PERGAMON

Intl. Trans. in Op. Res. 7 (2000) 211–230

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

www.elsevier.com/locate/orms

A scenario-based stochastic programming model for water supplies from the highland lakes

David W. Watkins Jr[a,*], Daene C. McKinney[a], Leon S. Lasdon[b], Soren S. Nielsen[c], Quentin W. Martin[d]

[a]*Department of Civil Engineering, The University of Texas at Austin, Austin, TX 78712, USA*
[b]*Department of Management Science and Information Systems, The University of Texas at Austin, Austin, TX 78712, USA*
[c]*Department of Mathematics, University of Copenhagen, DK-2100 Copenhagen 0, Denmark*
[d]*Lower Colorado River Authority, Austin, TX 78767, USA*

Received 4 January 1999; accepted 15 July 1999

Abstract

A scenario-based, multistage stochastic programming model is developed for the management of the Highland Lakes by the Lower Colorado River Authority (LCRA) in Central Texas. The model explicitly considers two objectives: (1) maximize the expected revenue from the sale of interruptible water while reliably maintaining firm water supply, and (2) maximize recreational benefits. Input data can be represented by a scenario tree, built empirically from a segment of the historical flow record. Thirty-scenario instances of the model are solved using both a primal simplex method and Benders decomposition, and results show that the first-stage ('here and now') decision of how much interruptible water to contract for the coming year is highly dependent on the initial (current) reservoir storage levels. Sensitivity analysis indicates that model results can be improved by using a scenario generation technique which better preserves the serial correlation of flows. Ultimately, it is hoped that use of the model will improve the LCRA's operational practices by helping to identify flexible policies that appropriately hedge against unfavorable inflow scenarios. © 2000 IFORS. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Stochastic programming; Water resources management; Scenario generation; Decision making under uncertainty

* Corresponding author. Michigan Technological University, Houghton, MI 49931, USA.
E-mail address: dwatkins@mtu.edu (D.W. Watkins Jr).

PII: S0969-6016(99)00021-0

1. Introduction

Until recently, surface waters in Texas have been managed primarily by individual river authorities with the sole purpose of m'eeting demands within their own basin. Now, due to rapid population growth in some areas with limited supplies, there is a growing trend toward inter-basin planning and management. An example of this trend is a recent agreement between the Lower Colorado River Authority (LCRA) and the Brazos River Authority to cooperate in water planning in Central Texas (Two Water Authorities Pledge to Cooperate, 1995). Another is the Trans-Texas Water Program, led by the Texas Water Development Board and sponsored by several of the state's river authorities, which includes planning studies of ways to meet growing water demands in cities such as Corpus Christi, Houston, and San Antonio. As suggested by the name of this program, inter-basin transfers are among the options being considered. However, it is likely that such transfers will occur only if they do not jeopardize benefits in the basin of origin.

The LCRA is a water conservation and reclamation district cestablished by the Texas Legislature in 1934. The district, shown in Fig. 1, is responsible for the control, storage,

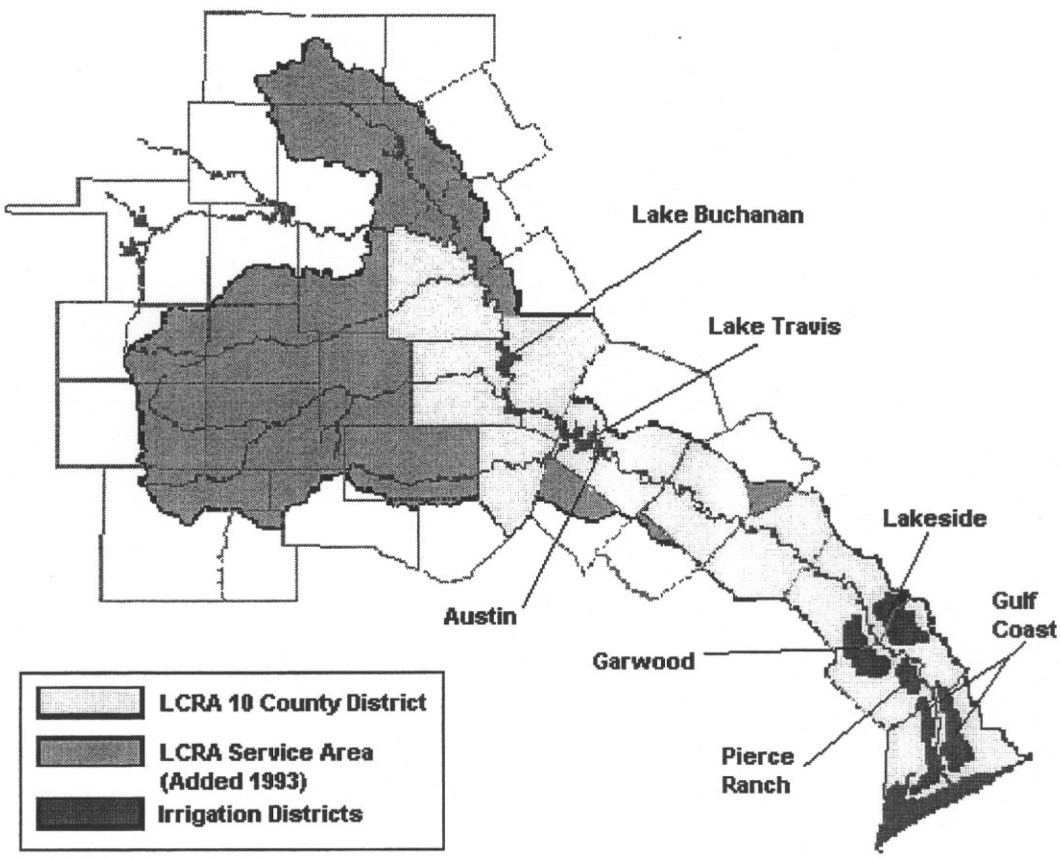

Fig. 1. The lower Colorado river authority district.

preservation, and distribution of the waters of the Colorado River in order to provide benefits such as flood control, electric power generation, water supply, recreation, and environmental protection (LCRA, 1993). To accomplish these tasks, the LCRA owns and operates a series of six reservoirs along the lower Colorado River. Only two of these reservoirs (Lakes Buchanan and Travis) provide storage for water supply; the other four are essentially constant level pools providing hydropower and recreational benefits.

At the beginning of each year, based on storage levels in Lakes Buchanan and Travis, the LCRA determines the amount of water available to meet firm and interruptible water demands in the coming year. Firm water is that which is diverted from storage under a contract or resolution issued by the LCRA Board to high priority users such as the City of Austin. Interruptible water contracts are issued on a shorter time scale (typically one year) with the condition that supplies may be interrupted or curtailed in the event that firm supplies become endangered. In allocating interruptible water, priority is given to irrigation operations downstream of Austin on the Texas Gulf Coastal Plain. If it is projected that the availability of interruptible water exceeds these irrigation needs, annual contracts can then be made with other entities within the Lower Colorado basin. Finally, if the consideration of inter-basin transfers becomes a policy of the LCRA Board, these may also be contracted at this time.

2. Stochastic programming approach

A scenario-based, multi-stage linear programming model is developed to support the LCRA's decision of how much interruptible water to contract for delivery in the coming year. In contrast to deterministic models, which select values of decision variables with perfect knowledge of the future, scenario-based stochastic programming models consider a number of possible futures. Essentially, stochastic programming models support the 'here and now' decision, while providing a number of 'wait and see' strategies dependent on which scenario unfolds. As such, an important requirement of these models is *non-anticipativity*: in each stage, decisions must be made without knowledge of the realizations of random variables in future stages.

Stochastic optimization models have been used in a number of reservoir operations studies. Alternative approaches to scenario-based stochastic programming include stochastic dynamic programming (e.g., Kelman et al., 1990) and stochastic optimal control (e.g., Hooper et al., 1991). The primary advantage of scenario-based stochastic programming over these other approaches is the flexibility it offers in modeling the decision process and defining scenarios, particularly if the state dimension is high. One disadvantage, however, is that the resulting math programming models can be very large, therefore requiring special solution algorithms. Previous applications of two-stage and multi-stage stochastic programming to reservoir management (Pereira and Pinto, 1985, 1991; Jacobs et al., 1995) have typically applied algorithms based on the L-shaped method (Benders, 1962; Van Slyke and Wets, 1969). This method is useful because it allows the large-scale problem to be decomposed by scenario. Moreover, using the technique in a nested manner allows multi-stage problems to be decomposed by both scenario and decision period (Birge, 1985; Gassman, 1990). Alternative solution techniques for multi-stage stochastic programs include regularized decomposition with constraint dropping (Ruszczynski, 1986), decomposition via augmented Lagrangian methods (Rockafellar and Wets, 1991; Mulvey and Ruszczynski, 1995), and direct solution by interior point methods (Carpenter et al., 1991; Lustig et al., 1994).

Typically, as done herein, the underlying probability distribution is sampled (i.e., scenarios are generated) prior to the solution procedure. Other approaches, such as sampling-based cutting plane methods (Dantzig and Glynn, 1990; Higle and Sen, 1991; Infanger, 1992) and stochastic quasi-gradient algorithms (Ermoliev, 1988; Gaivoronski, 1988), are based on 'internal' sampling — new scenarios are generated in each iteration of the algorithm. For more discussion of sampling-based methods and other modern developments in stochastic programming, the reader is referred to the review paper of Ruszczynski (1997) and the texts of Infanger (1994), Higle and Sen (1996) and Birge and Louveaux (1997).

3. Model conceptualization and formulation

The water supply planning model is based on a simplified representation of the Highland Lakes system consisting of only the two storage reservoirs and two tributaries of the Colorado River, as shown in Fig. 2. The first-stage decision in the model is the quantity of interruptible water to contract for the coming year. Subsequent-stage decisions involve monthly reservoir releases to meet both firm and interruptible demands, as well as annual water contracts signed in future years. Since maximizing the sale of interruptible water will likely conflict with other planning goals such as maintaining lake levels for recreational use and storage for firm water supply, the objective function is a weighted combination of two goals: (1) maximize revenue from interruptible water sales minus penalties for not meeting firm demands and interruptible contracts, and (2) maximize recreational benefits. Other purposes of the LCRA, such as hydropower generation and flood control, are not considered directly. Modeling to support operation for these purposes must be done using a time step smaller than one month —

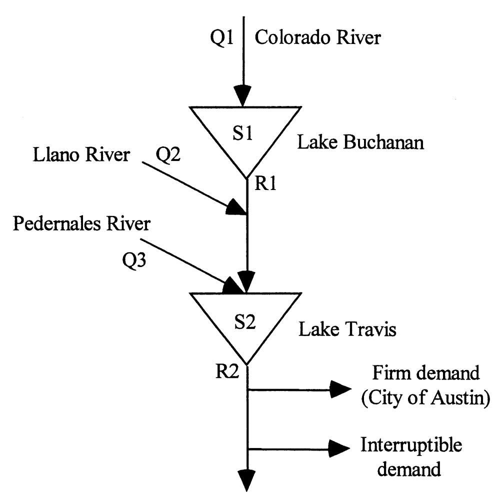

Fig. 2. Schematic of the highland lakes system.

possibly a day or even a few hours. Nonetheless, flood control is addressed implicitly in the model by not allowing carry-over (month-to-month) storage in the flood control pool, and Martin (1995) has developed a methodology for meeting peak power demands without significantly affecting long-term water availability or lake levels.

A scenario in the model is considered to be a sequence of monthly 'available' flow values for each of the above inflows, *Q*1, *Q*2, *Q*3, which is representative of flows which could occur in the future. Available flows are defined as those which are in excess of environmental needs and water rights senior to the LCRA's, as determined by a daily simulation model (LCRA, 1993). For simplification, firm monthly demands, the fraction of contracted interruptible water used each month, and monthly evaporation rates are considered deterministic and set to average values.

Monthly inflow data for *Q*1, *Q*2, *Q*3, for the period 1941–1965, which includes the 1947–1956 drought of record, were used to generate scenarios. The cumulative distribution of annual available flows over this 25-year period (the sum of *Q*1 through *Q*3) is shown in Fig. 3. As a first approximation, a scenario tree was constructed with five branches in the second stage (year 1), three branches in the third stage (year 2), and two branches in the fourth stage (years 3–5), as shown in Fig. 4. Flows were assigned to the branches according to the following procedure. First, sort the 25 total annual flows in order from lowest to highest. For year 1, select an annual flow vector [*Q*1, *Q*2, *Q*3] from each 20% quantile (approximately the 10, 30, 50, 70, and 90% flows). Assign the vectors of monthly flows associated with each annual flow vector to the five first-year branches emanating from the root (topmost) node of the tree. For year 2, for each of the five first-stage branches, randomly select an annual flow vector from the 0–35, 35–65, and 65–100% ranges of the annual flow distribution, and assign the three associated monthly flow vectors to each set of the three second-year branches. Proceed similarly for each set of two fourth-stage branches, choosing randomly from the 0–50 and 50–100% ranges of the distribution of three-year flows. For example, scenario 1 contains the flows from years 1950, 1962, and 1941–1943. Since this procedure is based on random sampling from

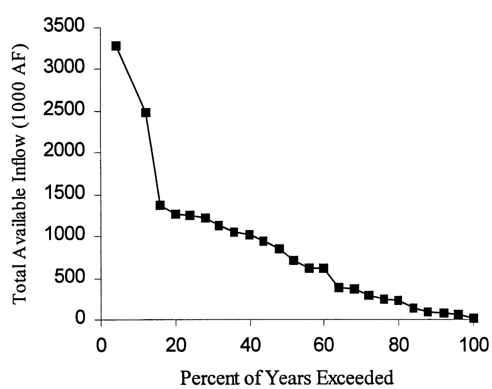

Fig. 3. Distribution of total annual available inflows to the Highland Lakes, 1941–1965. Min = 16,203 AF; median = 622,268 AF; max = 3,276,807 AF (1 AF = 1234 m^3).

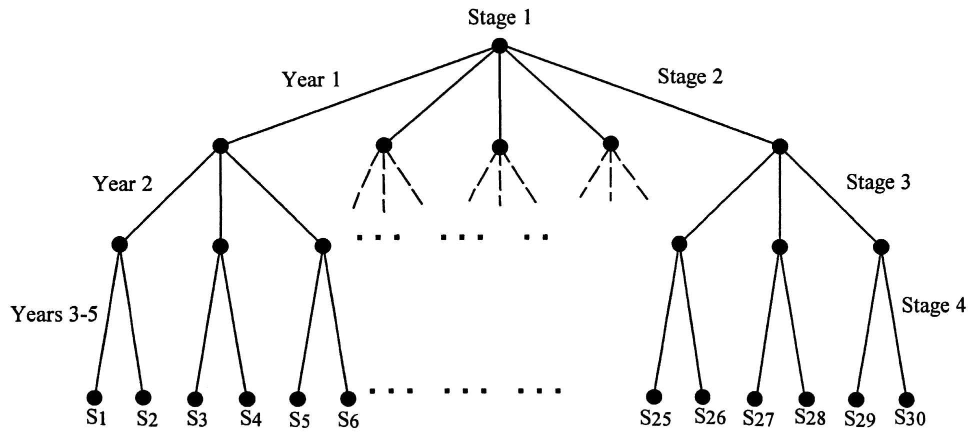

Fig. 4. Scenario tree for the multistage reservoir model.

equal-size ranges of the probability distribution, it is assumed that each of the scenarios (paths through the tree) is equally likely.

A favorable characteristic of this sampling procedure is that it preserves the historical spatial correlation of inflows. Monthly serial correlation is also preserved, but annual serial correlation is not, since the sequences in each stage are sampled independently of those in the preceding stage. There is evidence that such random ordering can lead to less severe inflow sequences than the historical order (Vaugh and Maidment, 1987). Furthermore, there is no guarantee that 30 scenarios assembled in this particular manner can adequately represent the inflow uncertainty. To address these potential limitations, sensitivity analyses are presented in a later section.

Along with these concerns, it is also important to note the difference between the decision stages in the model ('here and now', year 1, year 2, and years 3–5) and the operating periods (each month). Implicitly, the model assumes that the flows for the rest of the year become known immediately after making the annual contract decision in the beginning of January. This of course is not realistic, but it is believed to be a reasonable simplification since releases are driven by annual interruptible contracts and the (deterministic) firm demands. With this assumption, the multistage stochastic program is now formulated. Indices used in the formulation are listed in Table 1, variables in Table 2, and parameters in Table 3.

Table 1
Indices used in model formulation

Index	Definition
t	Operating time period (monthly), $1, 2,\ldots,T$ ($T = 60$)
τ	Contract time period (annual), $1, 2, \ldots, \frac{\lceil T \rceil}{\lfloor 12 \rfloor}$
s	Hydrologic scenario, $1, 2,\ldots,S$ ($S = 30$)

Table 2
Model variables

Variable	Definition
X_1	First-stage annual interruptible water contract amount (AF/year)
$X_{\tau s}$	τ-stage annual interruptible water contract amount, $\tau = 2, 3, 4$
$R1_{ts}$, $R2_{ts}$	Releases from lakes in period t, scenario s (AF)
$S1_{ts}$, $S2_{ts}$	Storage levels in lakes in period t, scenario s (AF)
UF_{ts}, UI_{ts}	Firm and interruptible supply deficits in period t, scenario s (AF)
$REC1_{ts}$, $REC2_{ts}$	Recreational benefit of water in lakes in period t, scenario s (\$)

Maximize

$$Z = c_o X_1 + \frac{1}{S}\sum_{s=1}^{S}\left[\sum_{\tau=2}^{[T/12]} c_o X_{\tau s} - \sum_{t=1}^{T}\left(c_i UI_{ts} + c_f UF_{ts}\right) + \omega \sum_{t=1}^{T}(REC1_{ts} + REC2_{ts})\right] \tag{1}$$

Subject to

$$S1_{t+1,\,s} = S1_{ts} + Q1_{ts} - \left[A1\left(\frac{S1_{t+1,\,s} + S1_{ts}}{2}\right) + B1\right]e1_t - R1_{ts} \quad \forall t,\ \forall s \tag{2a}$$

$$S2_{t+1,\,s} = S2_{ts} + Q2_{ts} + Q3_{ts} + R1_{ts} - \left[A2\left(\frac{S2_{t+1,\,s} + S2_{ts}}{2}\right) + B2\right]e2_t - R2_{ts} \tag{2b}$$

Table 3
Model parameters

Parameter	Definition
c_o	Price for selling interruptible water (\$/AF)
c_i	Cost (penalty) for not meeting interruptible contract (\$/AF)
c_f	Cost (penalty) for not meeting firm demand (\$/AF)
$K1$, $K2$	Capacities of Lakes Buchanan and Travis (AF)
$S1_0$, $S2_0$	Initial storages in Lakes Buchanan and Travis (AF)
$S1_{\mathrm{T}}$, $S2_{\mathrm{T}}$	Final storage targets for Lakes Buchanan and Travis (AF)
$RESV$	Reserve storage (AF)
$FIRM$	Firm annual demand (AF/year)
fi_t	Monthly fraction of annual interruptible contract
ff_t	Monthly fraction of annual firm demand
$e1_t$, $e2_t$	Evaporation constants for lakes in month t
$A1$, $A2$, $B1$, $B2$	Storage-area coefficients
$F1_t$, $F2_t$, $G1_t$, $G2_t$	Recreational benefit coefficients
$Q1_{ts}$, $Q2_{ts}$, $Q3_{ts}$	Random inflows to lakes in period t, scenario s (AF)
α, β	Equity and contract stability coefficients
ω	Weight placed on recreation benefits

$$R2_{ts} + UF_{ts} \geq ff_t FIRM \tag{3a}$$

$$R2_{ts} + UF_{ts} + UI_{ts} \geq ff_t FIRM + fi_t X_{\tau s} \quad \forall t,\ \forall s,\ \tau \in \left[\frac{t}{12}\right] \tag{3b}$$

$$RECi_{ts} \leq -Fi_t + Gi_t\left(\frac{Si_{ts} + Si_{t+1,\ s}}{2}\right) \quad i = 1, 2,\ \forall t,\ \forall s \tag{4}$$

$$RECi_{ts} \leq RMAXi \quad i = 1, 2,\ \forall t,\ \forall s \tag{5}$$

$$Si_{ts} \leq Ki \quad i = 1, 2,\ \forall t,\ \forall s \tag{6}$$

$$S1_{ts} + S2_{ts} \geq RESV \quad \forall t,\ \forall s \tag{7}$$

$$\frac{S1_{ts}}{K1} \geq \frac{(\alpha S2_{ts})}{K2} \quad \forall t,\ \forall s \tag{8}$$

$$\frac{1}{S}\sum_{s \in S} Si_{Ts} \geq Si_T \quad i = 1, 2 \tag{9}$$

$$X_{\tau,\ s} \geq \beta X_{\tau-1,\ s} \quad \tau \in \left[\frac{t}{12}\right], \quad t \geq 12,\ \forall s \tag{10}$$

$$X_{\tau,\ s} = X_{\tau,\ s'}, \quad R1_{t,\ s'} = R1_{t,\ s'}, \quad R2_{t,\ s} = R2_{t,\ s'} \quad \text{if } s(t) = s'(t),\ \forall t \tag{11}$$

$$X_\tau,\ S1_{ts},\ S2_{ts},\ R1_{ts},\ R2_{ts},\ UF_{ts},\ UI_{ts} \geq 0 \quad \forall t,\ \forall s \tag{12}$$

As discussed earlier, the objective function (1) includes two goals: maximize the revenue from the first year contract plus the expected revenue from future contracts, minus the expected cost of shortfalls; and maximize recreational benefits. In this case, the two objectives cannot be summed directly, even though a monetary value is assigned to each, for a number of reasons. First, the LCRA attempts to set water prices so that their operating expenses are just recovered, and these expenses may be significantly less than the value of the water to its customers. In fact, the penalty cost for contract shortfalls provides an upper bound on the value of water for irrigation; it equals the 'fine' that the LCRA would pay farmers to compensate them for lost crops. Second, irrigation is generally given priority over recreational uses for historical reasons (LCRA, 1993). Thus, the objective function includes the parameter ω, which represents the value of recreation relative to interruptible water supply, and the objective Z is the weighted sum of the two objective function values.

Constraints (2) are mass balance equations representing the physics of the hydrologic system. In each time period, inflows minus releases and evaporation equal the change in storage in each reservoir. Evaporation is assumed to be a linear function of average reservoir storage in each time period.

Demands and water supply deficits are related to contracts and releases by constraints (3). Deficits occur if releases from Lake Travis do not meet the firm and interruptible demands in each period. Due to the inequalities, UI_{ts} is positive only if there is an interruptible water demand deficit, while UF_{ts} is positive only if there is a firm water demand deficit (due to the higher cost of firm demand deficits, UF_{ts} will never be positive if UI_{ts} equals zero). These constraints also allow for the possibility of spills occurring.

Constraints (4) and (5), along with the maximization of recreational benefits in the objective function, allow recreational benefits to be approximated as piecewise linear functions of storage levels (see Fig. 5). The coefficients in these constraints are indexed by time since recreational benefits accrue at a higher rate in the summer months. The values of the coefficients are based on a study of the economic significance of boating visitation to the Highland Lakes, and therefore do not consider non-boating recreational benefits of the lakes (U.S. Army Corps of Engineers, 1994).

Reservoir storage constraints are defined by relations (6)–(9). Constraints (6) specify that storage levels in each reservoir are not allowed to exceed conservation capacity at any time, and constraints (7) prevent total storage in both reservoirs from dropping below a reserve storage level at any time. For equity purposes, constraints (8) require the two reservoirs to be drawn down by approximately the same amount. Finally, constraints (9) require the expected final storage levels (at $T = 60$) in each reservoir to meet or exceed a specified target level. The purpose of these constraints on expected final storage is to reduce end-of-horizon effects in the model. Without it, the model tends to drain the reservoirs in the last time period, since there is no value placed on storage beyond the planning horizon.

Constraints (10) require contracts from year 2 onward to equal or exceed some fraction of the previous year's contract. This conforms with a secondary goal of the LCRA, which is to maintain contract stability from year to year.

Constraints (11), written using the split-variable approach, are the non-anticipativity constraints which prevent decisions from anticipating future years' inflows. These specify that decisions must be the same for all scenarios (s and s') which have the same data up until time t. For example, with reference to Fig. 4,

$$X_{2,\,s} = X_{2,\,s+1}, \quad s = 1,\ldots,5 \tag{13}$$

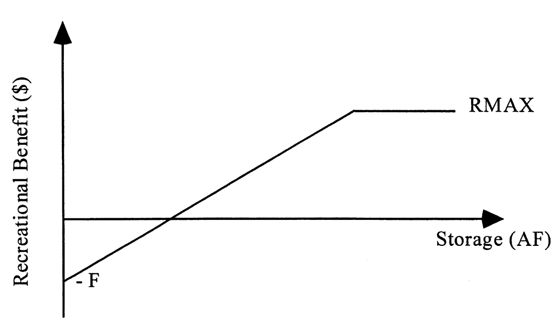

Fig. 5. Recreational benefit function, increasing linearly up to some maximum level.

That is, scenarios 1 through 6 all have a common inflow history during year 1, so the interruptible contract amounts at the start of year 2 must be equal across these scenarios. Similarly, $X_{3,1} = X_{3,2}$ because, at the start of year 3, scenarios 1 and 2 have the same inflow history. For reservoir releases, first-year releases from both reservoirs in scenarios 1 through 6 must be equal:

$$R1_{t,s} = R1_{t,s+1} \quad \text{and} \quad R2_{t,s} = R2_{t,s+1}, \quad t \leq 12,\ s = 1, \ldots, 5 \tag{14}$$

Constraints (11) also hold for $s = 7, \ldots, 11$; $s = 13, \ldots, 17$; etc.

Finally, constraints (12) specify that all decision variables are non-negative.

4. Solution procedure

The multistage stochastic planning model (comprising over 25,000 columns, 14,000 rows, and 70,000 non-zero elements) is solved using both GAMS/OSL (Brooke et al., 1992) and the Stochastic Programming Interface to the Optimization Subroutine Library, SP/OSL (King, 1994). The model was initially coded in GAMS and solved as a single large-scale linear program using GAMS/OSL. GAMS provided a convenient means of specifying the large model in a compact way, and the pre-solve capabilities of GAMS/OSL (removal of redundant rows and fixing of variables summing to zero) were useful in reducing the size of the split-variable model. Furthermore, multiobjective (parametric) analysis could be accelerated by applying the primal simplex method repeatedly within a single GAMS program, thereby providing an advanced basis start for each subsequent solve. Solution times on an IBM/RS 6000 were on the order of 6 to 8 CPU minutes for the first solve (11,000–13,000 simplex iterations) and from 60 to 90 sec for subsequent solves (1500–2000 simplex iterations).

Since many of the tasks related to the management of stochastic data structures, as well as special solution algorithms such as Benders decomposition, are not supported by GAMS, SP/OSL was also used to solve the model. The SP/OSL software library comprises portable C language subroutines for initialization and data input, scenario tree generation, solution by Benders decomposition, and user access to data and model results. The decomposition algorithm uses various facilities of the IBM Optimization Subroutine Library Guide (OSL, 1991) and implements a number of enhancements to Benders' original algorithm (Benders, 1962). These include a multicut version of the algorithm, an augmented master problem which includes the first subproblem in order to generate better proposals, and the addition of a simplex phase in the event that Benders decomposition is slow to converge. Specifying a maximum Benders optimality gap of 1×10^{-6}, this approach resulted in significantly faster 'cold start' solutions than with GAMS/OSL. Cold start solution times ranged from 45 to 75 CPU sec on the same IBM/RS6000 (6 to 10 major Benders iterations, with 6000–12,000 total simplex iterations needed to solve all subproblems and all master problems). It should be noted that the decomposition technique will likely become substantially more advantageous for scenario trees containing a greater number of nodes.

5. Results

The data used to generate model results are shown in Table 4. The firm water demand from storage, FIRM, is set to a value approximately equal to the City of Austin's current use. The reserve storage level, RESV, is set to the value at which interruptible demands are completely curtailed (LCRA, 1993). Results are generated for a range of initial reservoir storage levels and a range of importance weights placed on recreational benefits. As shown in Fig. 6, the 'worst case' solution with respect to interruptible water contracts corresponds to low initial storage levels and high recreational benefits. If average annual recreational benefits are to exceed \$60 million, then the LCRA should contract no interruptible water in the first-stage under any initial storage conditions (since $X_1 = 0$ for all initial storage levels). If recreational benefits of only \$55 million/year. are desired, but the lakes are only 70% full initially, the first-stage decision is also $X_1 = 0$ (labeled as Solution C). However, in this case, this does not mean that the LCRA should never contract for the sale of interruptible water: if year 1 turns out to be 'wet', the model calls for second-stage water contracts greater than zero. In any case, it is important to note that no water contract deficits occur in any of the model runs. The first-stage decision is always sufficiently cautious to prevent this highly undesirable event from occurring.

Table 4
Data for the LCRA Highland Lakes system

Item	Value
Cost coefficients	
c_o	4.50 \$/AF
c_f	600.00 \$/AF
c_i	120.00 \$/AF
Reservoir storage data	
$K1$, $K2$	918, 117 thousand AF
$S1_T$, $S2_T$	500, 700 thousand AF
RESV	200 thousand AF
Water supply data	
FIRM	75 thousand AF
fi_t (January–December)	0.0, 0.0, 0.031, 0.089, 0.184, 0.231, 0.149, 0.165, 0.136, 0.015, 0.0, 0.0
ff_t	0.062, 0.062, 0.065, 0.077, 0.080, 0.095, 0.116, 0.122, 0.098, 0.083, 0.072, 0.068
Evaporation data	
$e1_t$ (January–December)	0.08, 0.05, 0.18, 0.10, 0.14, 0.38, 0.66, 0.71, 0.43, 0.31, 0.22, 0.11
$e2_t$	0.06, 0.02, 0.24, 0.10, 0.17, 0.35, 0.60, 0.63, 0.35, 0.26, 0.18, 0.08
$A1$, $A2$, $B1$, $B2$	0.02194, 0.01354, 2100, 1200
Recreational benefit coefficients	
$F1_t$, $F2_t$ (June–August)	−0.91, −1.22
$F1_t$, $F2_t$ (September–May)	−3.65, −4.89
$G1_t$, $G2_t$ (June–August)	0.00087, 0.00175
$G1_t$, $G2_t$ (September–May)	0.00349, 0.00699
Equity and contract stability coefficients	
α, β	0.5, 0.5

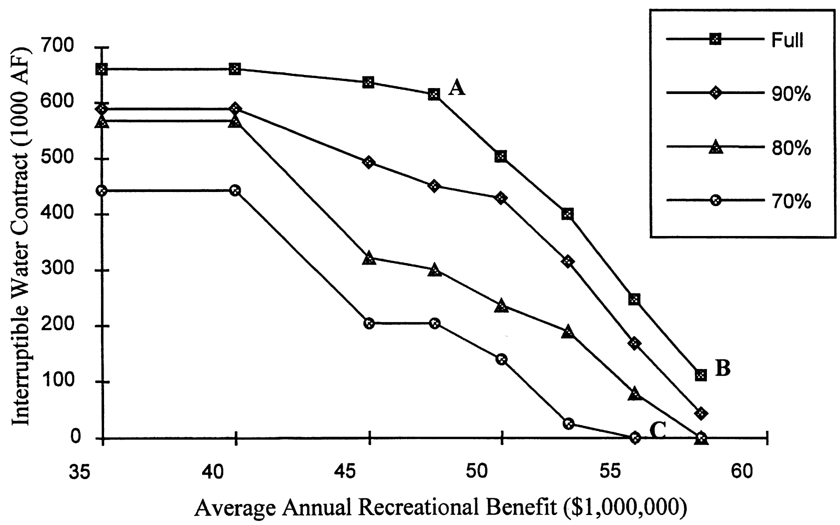

Fig. 6. First-stage contract decision, X_1, as a function of recreational benefits and initial storage levels.

The distribution of second- and subsequent-stage water contracts from Solution C are shown in Fig. 7. It can be seen that the second-stage decisions are still quite cautious, though not as cautious as the first-stage decision. 40% of the scenarios — those which have low inflows in the first year — call for little or no water to be contracted in the second year. Substantial second-year contracts (from 400 to 600 thousand AF) are made only when inflows are moderate to high in the first year. On the contrary, the third- and fourth-stage decisions appear rather reckless, with contracts as large as 3 million AF in some scenarios. Essentially, these decisions are made with perfect foresight since the scenario tree does not branch after the second year. The model can look ahead and see large inflows coming in years 4 and 5, appropriately deciding to sell an immense amount of water. Again, this is not a major concern,

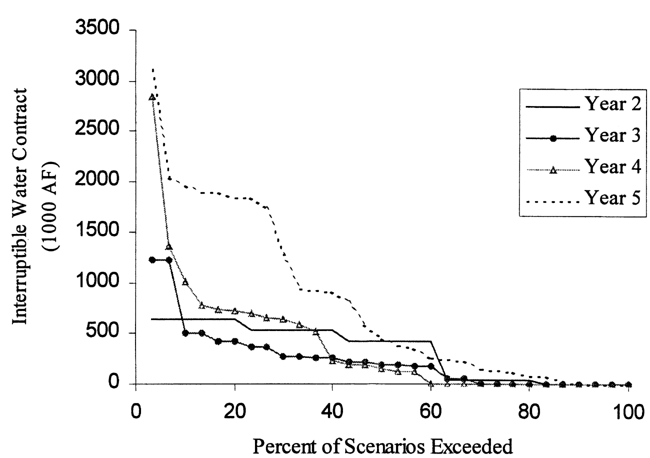

Fig. 7. Distribution of annual interruptible water contracts from Solution C (Recreational benefit = 55 million/year; lakes 70% full initially).

since from the standpoint of decision support it is important only that the first-stage decision hedge appropriately.

To further evaluate the utility of this model for decision support, it is useful to look at reservoir storage levels over time. Consider the solutions labeled A and B in Fig. 6. Solution A calls for a large first-stage water contract (X_1 = 614,000 AF) at the expense of recreational benefits, while Solution B calls for a small first-stage water contract (X_1 = 105,000 AF) so that recreational benefits can be maintained at higher levels. For these two solutions, the 10%, median, and 90% reservoir storage levels (in each time period) are shown in Figs. 8–11

From these figures, the end-of-horizon effects are quite apparent. In Solution A, Lake Travis is severely depleted at the end of the horizon in most scenarios, since there is no value placed on water supply beyond year 5. Meanwhile, levels in Lake Buchanan tend to drop in years 3 and 4, but increase in year 5 in order to meet the constraint on expected final storage ((9)). Lake levels generally remain higher in Solution B due to the high importance placed on recreational benefits. However, as seen in the 10% storage trace for Lake Travis (Fig. 11), there are still some rather reckless decisions in years 3–5. From these results, we can conclude that the decisions from the first and second stages of the model might be useful to the LCRA, while the primary purpose of the last stage of the model (years 3–5) appears to be the reduction of end effects.

6. Sensitivity analysis

As mentioned above, one limitation of this model is that the scenario generation procedure does not preserve the annual correlation of inflow. Although the serial correlation coefficient of total annual available inflow is rather small (approximately 0.17), Vaugh and Maidment (1987) found, for six-year sequences, that random ordering produces less severe sequences than does the historical order. Doubts as to the reliability of model results also arise from the ad hoc construction of the scenario tree and the relatively short record of observed flows used.

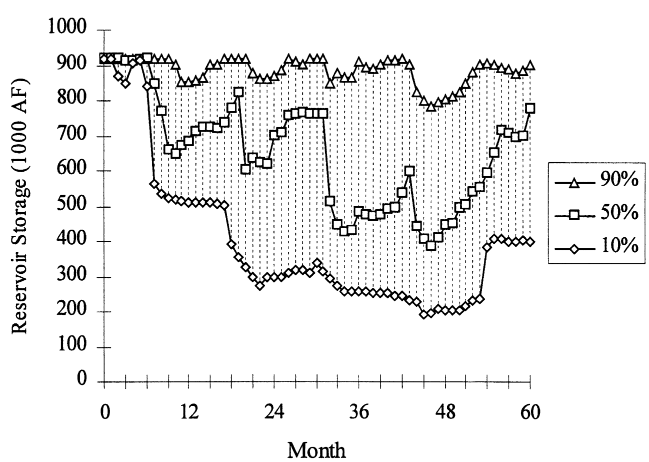

Fig. 8. Lake Buchanan storage levels for Solution A.

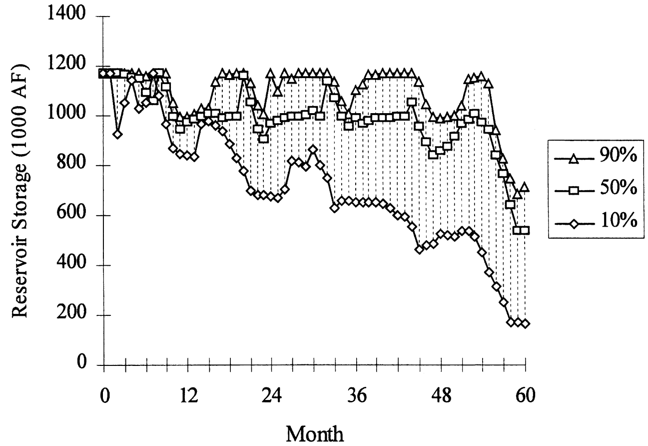

Fig. 9. Lake Travis storage levels for Solution A.

To evaluate the effect of neglecting annual inflow correlation, model results are compared to those using a new scenario generation technique that attempts to preserve the correlation of flows across stages. This new procedure is based on the nearest-neighbor bootstrap (NNB) algorithm developed and applied to hydrologic time series by Lall and Sharma (1996). Rather than sampling flows in a random manner, the NNB first identifies a set of historic 'nearest neighbors' to the current hydrologic state and then samples (with replacement) from their successors in an attempt to preserve serial correlation. This approach is illustrated in Fig. 12. Given a time series from $t = 1$ to 24, and assuming the stochastic process is adequately modeled as an autoregressive lag-one process, the five nearest neighbors to the time series value at time $t = 24$ are marked with circles. A scenario or forecast for time $t = 25$ is generated by sampling from the successors of the nearest neighbors, indicated by arrows.

Application of the NNB requires selection of the following: a 'feature vector' which characterizes the state of the system, $\mathbf{D}_t$; the number of nearest neighbors k whose successors will be considered for sampling; and a discrete 'sampling kernel' to weight the likelihood of

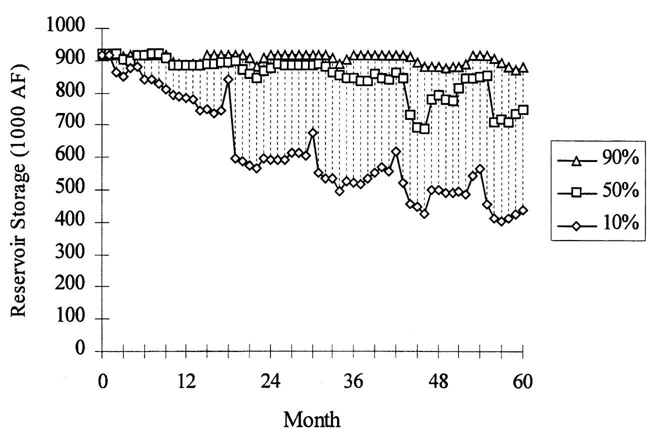

Fig. 10. Lake Buchanan storage levels for Solution B.

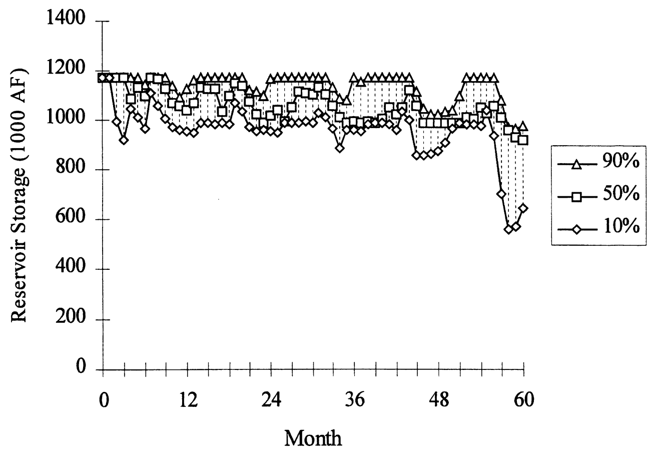

Fig. 11. Lake Travis storage levels for Solution B.

each successor being sampled based on the similarity of the nearest neighbor to the current feature vector, $\mathbf{D}_i$ (Lall and Sharma, 1996). In this study, the feature vector is actually a scalar, equal to the total annual inflow, and a model order of one is chosen (i.e., it is assumed that future annual flows depend only on the previous year's flow). Five nearest neighbors are considered ($k = 5$), and the discrete kernel for sampling is defined as

$$K\big(j(i)\big) = \frac{\frac{1}{j}}{\sum_{j=1}^{k}\frac{1}{j}} \tag{15}$$

where $K(j(i))$ is the probability that the successor of the jth nearest neighbor to $\mathbf{D}_i$ will be sampled. To construct a scenario tree, second stage sequences are sampled randomly, and the

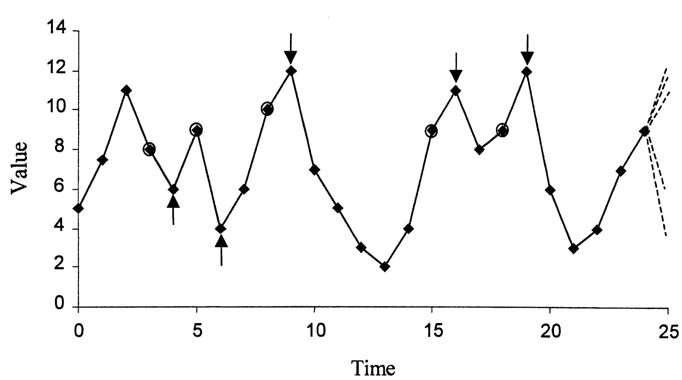

Fig. 12. Example of the nearest-neighbor bootstrap. Nearest neighbors of the value at time $t = 24$ are indicated by circles, and their successors are indicated by arrows.

bootstrap procedure is applied at each branch in the remaining stages using the parent branch to define the feature vector.

To test the performance of the NNB procedure, 500 25-year sequences are generated, along with 500 sequences of the same length using the original 'random' sampling procedure (in both cases, two single-year segments are followed by a three-year segment, as in the scenario tree). One metric for comparison is the serial correlation coefficient of annual available flows. Another, perhaps more useful, metric is the maximum cumulative deficit from the median annual flow. For an n-year sequence, this maximum cumulative deficit, $C^{\max}$, is computed as

$$C_1 = \max\{0, M - Q_1\} \quad (16a)$$

$$C_t = \max\{0, C_{t-1} + M - Q_t\}, \quad t = 2, \ldots, n \quad (16b)$$

$$C^{\max} = \max\{Ct{:}t = 1, \ldots, n\} \quad (16c)$$

where M is the median annual flow, and Q_t is the annual flow in year t. The performance of the two procedures based on these metrics is summarized in Table 5.

In general, the NNB procedure appears to outperform the random sampling procedure. As expected, random sampling does a poor job of reproducing the annual correlation of flows, and it tends to generate sequences with smaller maximum cumulative deficits than the historical sequence. However, the mean maximum cumulative deficit is only about 15% smaller than the historical maximum cumulative deficit. This may be due to the fact that there were two short periods (1949–1951 and 1962–1963) in which the cumulative deficit exceeded 100 thousand AF, and these periods might be sampled in the three-year segments. NNB does a better job of preserving the annual correlation of flows, and it actually tends to produce sequences with larger maximum cumulative deficits than the historical sequence. One reason for this might be that the drought of record ended with a year of extremely high flows.

More important than the performance of the two sampling procedures based on these metrics is the sensitivity of model results to the scenarios generated. To evaluate this sensitivity, model results with 20 scenario trees generated by NNB are compared to results with 20 trees generated by the random sampling procedure. These results are also compared to the results from the original tree, as summarized in Table 6.

These results indicate that solutions are quite sensitive to the flow sequences considered by

Table 5
Performance of the random and NNB sampling procedures

	Mean annual correlation	Mean maximum cumulative deficit (1000 AF)	Standard deviation maximum cumulative deficit (1000 AF)
Historical sequence	0.173	2598	–
Random sampling	0.024	2229	1024
NNB	0.121	2827	1228

Table 6
Sensitivity of model solutions (ω and initial storage corresponding to Solution A)

	Mean first-stage solution	Standard deviation solution	Mean objective value	Standard deviation of objective value
Original tree (Sol. A)	613.8	–	17,559	–
Other random trees (20)	581.1	91.1	19,128	1084
NNB-generated trees (20)	555.3	180.7	18,063	1275

the model. Both the values of the first-stage decision variables and the optimal objective function values can vary substantially depending on the inflow scenarios sampled. Solution A, in fact, does not appear to be very representative of the other solutions from scenario trees produced by random sampling. Although the first-stage decision value of Solution A is well within one standard deviation of the mean first-stage decision value produced by either sampling method, the Solution A objective value is about 1.5 standard deviations below the mean objective value produced by random sampling. As expected, model solutions based on randomly sampled scenarios hedge less and produce more optimistic objective function values than those based on NNB-generated scenarios. Solutions based on NNB-generated scenarios also demonstrate more variability, perhaps because the NNB scenarios themselves appear to have more variable drought characteristics, as shown in Table 5.

The sensitivity of model results to changes in the structure of the scenario tree can also be analyzed. Since SP/OSL provides facilities for selecting subsets of the nodes in the tree, automatically adjusting the probabilities assigned to each branch, two additional scenario trees are easily constructed from the 'full' four-stage tree: (1) a three-stage tree with 15 scenarios, in which flows on the fourth-stage branches of the full tree are averaged; and (2) a two-stage tree with five scenarios, in which flows on both the third- and fourth-stage branches of the full tree are averaged. Models based on these trimmed scenario trees can be expected to provide more optimistic solutions than the original four-stage model since their 'futures' are characterized by fewer, and less diverse, scenarios. This expectation is indeed borne out in the results. As shown in Table 7 for two values of ω, the fewer the stages and scenarios, the less the solution hedges (i.e., the larger the first-stage contracts), and the higher the optimal objective value.

Viewing Table 7 from the bottom up, the first-stage solution appears to be converging as the branches and stages in the tree are increased in number. This indicates that the structure of the original four-stage scenario tree may be adequate for representing of the variability of inflows observed in the 25-year record. Nonetheless, ensuring that uncertainty is modeled satisfactorily for future decision-making would require a more in-depth investigation of a larger hydrologic database for the region. Furthermore, before any actual decision is made, candidate policies should be tested in a more detailed simulation model considering a larger number of inflow scenarios.

Table 7
Summary of scenario tree sensitivity analysis[a]

Model	$\omega = 0.2$	$\omega = 0.4$
4-Stage (30 scenarios)	240, 55.9, 12.1	189, 58.3, 8.9
3-Stage (15 scenarios)	271, 56.0, 14.0	191, 58.3, 11.1
2-Stage (5 scenarios)	377, 56.6, 15.3	301, 58.6, 12.8

[a] Items are the first-stage contract amount (1000 AF), total expected recreational benefits ($ million/year), and total expected revenue from sale of interruptible water contracts ($ million).

7. Use of model results

Currently, water management at the LCRA is based largely on deterministic analysis of hydrologic conditions. In part, this is a consequence of not understanding the probabilistic nature of the variables, as well as difficulty in deciding on acceptable levels of risk even when probabilistic information is available. More importantly, the State of Texas, which regulates surface water, requires the LCRA and other regional agencies to use the 'firm yield' concept for estimating dependable water supplies. This approach considers the most severe historical drought conditions as a worst-case scenario for estimating dependable water supply. However, analysis of rainfall and other long-term data indicate that the historical critical drought for the Highland Lakes has a return interval greater than once a century. Therefore, basing all water supply operations on such an infrequent event could lead to operational practices which are overly conservative.

As discussed in this paper, the LCRA has departed to some extent from using the firm yield concept by allowing the use of interruptible stored water. Interruptible water is made available only on an annual basis, and its sale must not impact the water supply available to meet the firm water demands expected over the next ten years. Thus, a probabilistic estimate of any potential impact is beneficial in deciding how much interruptible water to commit to supply for the coming year. Use of a stochastic planning model can not only help to provide an estimate of the risks involved, but also help determine contract amounts which hedge appropriately against future inflow scenarios.

Using a stochastic model may also lead to more beneficial use of available water resources for some secondary purposes. Although water supply and flood control are the primary purposes for which the Highland Lakes were built, hydroelectric power generation and recreation are important secondary benefits. The lakes can be managed to optimize the use of water for these secondary purposes if it can be demonstrated that there is a low probability of a significant impact on water supply or flood control. Thus, the stochastic planning model developed in this paper may ultimately benefit LCRA water management by helping to add a degree of flexibility to their operating procedures.

8. Conclusion

A multi-stage stochastic programming model is developed to support the LCRA's decision of how much interruptible water to contract each year. This type of model can provide different insights than deterministic optimization models which, in essence, assume that future inflows are known with certainty. By explicitly considering a number of inflow scenarios, the stochastic model can determine a contract level that balances interruptible water supply and recreational goals while appropriately hedging against the effects of drought. However, care must be taken not to overstate the benefits of the stochastic model, since it considers only a moderate number of scenarios, and these are constructed from a rather small amount of statistical data. Compilation of a larger hydrologic database, model formulation with a larger scenario tree, sampling that preserves the serial correlation of inflows, and a decomposition technique for model solution are recommended for improved results.

Acknowledgements

The authors wish to thank Alan King of IBM Corp. for making available the SP/OSL software. Helpful discussions the authors had with David Maidment and David Morton of the University of Texas are also appreciated, as are the comments and suggestions of reviewers.

References

Benders, J.F., 1962. Partitioning procedures for solving mixed variables programming problems. Numerische Mathematik 4, 238–252.

Birge, J.R., 1985. Decomposition and partitioning methods for multistage stochastic programs. Oper. Res 33, 989–1007.

Birge, J.R., Louveaux, F.V., 1997. Introduction to Stochastic Programming. Springer-Verlag, New York.

Brooke, A., Kendrick, D., Meeraus, A., 1992. GAMS: A User's Guide, Release 2.25. The Scientific Press, Redwood, CA.

Carpenter, T., Lustig, I., Mulvey, J., 1991. Formulating stochastic programs for interior point methods. Opers. Res 39, 757–770.

Dantzig, G.B., Glynn, P.W., 1990. Parallel processors for planning under uncertainty. Annals of Opers. Res 22, 1–22.

Ermoliev, Y., 1988. Stochastic quasigradient methods and their applications to systems optimization. Stochastics 9, 1–36.

Gaivoronski, A. 1988. Stochastic quasigradient methods and their implementation. In: Ermoliev, Y., Wets, R.J.-B. (Eds.), Numerical Techniques for Stochastic Optimization. Springer-Verlag, Berlin.

Gassman, H.I., 1990. MSLiP: A computer code for the multistage stochastic linear programming problem. Math. Prog 47, 407–423.

Higle, J.L., Sen, S., 1991. Stochastic decomposition: an algorithm for two-stage linear programs with recourse. Mathematics of Opers. Res 16, 650–669.

Higle, J.L., Sen, S., 1996. Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming. Kluwer Academic Publishers, Dordrecht.

Hooper, E.R., Georgakakos, A.P., Lettenmaier, D.P., 1991. Optimal stochastic operation of Salt River project, Arizona. J. Water Resour. Plng and Mgmt., ASCE 117 (5), 566–587.

Infanger, G., 1992. Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. Annals of Opers. Res 39, 69–95.

Infanger, G., 1994. Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs. The Scientific Press, Boyd and Fraser, Danvers, MA.

Jacobs, J., Freeman, G., Grygier, J., Morton, D., Schultz, G., Staschus, K., Stedinger, J., 1995. SOCRATES: A system for scheduling hydroelectric generation under uncertainty. Ann. of Opns. Res 59, 99–134.

Kelman, J., Stedinger, J.R., Cooper, L.A., Hsu, E., Yuan, S.-Q., 1990. Sampling stochastic dynamic programming applied to reservoir operation. Water Resour. Res 26 (3), 447–454.

King, A.J., 1994. SP/OSL Version 1.0: Stochastic Programming Interface Library User's Guide. International Business Machines Corp, Yorktown Heights, NY.

Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour. Res 32 (3), 679–693.

LCRA, Lower Colorado River Authority, 1993. Water Management Plan for the Lower Colorado River Basin. Austin, TX.

Lustig, I.J., Marsten, R.E., Shanno, D.F., 1994. Interior point methods for linear programming: computational state of the art. ORSA Journal on Computing 6, 1–14.

Martin, Q.W., 1995. Optimal reservoir control for hydropower on Colorado River, TX. J. Water Resour. Plan. Mgmt., ASCE 121 (6), 438–446.

Mulvey, J.M., Ruszczynski, A., 1995. A new scenario decomposition method for large-scale stochastic optimization. Oper. Res 43 (3), 477–490.

Optimization Subroutine Library Guide and Reference, 1991. International Business Machines Corp., SC23-0519-02.

Pereira, M.V.F., Pinto, L.M.V.G., 1985. Stochastic optimization of a multireservoir hydroelectric system: a decomposition approach. Water Resour. Res 21 (6), 779–792.

Pereira, M.V.F., Pinto, L.M.V.G., 1991. Multi-stage stochastic optimization applied to energy planning. Math. Prog 52, 359–375.

Rockafellar, R.T., Wets, R.J.-B., 1991. Scenarios and policy aggregation in optimization under uncertainty. Math. Oper. Res 16, 119–147.

Ruszczynski, A., 1986. A regularized decomposition method for minimizing a sum of polyhedral functions. Math. Prog 35, 309–333.

Ruszczynski, A., 1997. Decomposition methods in stochastic programming. Math. Prog 79, 333–353.

Two Water Authorities Pledge to Cooperate, 1995. The Austin American-Statesman. November 21, p. A1.

U.S. Army Corps of Engineers, 1994. The Economic Significance of Boating Visitation to the Highland Lakes. U.S. Army Eng. Dist., Fort Worth, TX.

Van Slyke, R., Wets, R.J.-B., 1969. L-shaped linear programs with applications to optimal control and stochastic programming. SIAM J. Appl. Math 17, 638–663.

Vaugh, S.K., Maidment, D.R., 1987. Projecting storage in the Highland Lakes reservoir system. J. Water Resour. Plann. Mgmt., ASCE 113 (5), 659–676.