0	50
9	10710	10311	349	0	50
10	11044	10341	703	0	0
11	10798	10361	487	0	−50
12	11098	10411	737	0	−50
13	11105	10451	604	0	50
14	10696	10441	205	0	50
15	10399	10399	0	0	0
16	10149	10199	0	0	−50
17	10098	10148	0	0	−50
18	9763	9813	0	0	−50
19	9639	9689	0	0	−50
20	9724	9774	0	0	−50
21	9935	9985	0	0	−50
22	9764	9814	0	0	−50
23	9764	9814	0	0	−50
24	9543	9593	0	0	−50
		Optimal value:			8.709516E+06

Table 9

Stochastic power dispatch model: optimal solution for $q^{++} = 50$, $q^{--} = 0$.

Time	Optimal tender	Thermal units	Pumped storage plants generation	Pumped storage plants pumping	Energy contract
t	$[Ax_{\mathrm{opt}}]_t$	$\sum_{i \in I_t} y^i_{t,\mathrm{opt}}$	$\sum_{j=1}^{N_2} s^j_{t,\mathrm{opt}}$	$\sum_{j=1}^{N_2} w^j_{t,\mathrm{opt}}$	$z^1_{t,\mathrm{opt}}$
1	8987	8937	0	0	50
2	8439	8752	0	363	50
3	8147	8782	0	685	50
4	8043	8837	0	844	50
5	8040	8887	0	897	50
6	8234	8917	0	733	50
7	8198	8927	0	779	50
8	9589	9539	0	0	50
9	10917	10523	344	0	50
10	11200	10553	686	0	−39
11	10997	10573	474	0	−50
12	11300	10623	627	0	50
13	11364	10663	651	0	50
14	10956	10653	303	0	0
15	10637	10648	0	0	−11
16	10384	10434	0	0	−50
17	10357	10407	0	0	−50
18	9874	9924	0	0	−50
19	9814	9864	0	0	−50
20	9802	9852	0	0	−50
21	10100	10150	0	0	−50
22	9932	9982	0	0	−50
23	9895	9945	0	0	−50
24	9658	9708	0	0	−50
		Optimal value:			8.764862E+06

Acknowledgements

The authors wish to thank Dr. Heinz Balzer (Vereinigte Energiewerke AG, Berlin) for the fruitful cooperation. Further thanks are due to the referees and the guest editor for valuable comments that led to improvements over the first draft of this paper. Computations were partially carried out by the first author while at the Delft University of Technology (The Netherlands).

References

[1] A. Azzalini, A note on the estimation of a distribution function and quantiles by a kernel method, Biometrika 68(1981)326–328.
[2] R. Bacher, Power system models, objectives and constraints in optimal power flow calculations, in [14], pp. 159–198.
[3] J.R. Birge, The value of the stochastic solution in stochastic linear programs with fixed recourse, Math. Progr. 24(1982)314–325.
[4] J.R. Birge, Stochastic programming: Optimizing the uncertain, The University of Michigan, Department of Industrial and Operations Engineering, Technical Report 92–24 (1992).
[5] A. Berlinet, Hierarchies of higher order kernels, Prob. Theory Related Fields 94(1993)489–504.
[6] P.P.J. van den Bosch, Optimal static dispatch with linear, quadratic and nonlinear functions of the fuel costs, IEEE Trans. Power Apparatus Systems 104(1985)3402–3408.
[7] J. Böttcher, *Stochastische lineare Programme mit Kompensation*, Mathematical Systems in Economics, Vol. 115 (Athenäum, Frankfurt am Main, 1989).
[8] D.W. Bunn and S.N. Paschentis, Development of a stochastic model for the economic dispatch of electric power, Euro. J. Oper. Res. 27(1986)179–191.
[9] G.B. Dantzig and P.W. Glynn, Parallel processors for planning under uncertainty, Ann. Oper. Res. 22(1990)1–21.
[10] L. Devroye, *A Course in Density Estimation* (Birkhäuser, Boston, 1987).
[11] J. Dupačová, Applications of stochastic programming under incomplete information, J. Comp. Appl. Math. 56(1994)113–125.
[12] Y.M. Ermoliev and V.I. Norkin, Normalized convergence in stochastic optimization, Ann. Oper. Res. 30(1991)187–198.
[13] Y.M. Ermoliev and R.J-B Wets (eds.), *Numerical Techniques for Stochastic Optimization* (Springer, Berlin, 1988).
[14] K. Frauendorfer, H. Glavitsch and R. Bacher (eds.), *Optimization in Planning and Operation of Electric Power Systems*, Lecture Notes of a SVOR/ASRO-Tutorial, Thun, Switzerland, October 14-16, 1992 (Physica-Verlag, 1993).
[15] H. Glavitsch, Use of linear and quadratic programming techniques in exploiting the nonlinear features of the optimal power flow, in [14], pp. 199–235.
[16] N. Gröwe and W. Römisch, A stochastic programming model for optimal power dispatch: Stability and numerical treatment, in: *Stochastic Optimization*, ed. K. Marti, Lecture Notes in Economics and Mathematical Systems, Vol. 379 (Springer, Berlin, 1992) pp. 111–139.
[17] J. Guddat, W. Römisch and R. Schultz, Some application of mathematical programming techniques in optimal power dispatch, Computing 49(1992)193–200.
[18] P. Kall, *Stochastic Linear Programming* (Springer, Berlin, 1976).
[19] P. Kall and D. Stoyan, Solving stochastic programming problems with recourse including error bounds, Optimization 13(1982)431–447.
[20] K. Mai and J. Polzehl, Regression-based short term prediction of electrical load for a power system, Humboldt-Universität Berlin, Fachbereich Mathematik, Preprint Nr. 91-14 (1991).

[21] B.A. Murtagh and M.A. Saunders, MINOS 5.1 User's Guide, System Optimization Laboratory, Stanford University (1987).
[22] J.L. Nazareth, Design and implementation of a stochastic programming optimizer with recourse and tenders, in [13], pp. 273–294.
[23] J.L. Nazareth and R.J-B Wets, Algorithms for stochastic programs: The case of nonstochastic tenders, Math. Progr. Study 28(1986)1–28.
[24] B.L.S. Prakasa Rao, *Nonparametric Functional Estimation* (Academic Press, New York, 1983).
[25] A. Prékopa, Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probability distribution, Zeits. Oper. Res. 34(1990)441–461.
[26] R.T. Rockafellar, Integral functionals, normal integrands and measurable selections, in: *Nonlinear Operators and the Calculus of Variations*, ed. G.P. Gossez et al., Lecture Notes in Mathematics, Vol. 543 (Springer, New York, 1976) pp. 157–207.
[27] R.T. Rockafellar and R.J-B Wets, A dual solution procedure for quadratic stochastic programs with simple recourse, in: *Numerical Methods*, ed. V. Pereyra and A. Reinoza, Lecture Notes in Mathematics, Vol. 1005 (Springer, Berlin, 1983) pp. 252–265.
[28] W. Römisch and R. Schultz, Stability analysis for stochastic programs, Ann. Oper. Res. 30(1991) 241–266.
[29] W. Römisch and R. Schultz, Lipschitz stability for stochastic programs with complete recourse, SIAM J. Optim., to appear.
[30] R.J. Serfling, *Approximation Theorems of Mathematical Statistics* (Wiley, New York, 1980).
[31] G.R. Shorack and J.A. Wellner, *Empirical Processes with Applications to Statistics* (Wiley, New York, 1986).
[32] H. Wacker (ed.), *Applied Optimization Techniques in Energy Problems* (Teubner, Stuttgart, 1985).
[33] R.J-B Wets, Solving stochastic problems with simple recourse, Stochastics 10(1983)219–242.
[34] R.J-B Wets, Stochastic programming, in: *Handbooks in Operations Research and Management Science*, Vol. 1, *Optimization*, ed. G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd (North-Holland, Amsterdam, 1989) pp. 573–629.
[35] B.B. Winter, Convergence rate of perturbed empirical distribution functions, J. Appl. Prob. 16(1979)163–173.