Southwest Fisheries Center Administrative Report H-79-15

A SYSTEMATIC APPROACH TO DETERMINING MEAN-VARIANCE TRADE OFFS WHEN MANAGING RANDOMLY VARYING POPULATIONS

Roy Mendelssohn
Southwest Fisheries Center
National Marine Fisheries Service, NOAA
Honolulu, Hawaii 96812

August 1979

1. INTRODUCTION

In several recent papers, Beddington and May [1] and May et al. [5] have raised the issue of high variances in the equilibrium harvests when stochastic populations are harvested at high levels. Their concerns suggest that there may be desirable trade offs between the average per period harvest and the long-run (ergodic) variance of the harvests. In order to determine this trade off in a practical manner, what is needed is a systematic method to determine the mean-variance trade off among some subset of policies with desirable properties.

In this paper, such a methodology is suggested. New results by Henig [3] and White and Kim [11] on multiobjective Markov decision problems are combined with an approach to "smooth out" year-to-year fluctuations in the harvest size, an approach studied analytically in Mendelssohn [6]. This leads to a systematic trade off of the mean and variance for only a subset of policies that are Pareto optimal policies for two well defined objectives. The algorithm is applied to a model suggested by Mathews [4] for salmon runs off Bristol Bay, Alaska. Section 2 describes the model and the objective functions to be considered. Section 3 discusses the solution procedure and computational considerations. Section 4 shows the results of the analysis when applied to the example.

2. THE MODEL

A population is to be managed over an infinite planning horizon. At the start of each period t, an initial population size x_t is observed. During period t an amount z_t is harvested, leaving a population size of $y_t = x_t - z_t$ at the end of the period. The population size x_{t+1} at the start of the next period is a random function of y_t and of a random variable D_t, that is:

$$x_{t+1} = s[y_t, D_t] \tag{2.1}$$

where the random variables $D_1, D_2, D_3, \ldots$ are assumed to be independent and identically distributed as the generic random variable D. If a population size x is observed and a decision y is made, an expected one-period return of g(x, y) is received. The returns are discounted by a factor α, $0 \leq \alpha < 1$, and letting E be the expectation operator, it is desired to:

$$\begin{aligned} &\text{maximize } E \sum_{t=1}^{\infty} \alpha^{t-1} g(x_t, y_t) \\ &\text{s.t. } (2.1);\ y_t \in Y(x_t) \end{aligned} \tag{2.2}$$

where Y(x) constrains the population size y that can be left given x. For $\alpha = 1$, the summation in (2.2) will not converge. The results presented can be used in this instance by allowing α to approach 1 from below. A simple example of (2.2) is to maximize total expected discounted yield, that is

$$\begin{aligned} &\text{maximize } E \sum_{t=1}^{\infty} \alpha^{t-1} (x_t - y_t) \\ &\text{s.t. } (2.1);\ 0 \leq y_t \leq x_t \end{aligned} \tag{2.3}$$

A policy that maximizes (2.2) or (2.3) may produce harvest sizes that fluctuate sharply from period to period. This may not be undesirable. However, it is natural to desire to obtain a feel of what is lost if these fluctuations are smoothed out. One method [6] to "smooth out" the period-to-period fluctuations in harvest size is to assess a cost for any change between z_{t-1} and z_t. Specifically, suppose for any decrease in the harvest size a cost of $\lambda \cdot (z_{t-1} - z_t)$ is assessed, and for any increase in the harvest size a cost of $\varepsilon \cdot (z_t - z_{t-1})$ is assessed. Letting $e = \frac{\lambda - \varepsilon}{2}$ and $c = \frac{\lambda + \varepsilon}{2}$, this cost can be compactly represented as $e \cdot (z_{t-1} - z_t) + c \cdot |z_t - z_{t-1}|$. In [6, 7] the approach is to combine the two objective functions. That is, the new optimization problem becomes:

$$\text{maximize } E \sum_{t=1}^{\infty} \alpha^{t-1}\Big(g(x_t, y_t) - e\big(z_{t-1} - (x_t - y_t)\big) - c \cdot \big|(x_t - y_t) - z_{t-1}\big|\Big) \tag{2.4}$$

$$\text{s.t. } (2.1);\ y_t \in Y(x_t);\ z_t = x_t - y_t$$

To smooth the total fluctuations, it is natural to assume $\lambda = \varepsilon$. This implies $e = 0$ and $c = \lambda = \varepsilon$, so that the one-period return for any state (x, z) is:

$$g(x, y) - \lambda \cdot |x - y - z| \tag{2.5}$$

The desired trade off could be approximated by varying λ, and solving (2.4) for each value of λ, and then calculating the ergodic distribution of each policy found. However, there is a more systematic method to achieve the same ends. Assume each period there are two separate returns, $g(x, y)$ and $-\lambda \cdot |x - y - z|$. Then it is desired to find the set of Pareto optimal values and associated policies.

For any policy $\delta(x)$ defined for all x in some set of possible states X, let $v_\delta = \{v_\delta^i\}$ be the vector value of expected returns when following policy δ, that is it is a two-dimensional vector where v_δ^i is the expected return for objective i, i = 1, 2, when following policy δ. Let V be the set of all possible return vectors. Then ν_{δ^*} is Pareto optimal and δ^* is a Pareto optimal policy if there exists no other $\nu_\delta \in V$ such that:

$$v_\delta^i \geq v_{\delta^*}^i \quad \text{for } i, j = 1, 2 \tag{2.6}$$

$$v_\delta^j > v_{\delta^*}^j \quad \text{for some } j$$

Intuitively, (2.6) says that a policy δ^* is a Pareto optimal policy if any other policy δ that increases the expected return for one objective decreases the expected return for the other objective. Thus when determining the mean-variance trade off, policies are included if they lower the costs due to fluctuations as they lower the "true" value of the harvest, or vice versa, increase the "true" value of the harvest if they increase the assessed cost due to fluctuations in harvest size.

Given the set of Pareto optimal policies, the ergodic distribution for each policy is calculated. The mean and variance are then readily calculated from the ergodic distributions.

3. COMPUTATIONAL TECHNIQUES AND CONSIDERATIONS

In this section an approach to finding the set of Pareto optimal policies is described. Henig [3] has provided the theoretical foundation and White and Kim [11] an algorithm for efficiently calculating the Pareto optimal set. Suppose the problem has been redefined on a discrete grid. Let $p(i, j, a_i)$ be the probability of going to state j from state i when the action chosen is a_i. Here a state i refers to a particular vector pair (x, z), that is the present population size and the harvest size last period. Let $P[\delta]$ be the transition matrix for policy δ and let $P_{(x, z)}[\delta]$ be the row of $P[\delta]$ for state (x, z). Let $r[(x, z), \delta]$ be the two-dimensional row vector of rewards when (x, z) is the state and δ is the policy, that is $r^i[(x, z), \delta]$ is the immediate reward for objective i when (x, z) is the state and policy δ is being followed. Let γ be a two-dimensional vector such that $0 \leq \gamma^i \leq 1$ and $\sum_i \gamma^i = 1$. Let f be any real-valued two-dimensional vector, and define the following operators:

$$[L_\delta f](x, z) = r[(x, z), \delta]\gamma + \alpha P_{(x, z)}[\delta]f \tag{3.1a}$$

and

$$[Uf](x, z) = \underset{\delta}{\text{maximum}}\ [L_\delta f](x, z) \tag{3.1b}$$

Let $\bar{f}$ be the unique vector where equality is obtained in (3.1b), and let δ^* be a maximizing policy. Then Henig [3, p. 60-61] proves that δ^* is a Pareto optimal policy if and only if it is a policy where the maximum is obtained in (3.1b) for some value of γ. Writing out (3.1) specifically in our context, a policy is Pareto optimal if and only if it achieves, for some γ, the maximum of:

$$\bar{f}(x, z) = \underset{y \in Y(x)}{\text{maximum}} \{\gamma g(x, y) - (1 - \gamma)\lambda \cdot |x - y - z| + \alpha E\bar{f}(s[y, D], x - y)\}. \tag{3.2}$$

White and Kim [11] show that by including γ in the state vector, (3.2) can be treated as a partially observed Markov decision process, and use the special structure inherent in this problem to derive an efficient version of Sondik's [10] algorithm for partially observed MDPs.

The algorithm in White and Kim [11] solves a discrete state and action version of (3.2) simultaneously for all values of γ in the interval [0, 1]. This readily can be seen to be the equivalent of letting λ vary in (2.5). Therefore, without loss of generality, λ can be set equal to one (which can also be seen by rescaling the return function g at the outset).

Given the set of Pareto optimal policies, which must be finite in number, the ergodic probability distribution, assuming an irreducible transition matrix, can be calculated readily. Let π_0 be an initial guess at this distribution, and calculate successively for a given δ^*:

$$\pi_n = \pi_{n-1}^T P[\delta^*] \tag{3.3}$$

where T denotes the vector transpose. The iterations stop when the maximum difference between π_n and π_{n-1} is at a satisfactory level.

While the iterations described in (3.3) are straightforward, they are inefficient and impractical. Suppose the population size is defined on a 25-point grid. Then there are 25 possible harvest sizes, so that there are 625 states. This implies for any policy δ^* that $P[\delta^*]$ has 390,625 entries, which would require roughly 1,562K of computer storage. However, $P[\delta^*]$ is sparse since if $x - \delta^*(x, z) = z'$, then a transition only can be made to states (x, z') next period. Permute row and columns of $P[\delta^*]$ so that the columns run through all values of x for each fixed value of z, and order the rows so that all states that harvest the same amount

Table 1 when following δ^* are together. This yields a blocked matrix as in Table 1. A solution procedure is to iteratively solve (3.3) for each block in order until convergence is reached. This is precisely block Jacobi iterates as described in Young [12]. The advantage is both a reduction in computer storage, since only the smaller blocks are used one at a time, and also an acceleration of convergence [12, chapter 14].

4. EXAMPLE

Mathews [4] has suggested the following model for salmon runs on the Wood River off Bristol Bay, Alaska. Let x_t be the number of recruits at the start of period t, and y_t the number of spawners at the end of period t. Then:

$$x_{t+1} = \exp\{D_t\}4.077y_t\exp\{-0.800y_t\}$$
$$D \sim \text{Normal } (0, 0.2098) \quad (4.1)$$

The example problem is defined on a grid of 15 equally spaced points on (0, 5] in units of 10^6 fish. The procedure used is described in detail in [7]. The constraint set Y(x) is assumed to be $Y(x) = \{y : 0 \leq y \leq x\}$, α is arbitrarily set at 0.97, and the primary one-period return is simply yield, $x_t - y_t$, while the second objective is $-\lambda \cdot |x_t - y_t - z_{t-1}|$ with $\lambda = 1$. In applications, it will be necessary to vary α to test the sensitivity of the results to the change in the discount factor.

The mean-standard deviation trade off curve using the techniques

Fig. 1 of section 3 is shown in Figure 1. The points where γ = 0.0, 0.25, 0.5, 0.75, 1.0 are identified in order to show how the trade off varies with γ. It is clear from Figure 1 that substantial reductions can be made in the standard deviation with only minimal reductions in the mean per period harvest. At $\gamma = 0.75$ as compared with $\gamma = 1$, the mean per period harvest has only been reduced from 1.26×10^6 to 1.2×10^6 fish (a reduction of 4.76%), while the standard deviation has been reduced from 0.83×10^6 to 0.70×10^6 fish, a reduction of 15.66%.

Fig. 2 Figure 2 shows the change in the ergodic distribution for selected values of γ. As γ decreases, the tails become shorter, the mode higher and lower. The distribution is bimodal in the approximate range 0.8-0.6, and becomes extremely steep and narrow as γ nears zero.

Fig. 3 Figures 3(a)-(e) show an optimal policy for these same values of γ. The figures are read by finding the appropriate (x, z) point on the figure, following the arrow in that region, and the resulting z value is an optimal harvest for that state. Region I is the region in which the harvest size decreases, in region II the harvest size remains the same, and in region III the harvest size increases. Figures 2 and 3 show that as γ varies from 1.0 to 0.0, region II increases at the expense of region III, until region III disappears entirely, and while region I does not change in size, the probability of ever being in that region decreases to zero.

DISCUSSION AND SUMMARY

New results on multiobjective Markov decision problems have been combined with an approach to smoothing harvests to produce a technique for determining the mean-variance trade off among the "best" subset of harvesting policies. The technique has been found to be computationally reasonable, and produces the maximum amount of information for the decisionmaker.

The suggested technique greatly extends the analysis in Beddington and May [1] and May et al. [5]. While the particular example used to illustrate the algorithm has Gaussian noise, the general formulation is not restricted to this class of models as in the cited references. Moreover, in these references, they set the growth term equal to zero, and assume that the one-period distribution of returns from any policy is equivalent to the long-run distribution of returns when following the same policy (i.e., they examine behavior when $dx/dt = 0$). This assumption is very often false. The proposed optimization methods calculate explicitly the long-run distribution that arises when following a given policy. Also, by restricting the decision set $Y(x)$, the proposed models include as a subset models that require fixed effort or fixed harvests. However, since $Y(x)$ need not be so restricted, the class of decision processes that can be considered is much richer.

Moreover, fixed effort or fixed harvest policies arise from conclusions based on deterministic models of population dynamics, conclusions that may not be appropriate for randomly varying populations. (In practice, most fisheries management, for example, varies the estimate of maximum sustained yield each year as the population changes.) By not being restricted to fixed effort or fixed harvest policies, it is possible to determine what the loss in value is from using such policies. The example in section 4 suggests that the best fixed harvest policy reduces significantly the average per period harvest. It is proven analytically in [2, 8, 9] that <u>optimal</u> harvesting policies for randomly varying populations will rarely be of the fixed effort or fixed harvest type.

The determination of the mean-variance trade off is only as good as the model that is used in the analysis. However, by providing an easy to understand trade off curve as well as much ancillary information about other characteristics of the policies under consideration, the decision-maker is left in the best position to integrate this knowledge with other information available, both subjective and objective in nature.

ACKNOWLEDGMENT

I am extremely grateful to Chelsea C. White and K. W. Kim for allowing me to see a preliminary version of their paper, and for several invaluable discussions on implementing their algorithm. Mordechai Henig also supplied me with a copy of his Ph.D. dissertation, for which I am grateful.

REFERENCES

1 J. R. Beddington and R. M. May, Harvesting natural populations in a randomly fluctuating environment. Science (Wash. D.C.) 197:463-465,

2 A. Gleit, Optimal harvesting in continuous time with stochastic growth. Math. Biosci. 41:111-123 (1978).

3 M. Henig, Multicriteria dynamic programming. Ph.D. Dissertation, Yale University (1978).

4 S. B. Mathews, The economic consequences of forecasting sockeye salmon runs to Bristol Bay, Alaska: A computer simulation study of the potential benefits to a salmon canning industry from accurate forecasts of the runs. Ph.D. Dissertation, Univ. Wash. (1967).

5 R. M. May, J. R. Beddington, J. W. Horwood, and J. G. Shepherd, Exploiting natural populations in an uncertain world. Math. Biosci. 42:219-252 (1978).

6 R. Mendelssohn, Harvesting with smoothing costs, submitted for publication.

7 R. Mendelssohn, Using Markov decision processes and related techniques for purposes other than simple optimization Analyzing the consequences of policy alternatives on salmon runs, To appear, Fishery Bulletin, U.S.

8 R. Mendelssohn and M. J. Sobel, Capital accumulation and the optimization of renewable resource models, To appear, Journal of Economic Theory.

9 W. J. Reed, A stochastic model for the economic management of a renewable animal resources. Math. Biosci. 22:313-337 (1974).

10 E. J. Sondik, The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. Oper. Res. 26:282-304 (1978).

11 C. C. White and K. W. Kim, Two solution procedures for solving vector criterion Markov decision processes. To appear in *Large Scale Systems*.

12 D. M. Young, *Iterative Solutions of Large Linear Systems*, Acad. Press, New York, 1971.

TABLE 1.--Block form of the transition matrix.

	$(x_1, z_1)\ (x_2, z_1) \ldots (x_N, z_1)$	$\ldots\ldots$	$(x_1, z_i)\ (x_2, z_i) \ldots (x_N, z_i)$	$\ldots\ldots$	$(x_1, z_N)\ (x_2, z_N) \ldots (x_N, z_N)$
States that have z_1 as optimal harvest $\{(x, z)$	$P_{(x, z)}(\delta^*)$		$\bigcirc$		$\bigcirc$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
States that have z_i as optimal harvest $\{(x, z)$	$\bigcirc$		$P_{(x, z)}(\delta^*)$		$\bigcirc$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
States that have z_N as optimal harvest $\{(x, z)$	$\bigcirc$		$\bigcirc$		$P_{(x, z)}(\delta^*)$

where $P_{(x, z)}(\delta^*)$ are the appropriate rows of $P(\delta^*)$ for an optimal policy δ^*.

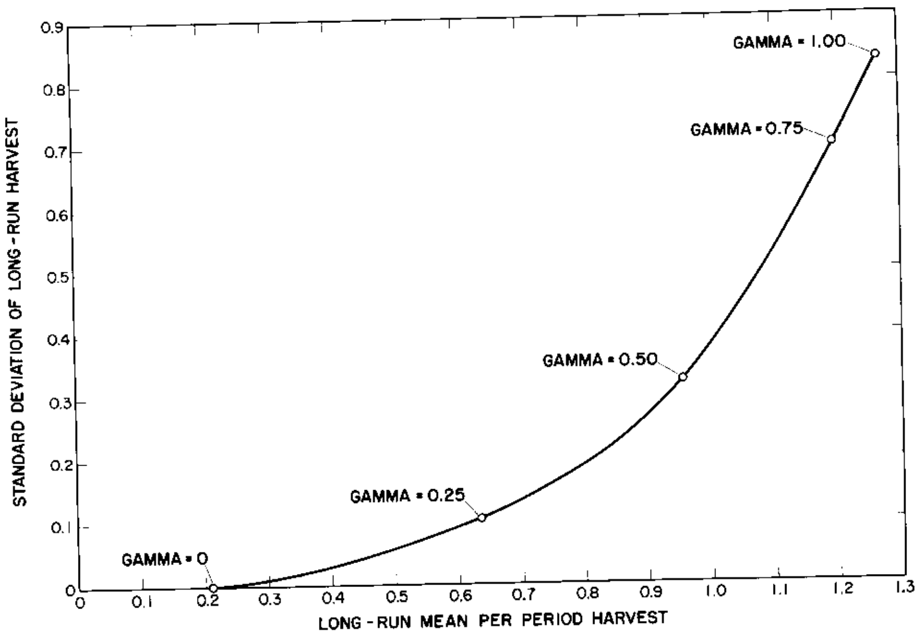

Figure 1.--Long-run (ergodic), mean-standard deviation trade off for the salmon model.

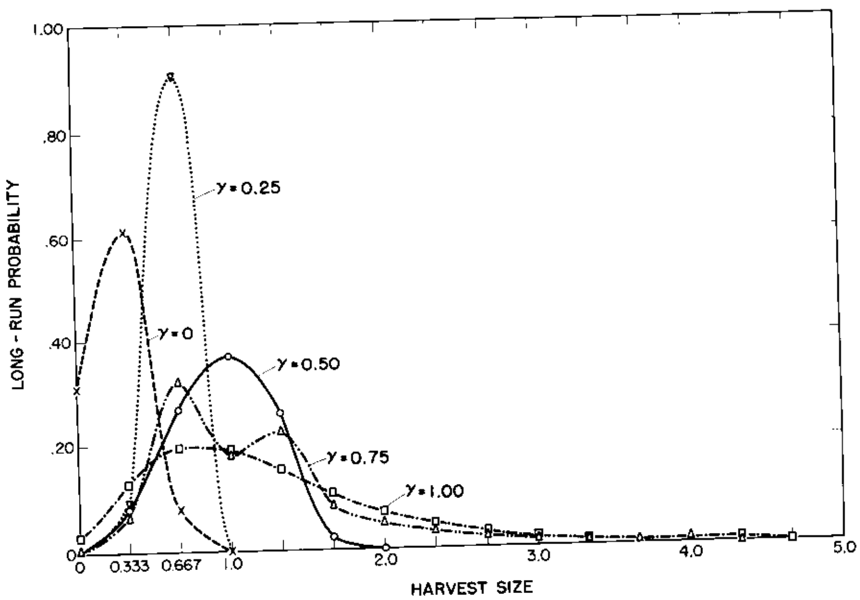

Figure 2.--Long-run (ergodic) distribution for the salmon model as γ varies.

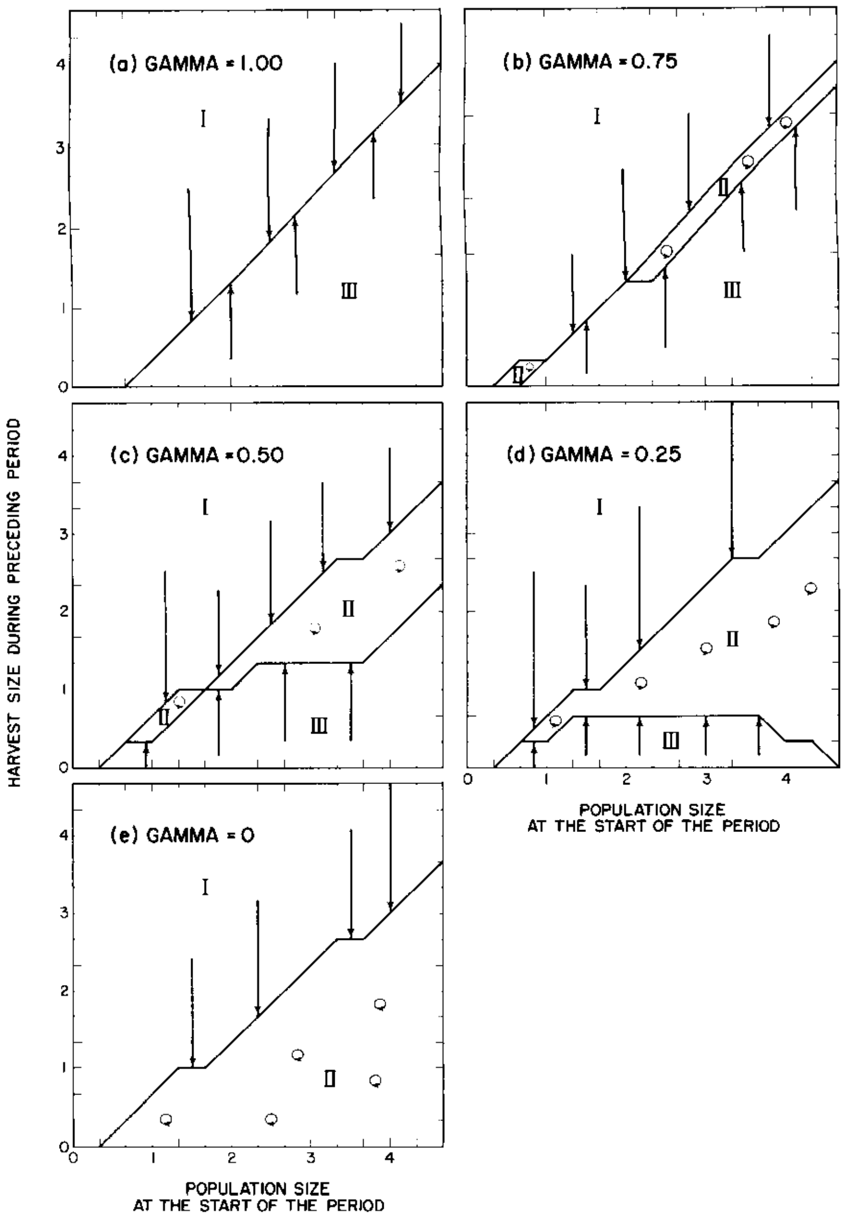

Figure 3.--Optimal policy functions for the salmon model for γ fixed at five levels.