Stochastic linear programming and Conditional Value-at-Risk for water resources management

R. B. Webby[1] J. Boland[2] P. G. Howlett[3]
A. V. Metcalfe[4]

(Received 1 September 2006; revised 16 January 2008)

Abstract

A mathematical analysis is presented for decision support for managing water resources in a water-limited environment. The water sources include rainfall, either direct or that held in reservoirs, shallow aquifers, river water withdrawal entitlements, and recycled water. Water from each source has its own characteristics of quality and thus suitability for use, quantity, temporal availability, environmental impact of use and cost to access. Water availability is modelled by a multivariate probability distribution. Relative values for salinity levels and nutrient or mineral loads are given and other water characteristics are summarised by a price for water from each source. We formulate

See http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/136 for this article, Published January 30, 2008. ISSN 1446-8735

and solve a stochastic linear program to find the optimal blend of the available sources while meeting quality and supply constraints. We apply these techniques to a common water resource management problem facing an Australian farmer, that of growing a summer crop usually reliant on irrigation. We compare alternate cropping decisions based on their risk of failing to meet supply or quality standards. Our measure of risk is Conditional Value-at-Risk.

Contents

1 Introduction

To illustrate the use of Conditional Value-at-Risk (CVaR) as a decision support tool for water resource managers, we present an application focussing on the irrigation requirements of a summer crop in a water limited environment. In this situation, water may be available from a number of sources such as rainfall, shallow aquifer groundwater, an entitlement to withdraw river water, and tailwater, that is, water collected from previous crop irrigation operations and recycled. This is a study to explore what questions can be asked using this approach and we present a simple model. The results are more to support intuition than to make reliable decisions.

Yamout and El-Fadel [3] formulated a linear program for a domestic water supply problem for Greater Beirut. Water supplies were deterministic and they included socio-environmental practices as constraints. Linear and non-linear programming algorithms have been used in coal blending for power generation, treating sources of coal as having known quality and quantity characteristics [1, e.g.]. Here we allow water from some sources to be stochastic in availability. We solve a linear program to minimise the cost of providing water which must meet quantity and quality constraints. We evaluate alternate decisions in terms of the linear program solutions and the CVaR values calculated from a distribution for minimum cost built up from sampling instances of the stochastic variable. CVaR has been applied in crop selection [4], where a maximum value of CVaR was included as a constraint in a linear program.

In deciding to grow a summer crop a farmer determines whether sufficient water is available to bring the crop to harvest, and compares the cost of that water and other input costs against the expected return. However, water is a crucial input to producing a crop and in this stochastic linear program formulation of the decision problem we focus on the frequency of seriously adverse events. The information from our solutions could be used to guide future practical farm works, and also the level of hedging (crop insurance or futures products) that might be applied to cover the investment in the crop.

2 Model definition

2.1 Definition of VaR and CVaR

Value-at-Risk (VaR) is a measure of risk developed in the finance industry for evaluating the risk exposure of a portfolio of financial instruments such as shares, bonds and derivatives. VaR is defined as the maximum loss expected to be incurred over a given time horizon at a specified probability level. Mathematically, let $x \in X \subset \mathbb{R}^n$ be a decision vector and $y \in Y \subset \mathbb{R}^m$ be a vector representing the values of a contingent variable influencing the loss. Let $z = f(x, y)$ be a function that describes the loss generated by x and y. At probability level $\alpha \in (0, 1)$, the VaR_α of the loss associated with a decision x is defined as [2]

$$\text{VaR}_\alpha(x) = \inf\{z \mid G(x, z) \geq \alpha\}\,, \tag{1}$$

where $G(x, z)$ is the cumulative density function for loss associated with decision x.

VaR gives the value of the specified quantile of the distribution but does not give any information about the upper tail beyond that value. That is, VaR describes the frequency of a sizable loss but not the likely severity of such a loss. CVaR does contain information about losses in the upper tail. CVaR is the expected loss, given that a loss greater than or equal to the threshold VaR occurs. The CVaR_α of the loss associated with a decision x is defined as [2]

$$\text{CVaR}_\alpha(x) = \mathrm{E}\{z \mid G(x, z) \geq \alpha\}\,, \tag{2}$$

where E denotes the expectation operator.

In this article we generate a cost, rather than loss, distribution through simulation of a mathematical model of the system. VaR is then found as the

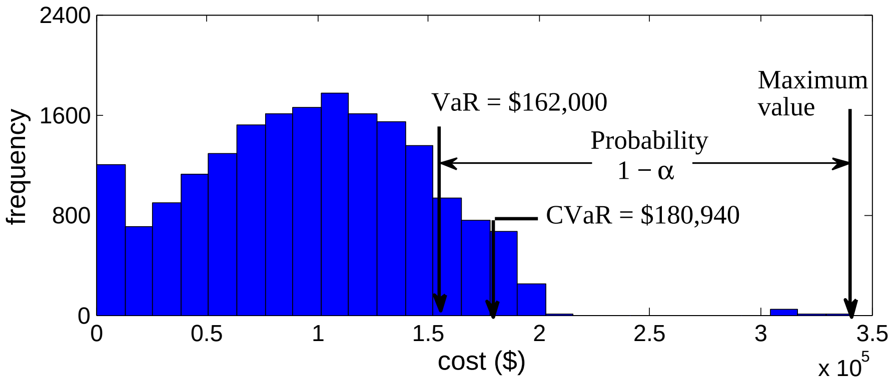

Figure 1: An example of the cost distributions simulated in Section 3 with VaR and CVaR indicated.

αth proportional value of the ordered distribution, and CVaR as the mean of the values equal to or beyond VaR. Figure 1 shows VaR and CVaR values for an empirical cost distribution generated by our model for Section 3. The mean cost is $96,095 and although most of the simulated costs are less than $200,000, there is a positive probability of experiencing costs of $3\frac{1}{2}$ times the average. For this distribution VaR is $162,000 and CVaR approximately $181,000. CVaR will always be greater than or equal to VaR.

2.2 Stochastic linear programming

Linear programming involves problems of the form

$$\begin{aligned} \min \mathbf{c}^{\mathbf{T}}\mathbf{x}\,, \\ \text{subject to}\quad A\mathbf{x} \leq \mathbf{b}\,, \\ \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\,, \end{aligned}$$

where $\mathbf{c}^{\mathbf{T}}\mathbf{x}$ is a cost function, $\mathbf{l}$ is a lower bound and $\mathbf{u}$ an upper bound for $\mathbf{x}$. The cost function is minimised subject to constraints which may be equality or inequality constraints. Stochastic linear programming allows for some elements of the constraint equations to be stochastic. In this application some elements of $\mathbf{b}$ are stochastic.

One approach to solving stochastic linear programs is to take particular values for the stochastic variables and solve the resulting deterministic problem. Values typically chosen are the expected value of the variable, its expected value plus and minus one or two standard deviations, or simply a spread of possible values of the variable. Another approach is to sample values from the distributions of the random variables and again solve a deterministic program. This method is particularly suited where there are correlations between the stochastic variables. Our approach, this latter one, involved specifying a multivariate normal distribution for the availability of rainfall and groundwater, allowing us to incorporate correlation between the random variables. Methods for generating samples for the multivariate normal are readily available but other distributions could be used. A copula or the empirical Gibbs sampler could also be used to generate multivariate data from arbitrary distributions. After sampling values from the input distributions, we use linear programming to find the optimal blend of water from the four sources to obtain the lowest cost for producing the crop. The program is run multiple times to build up an empirical distribution for the minimum cost and calculate CVaR values for the distribution.

We set x_j, $j = 1, \ldots, J$, to represent the amount of water taken from each source j. The cost of the water is c_j, and the amount of water available from each source in a given summer is a_j. Each source has a particular salinity concentration, s_j, and mineral or nutrient load, m_j, and we set maximum levels for these in the blended water of S and M respectively. We consider an individual crop with a water requirement for full potential productivity across a crop area of H hectare of X Ml. Expressed as a linear program, the water blending problem is

$$\begin{aligned} \min \quad & \sum_j c_j x_j \,, \\ \text{such that} \quad & x_j \leq a_j \,, \end{aligned}$$

$$
\begin{aligned}
&\sum_j s_j x_j \Big/ \sum_j x_j \le S\,, \\
&\sum_j m_j x_j \Big/ \sum_j x_j \le M\,, \\
&\sum_j x_j \ge X\,, \\
&x_j \ge 0 \quad \text{for } j = 1, \ldots, n\,.
\end{aligned}
$$

2.3 Water characteristics

We characterize the various water supplies as shown in Table 1. The salinity values are typical values encountered in inland cropping areas of Australia and here are fixed as a summer average, although they could also be made stochastic. For example, bore and river water may increase non-linearly in salinity throughout a summer. The mineral or nutrient loads are typical relative values for each source, and could represent sodicity levels in soil water or nitrate levels in recycled water. We use a bivariate normal distribution to represent the amounts of rainfall and groundwater available and model them as being correlated with a coefficient of 0.7. Cost per Ml of water is intended to represent the relative cost of accessing water from the respective sources. It then would include pumping, storage and application costs, and assumes the same application method is used for each crop, as well as costs to represent the environmental cost of using water from a given source. We are not certain of the accuracy of some of our parameters so have not carried out sensitivity tests on them.

Table 1: Relative values for water characteristics.

Source	Salinity	Mineral load	Availability	Cost
rainfall	0.035	0.01	stochastic	1
bore	3.2	1.0	stochastic	500
river	0.6	0.1	deterministic	500
recycled	1.4	2.0	deterministic	50

3 Simulation results

Throughout this application we set α to be 0.90 and the time horizon to be the life of the crop. The decision variable is a vector of the alternate actions that could be taken: for example, grow a relatively thirsty crop with higher returns, like cotton; or grow a relatively hardy crop with lower returns, like wheat; or not grow any crop. For each action there is a different cost distribution, and a CVaR value calculated for each one. To minimise exposure to risk, managers should choose the action that has the lowest CVaR value.

3.1 Feasibility of supply

To the question of whether or not to grow a crop, the results (Figure 2) show there is a 99% chance of successfully supplying at least 300 Ml of water under the model conditions. Alternately, the result says that supply does not meet a demand of 300 Ml on 1% of occasions. This increases to a 9% failure rate for a crop requiring 500 Ml of water to reach harvest at full potential.

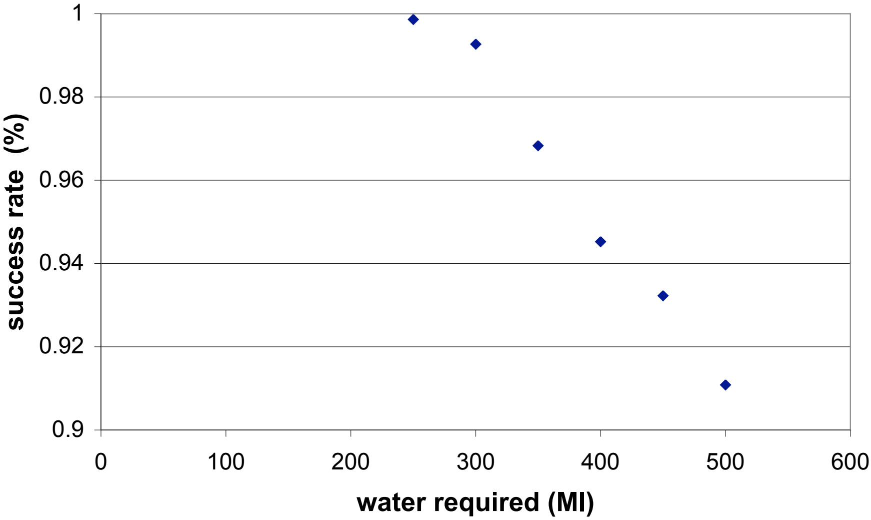

Figure 2: Percentage of simulations meeting various crop water requirements.

3.2 Water requirement of crop

Given that it is decided to grow a crop, should it be a relatively high water demanding crop? or a relatively low water demanding one? Expressed another way the problem is: given that we are able to grow a range of crops with specific water requirements for full growth potential, what area of each crop should be grown? As Figure 3 shows, the cost distribution of producing the thirsty crop has high variability and a bias toward higher values, while the bulk of the simulated costs for a hardy crop are low and the distribution is exponential in nature. The $\text{CVaR}_{0.90}$ value for the more thirsty crop is higher (\$239,459 as against \$79,377) as intuition would suggest. In effect, the CVaR values for both crops and particularly the thirsty crop are higher than stated as we have excluded the infeasible solutions from their calculation. Costs cannot be found for the infeasible solutions; however, they would be at least as great as the highest costs for feasible solutions. They could be

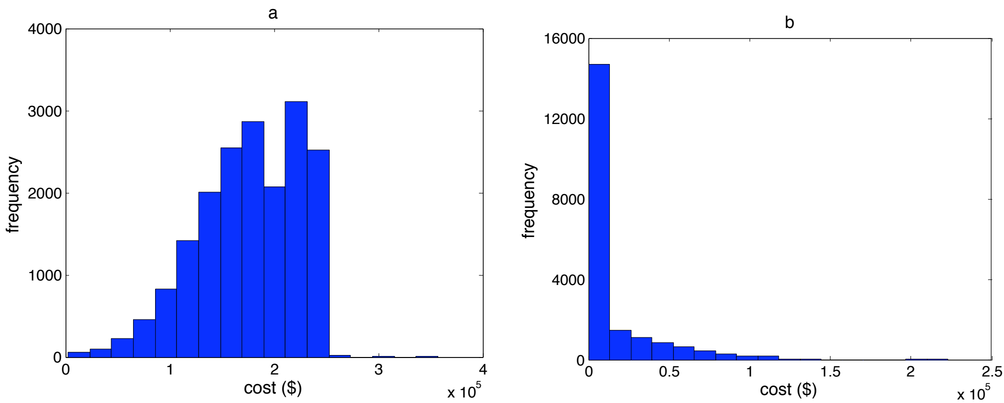

Figure 3: Cost distributions for (a) a relatively high water demanding crop and (b) a relatively low water demanding crop. Neither distribution includes costs for infeasible solutions which occurred at a rate of 9% (a) and 0.2% (b).

much higher in reality if, for example, extra water was purchased to supplement existing supplies. This is one of the advantages of using CVaR as a risk measure over VaR. CVaR does take into account the extreme values in the tail of the cost distribution.

3.3 CVaR and expected return

We illustrate the trade-off between CVaR and expected return by considering gross income from growing a single crop on the H hectare of, say, \$2.0 million for cotton and \$1.2 million for wheat. Each estimated income is multiplied by the probability of achieving full potential yield at harvest, from Section 3.1 above. We estimate total costs at \$476,935 and \$87,040 for cotton and wheat respectively. Expected return, found from expected income minus costs, is \$1,343,065 for cotton and \$1,110,560 for wheat. The net returns should be adjusted by the relative risks involved in irrigating the crop, that is, we subtract the CVaR values found in Section 3.2 and obtain values of \$1,103,606

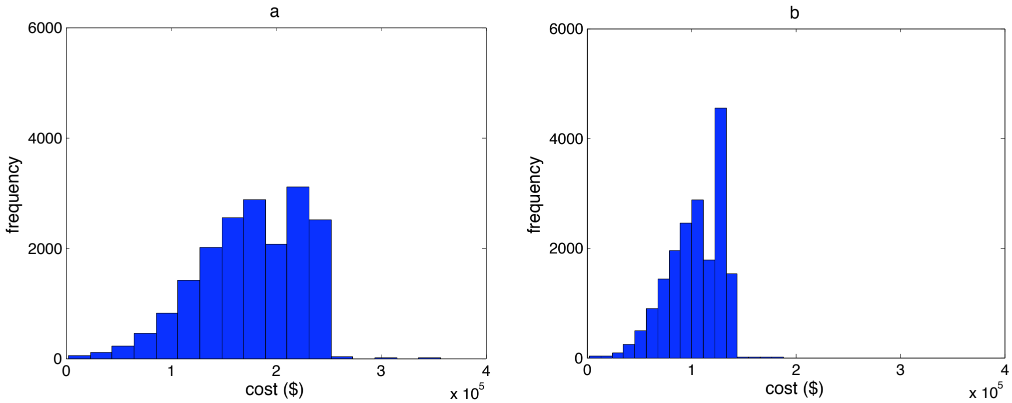

Figure 4: Cost distributions for (a) river water valued at a nominal rate and (b) river water valued at two times the nominal rate.

and $1,031,183, for a financial advantage of cotton over wheat of $72,423.

3.4 Value of entitlement

River water entitlements may become more valuable if water can be sold to other users. For this analysis, we double the cost of river water to represent the opportunity cost of not selling the water. Growing a crop that requires 500 Ml of water (Figure 4), the two cost distributions have a similar shape but are shifted along the horizontal axis. There is about a 75% increase in the cvar value for the higher valued water.

3.5 Model extension

The model described here can be easily extended to consider growing of a range of crops in the one season. The farmer would grow k crops, $k = 1, \ldots, K$, with area h_k under each crop. The decision variable is the relative proportion of the total cropping area to allocate between crops that require differing amounts of water. Then our linear program has added constraints $h_k \geq 0$ and $\sum_k h_k \leq H$ for $k = 1, \ldots, K$. The constraint that supply from all sources, $\sum_j x_j$, at least equals demand, X, is required for a single crop and for a mixture of crops. It is possible to implement constraints representing individual salinity (or mineral load) tolerances for different crops as $\sum_j \sum_k s_j x_{jk} / \sum_j \sum_k x_{jk} \leq S_k$ for $k = 1, \ldots, K$. This multiple-crop problem is not solved here but Liu et al. [4] give a related example.

4 Conclusion

Management of water, on farm and off, is becoming more critical due to the increasing demand, increasing value and, in some areas, decreasing availability of the resource. We present a mathematical analysis for a typical farm water blending problem where water from a variety of sources must meet quantity and quality specifications for crop production. A stochastic linear optimisation model represents the variability in water availability and crop requirements. Monte Carlo simulation is used to test a range of actions relevant to a farming operation and identify the preferred options. We make use of a conservative risk measure, CVaR, which reveals the exposure to risk of possible rare but devastating events. Our model quantifies the rate at which supply fails to meet demand; we generate cost distributions and calculate their CVaR values. While the application of our model in this article is general, using values encountered in the Narrabri region, its parameters could be specified to match conditions applying to any particular farm property.

Acknowledgements We thank the Australian Research Council for supporting this research under grant number DP0559399.

References

[1] F. J. Vasko, D. D. Newhart, and A. D. Strauss. Coal blending models for optimum cokemaking and blast furnace operation. *Journal of the Operational Research Society*, 56: 235–243, 2005. doi:10.1057/palgrave.jors.2601846 C887

[2] R. T. Rockafellar, and S. Uryasev. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26: 1443–1471, 2002. doi:10.1016/S0378-4266(02)00271-6 C888

[3] G. Yamout, and M. El-Fadel. An optimisation approach for multi-sector water supply management in the Greater Beirut area. *Water Resources Management*, 19: 791–812, 2005. doi:10.1007/s11269-005-3280-6 C887

[4] J. Liu, C. Men, V. E. Cabrera, S. Uryasev, and C. W. Fraisse. CVaR model for optimizing crop insurance under climate variability. *Research Report 2006-1, Risk Management and Financial Engineering Lab, University of Florida*, 2006. `http://www.ise.ufl.edu/uryasev/pubs.html` C887, C896

Author addresses

1. **R. B. Webby**, School of Mathematical Sciences, University of Adelaide, Adelaide, Australia.
   mailto:roger.webby@adelaide.edu.au
2. **J. Boland**, Centre for Industrial and Applied Mathematics, University of South Australia, Adelaide, Australia.
3. **P. G. Howlett**, Centre for Industrial and Applied Mathematics, University of South Australia, Adelaide, Australia.
4. **A. V. Metcalfe**, School of Mathematical Sciences, University of Adelaide, Adelaide, Australia.