A PROGRAM OPTIMIZATION SYSTEM FOR THE CLEANUP OF DOE HAZARDOUS WASTE SITES AN APPLICATION TO FY 1990 FUNDING DECISIONS

Miley W. Merkhofer
Applied Decision Analysis, Inc.
Menlo Park, California

Thomas A. Cotton
JK Associates, Inc.
Silver Spring, Maryland

Karen E. Jenni
Applied Decision Analysis, Inc.
Menlo Park, California

John C. Lehr, Director
Hazardous Waste and Remedial Actions Division
U.S. Department of EnergyDefense Programs
Washington, D.C.

Thomas P. Longo
Hazardous Waste and Remedial Actions Division
U.S. Department of EnergyDefense Programs
Washington, D.C.

ABSTRACT

This paper describes a formal system used by the Department of Energy (DOE) as an aid for allocating funds for cleaning up hazardous waste sites. The system, called the Program Optimization System (POS), is based on multiattribute utility analysis and was developed for DOE's Hazardous Waste and Remedial Actions Division (HWRAD). HWRAD has responsibility for recommending Environmental Restoration (ER) activities to the Assistant Secretary of Energy. Recently, the POS was used to analyze and recommend funding levels for FY 1990 cleanup activities at DOE Defense Program facilities.

INTRODUCTION

With oversight provided by seven field offices, the DOE operates 17 facilities that produce or support the production of defense-related nuclear weapons and materials (Table I). Operating these facilities has resulted in radioactive and hazardous chemical wastes whose storage or disposal at the facilities has, in many instances, produced locations or areas that have been designated inactive hazardous waste sites. These waste sites are now subject to regulation under the Resource Conservation and Recovery Act (RCRA) and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). It has been estimated that the cost of bringing the sites at these facilities into compliance with current environmental standards could be as high as $50 billion over a 25-year period.

Cleaning up the hazardous waste problems presents several challenges for the DOE, including justifying the allocation of the necessary funds, setting priorities, and defending funding decisions. Recognizing the need for a systematic, rational approach to decision making, Congress directed DOE to establish a priority system that would allow decisions to be made "so as to achieve the greatest benefits to health, safety, and protection of the environment from further damages."

To develop the system requested by Congress, a task group consisting of DOE and outside experts was established. This group undertook a deliberate process involving: (1) review of existing prioritization systems, (2) development of a "wish list" of desired system characteristics, (3) preparation of an options paper outlining four alternative system designs, (4) selection of a preferred design by HWRAD management, (5) trial implementation of the preferred design on a pilot scale, and (6) full-scale application to the FY 1990 ER budget. Two full-scale applications were conducted to meet the needs of the budgetary planning cycle. A first application was conducted to help establish the total ER budget request, which must be made early in the planning cycle. The second application was conducted roughly six months later to provide a basis for finalizing allocations among field offices. The second application was meant to update the initial analysis and to account for revisions in planned FY 1989 funding levels, which occurred after the initial application. The numerical results displayed in this paper represent the second, revised FY 1990 application and differ slightly from those presented in an earlier paper based on the initial application only (1).

APPLICATION OF THE POS TO FY 1990 BUDGETING DECISIONS

Development and application of the POS to FY 1990 budgeting decisions involved six steps:

Step 1: Specify Funding Objectives

The first step was to specify the basic objectives to be achieved through the funding of environmental restoration activities at DP facilities. Candidate objectives were identified in a series of meetings of the task group, considering relevant statutory language, mission statements,

TABLE I

LIST OF DOE FACILITIES SCORED IN THE POS STUDY

Field Office	Facility
Albuquerque (AL)	Kansas City (KC)
	Los Alamos (LANL)
	Mound (Mnd)
	Pantex (Pan)
	Pinellas (Pin)
	Rocky Flats (Rock)
	Sandia - Albuquerque (S-A)
	Sandia - Livermore (S-L)
Idaho (ID)	Idaho (INEL)
Nevada (NEV)	Nevada Operations Office (Nev)
Oak Ridge (OR)	Fernald (FMPC)
	Oak Ridge Gaseous Diffusion Plant (ORGDP)
	Oak Ridge (ORNL)
	Y-12 (Y-12)
Richland (RL)	Hanford (Han)
San Francisco (SAN)	Lawrence Livermore (LLNL)
Savannah River (SR)	Savannah River (SavR)

and policy guidelines. Table II lists the four objectives considered most important for guiding ER funding decisions -- public health and safety, regulatory requirements, public concerns, and costs. In addition, one constraint was established for the analysis -- to avoid program funding levels that pose a threat to the continued execution of important elements of the DP mission.

Step 2: Develop Scales for Measuring Performance Against Objectives

As summarized by Table III, measurement scales were developed to quantify the attainment of objectives. To quantify future costs, two cost measures were defined. The first, remaining costs to complete ongoing remedial actions, was established to account for the "mortgage" associated with undertaking multiyear remedial actions. Since the benefits of such actions are assessed assuming completion of the activity, remaining costs must be estimated to ensure that such benefits are balanced against their full costs.

The second cost measure, impact on the costs of future activities, is designed to account for problems that become more expensive to solve with time. An example is a site with a contaminant plume that is spreading. If reduced funding causes cleanup delays that increase ultimate cleanup costs, then an estimate of the increase, compared to what the costs would be without such delays, is provided through use of this scale.

For the objectives other than cost, no natural measurement scales exist. It was, therefore, necessary to construct scales by defining various levels of impact. Constructed scales were developed in a series of meetings involving decision analysts and individuals with expertise in fields relevant to the objectives. The final versions of these scales are such that risk to public health and safety and level of public concern are measured on 1-to-5 scales. Regulatory responsiveness is measured on a 1-to-7 scale. Threat to the DP mission is a no/yes (zero/one) scale. In the case of all constructed scales, the lowest possible score signifies the best state of affairs (e.g., no significant risk to the public), while the highest score signifies the worst state of affairs that might realistically exist (e.g., a level of public health risk that is as high as that resulting from any DOE DP Defense Programs facility in the country).

For illustration, Table IV shows the scale developed to measure regulatory responsiveness. The scale, which was developed with the help of DOE legal counsel, indicates gradations in the seriousness of failures to achieve various types of regulatory requirements. To illustrate the nature of the scale for public health and safety, Table V

TABLE II

MAJOR OBJECTIVES AND CONSTRAINT IDENTIFIED FOR DEFENSE PROGRAMS

ENVIRONMENTAL RESTORATION FUNDING DECISIONS

Objectives:

1. Minimize health and safety risk to members of the public
2. Minimize potential for being judged nonresponsive to regulatory requirements
3. Minimize current and future public, state, and community concern
4. Minimize future costs

Constraint:

1. Avoid threats to the DP mission

NOTE:The word "minimize" is used to indicate the direction of desired movement in the case of objectives that must be traded off against one another.minimize health risks and other undesired consequences while at the same time minimizing costs, tradeoffs are necessary. The word "avoid" is used to indicate that funding alternatives that would result in threats to the DP mission were judged unacceptable and not considered by the POS.

TABLE III

SCALES USED TO QUANTIFY PERFORMANCE

Objective or Constraint	Measurement Scale	Units
Health and Safety	Constructed (Table V)	Number between 1 & 5
Regulatory Reponsiveness	Constructed (Table IV)	Number between 1 & 7
Public Concern	Constructed	Number between 1 & 5
Future Costs Remaining costs to complete ongoing remedial actions	Millions of dollars complete ongoing remedial actions	
	Impact on costs of future tasks	
DP Mission	Constructed	Zero or one

shows the description corresponding to a score of 3 (moderate level of risk). The table illustrates that there are several alternative justifications that might be adopted to substantiate a judgment that a specified health and safety score is appropriate. To facilitate the application of this scale, an estimate of facility-wide health risks derived from a health risk simulation model called RAPS (2) was provided to each facility.

Application of the health risk scale, as well as the other scales, requires an absolute judgment of the level of risk. However, the measure of benefit is the extent to which the ER program at the facility decreases the risk score. Thus, what is important is not the absolute level of the score at any one proposed budget level, but how the different funding levels would change the score.

Step 3: Score Alternative Facility Funding Levels Against the Measurement Scales

Each facility evaluated four to six alternative funding levels for FY 1990. All facilities scored at least four funding levels:

1. A "null" program -- no environmental restoration activities to be funded at the facility in the budget year.
2. A minimum program -- activities that would be undertaken if the budget was set at the minimum level that would produce some good. Unless otherwise specified,

the minimum program was defined to be a level of funding equal to 80% of the target program.

3. A target program -- activities that would be undertaken if the proposed budget was funded.
4. A maximum program -- activities that would be undertaken if the proposed budget was increased by 100% or more so as to meet all program needs.

In addition, facilities with an increment of $10M or more between their minimum and target programs were asked to score a decremented program, defined as a funding level halfway between their minimum and target program. Facilities with a $10M or more increment between their target and max programs were asked to define and score an enhanced program, halfway between their target and max programs.

Field office personnel were asked to identify the environmental restoration activities that would be conducted in FY 1990 under each program funding level. Assuming these activities, field office personnel then independently scored each funding level, following a detailed set of written scoring instructions and using the constructed scales and cost definitions discussed above.

As a quality assurance check, a three-day workshop was held to finalize the scores. Each facility's scores were finalized when workshop attendees agreed that the scores seemed appropriate and defensible. Figure 1 provides an example of how finalized scores were displayed. It shows the health and safety scores for various program funding levels at four of the facilities.

Such plots were found useful because they clearly illustrate differences in the ability of facilities to improve their scores through increased funding. Facilities whose scores decline with increased funding are able to make immediate progress in solving their problems. Facilities whose scores do not decline might still use increased funds effectively (e.g., to better comply with regulatory

TABLE IV

Performance Measure Scale for Potential for Being Judged

NONRESPONSIVE TO REGULATORY REQUIREMENTS

1 =	The program funding level makes it almost certain that all legally binding and nonbinding regulatory requirements and agreements can be met during FY90.
2 =	The program funding level makes it likely that during FY90 DOE will be unable to meet some nonbinding requirement or agreement (e.g., a consent agreement without enforceability provisions, or anot-legally-applicable state law that DOE has agreed to observe as a matter of comity). If that occurs, complaints are likely to be raised.
3 =	The program funding level poses a significant probability (greater than 50%) that during FY90 or FY91 there will be a clear perception that DOE has not met some important legally binding requirement (e.g., an interagency agreement with enforceability provisions, an RCRA Part B permit schedule, or a legally applicable state law or regulation).
4 =	The program funding level makes it almost certain that during FY90 DOE will be unable to meet some important legally binding requirements (described in 3).
5 =	It is highly likely that an important legally binding requirement (described in 3) will have been missed prior to FY90, and the program funding level is inadequate to rectify the situation. If that occurs, a court order to comply may result.
6 =	The program funding level is almost certainly inadequate to allow timely compliance with all of the requirements that have already been imposed by a court order.
7 =	It is highly likely that a requirement imposed by a court order will have been missed prior to FY90, and the program funding level is inadequate to rectify the situation.

TABLE V

A Portion of the Performance Measure Scale for Health and Safety Risk -- The Definition of a Score of 3

3 = The public health and safety risks occurring under the program funding level are MODERATE. If the various defense facilities were ordered according to the level of public health and safety risk that each poses, then this facility would be between the 40 percent and 60 percent fractiles (i.e., it would be among 20 percent of fractiles closest to the average in terms of public health risk), assuming the program funding level. Thus, either the health and safety risks attributable to the sites are inherently only moderate, or they are reduced to moderate levels as a result of the activities included in the program. The conclusion that the risks are moderate is substantiated by calculations using data and/or models.

- If models are used to substantiate the conclusion that risks are moderate, then it must be the case that when the models are applied using inputs consistent with the conditions and scenarios that might reasonably exist under the program, the results indicate current and future impacts to public health (attributable to the sites) that are about the average of what are typically computed.

- Suppose the aggregate HPI (Hazard Potential Index) score for all sites associated with the facility was calculated using appropriate inputs. The result would have to be around 60 to add credibility to the judgment that risks are moderate.

- If an HRS (Hazard Ranking System) score was computed for the facility using appropriate inputs, the result would be around 40.

- If such calculations were to be made, the individual risk to the most highly exposed member of the public would be less than 10^{-4}. Also, if calculated, the total expected annual health effects (computed by weighing estimated health effects under alternative scenarios by the probabilities of the scenarios) would be around 0.1.

- Alternatively, the conclusion of moderate risk might be justified if it can be convincingly argued that there are low-probability scenarios producing significant public health impacts, or, if the probability of such scenarios is moderate, then the only significant health effects that would occur would take place in the distant future.

requirements), but allocating money for the cleanup of such sites will not reduce health risks.

Step 4: Develop a Multiattribute Utility Function for Aggregating Scores

According to multiattribute utility theory, an additive equation for the utility function is appropriate if the measures established for objectives are additive independent (3). By applying standard tests, the following additive form was justified:

$$U = w_{hs}U_{hs}(x_{hs}) + w_{rr}U_{rr}(x_{rr}) + w_{pc}U_{pc}(x_{pc}) - w_{rc}C_{rc} - w_{fc}C_{fc} \quad \text{(Eq.1)}$$

The x's are the program scores on the performance measurement scales (hs = health and safety, rr = regulatory responsiveness, pc = public concern), and the U's are component utility functions reflecting the value to decision makers of each additional unit of improvement on the scales. C_{rc} and C_{fc} are remaining costs and impacts on future costs, respectively. The w's are positive scaling factors representing value tradeoffs between units of the corresponding performance measures and FY 1990 costs.

The component utility functions and scaling factors were elicited from the Director of HWRAD using standard decision analysis techniques. The base-case scaling factors were consistent with a willingness to spend up to:

- $3.5 billion dollars in additional costs to eliminate extremely high health risks posed by a facility (i.e., to take a health and safety risk score of 5 and reduce it to 1).
- $175 million in additional costs to eliminate severe regulatory problems at a facility (i.e., to take a regulatory responsiveness score of 7 and reduce it to 1)
- $50 million in additional costs to eliminate very high public concern at a facility (i.e., take a public concern score of 5 and reduce it to 1).

The base-case scaling factor for public health and safety is approximately consistent with a judgment that a statistical public fatality is as undesirable as an additional $5 million in FY 1990 costs. This tradeoff is similar to that used in another prominent DOE analysis (4). Ranges of tradeoffs judged reasonable were also elicited and served as a basis for developing alternative scaling factors for sensitivity analysis.

Plots of utility versus FY 1990 costs were generated for each facility, with utility computed using Eq.(1). Figure 2, which shows the base-case results for two of the facilities, illustrates the nature of the insights generated.

In the case of some facilities, such as the Y-12 plant on the Oak Ridge Reservation in Tennessee, utility increases steadily with program costs. In the case of other facilities, such as the Los Alamos National Laboratory, large increments to utility are only obtained at certain funding levels. In most cases, large positive increments in program utility (large positive slopes) reflect the addition of a critical activity to the program that produces significant benefits. Conversely, small positive increments or negative increments (low slopes or negative slopes) reflect increased funding for activities producing small benefits or the addition of activities having very large remaining costs. Inspection of plots such as those of Fig. 2 was judged useful in suggesting alternative program mixes that are more effective at meeting objectives.

Step 5: Implement the Methodology as a Computer Program

HQ funding allocations can consist of any combination of facility-funding programs. Since there are 17 facilities and each could, in theory, have any one of the five possible funding levels, there are 517 or nearly one

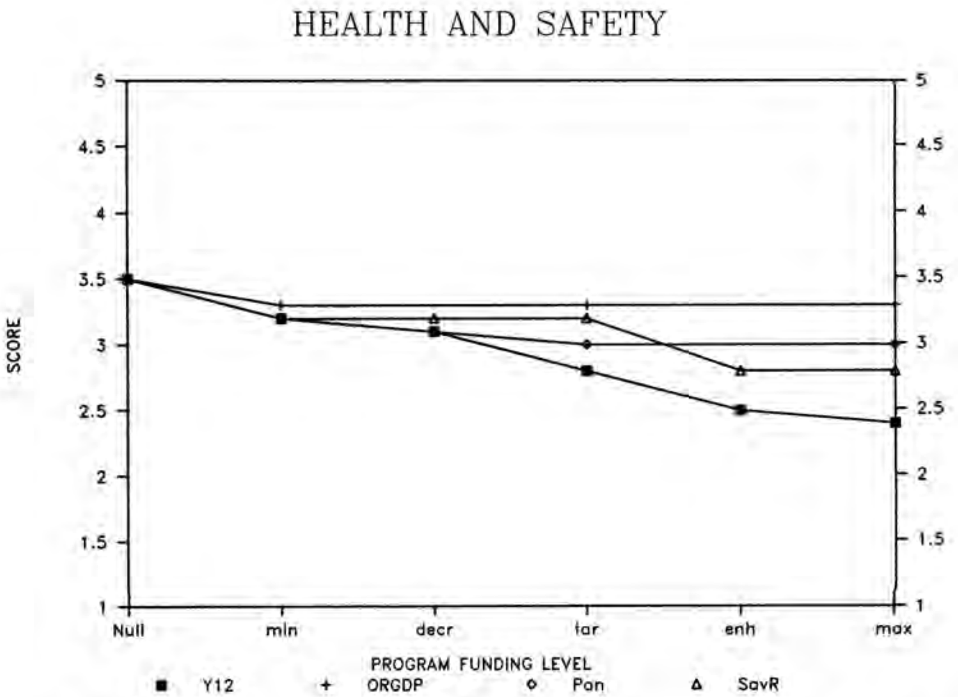

Fig. 1. Health and Safety Scores for Y-12, Oak Ridge Gaseous Diffusion Plant, Pantex, and Savannah River Facilities.

trillion funding possibilities. Obviously, a computer is needed to evaluate and compare the options.

In designing the computer program it was recognized that one trillion combinations would be time-consuming if not impossible to evaluate, and if even a small fraction were designated as optimal, the results would be difficult for decision makers to interpret. Consequently, to simplify the analysis, three exclusionary rules were used to eliminate from consideration combinations that were judged clearly undesirable. First, funding combinations that cause any facility to have a level of funding creating a threat to the DP mission were eliminated. Second, combinations that cause any facility to have zero FY 1990 funding (the null program) were excluded. Third, "inequitable" distributions -- that is, distributions in which some facilities receive funding levels far above their proposed targets while others receive funding levels below their

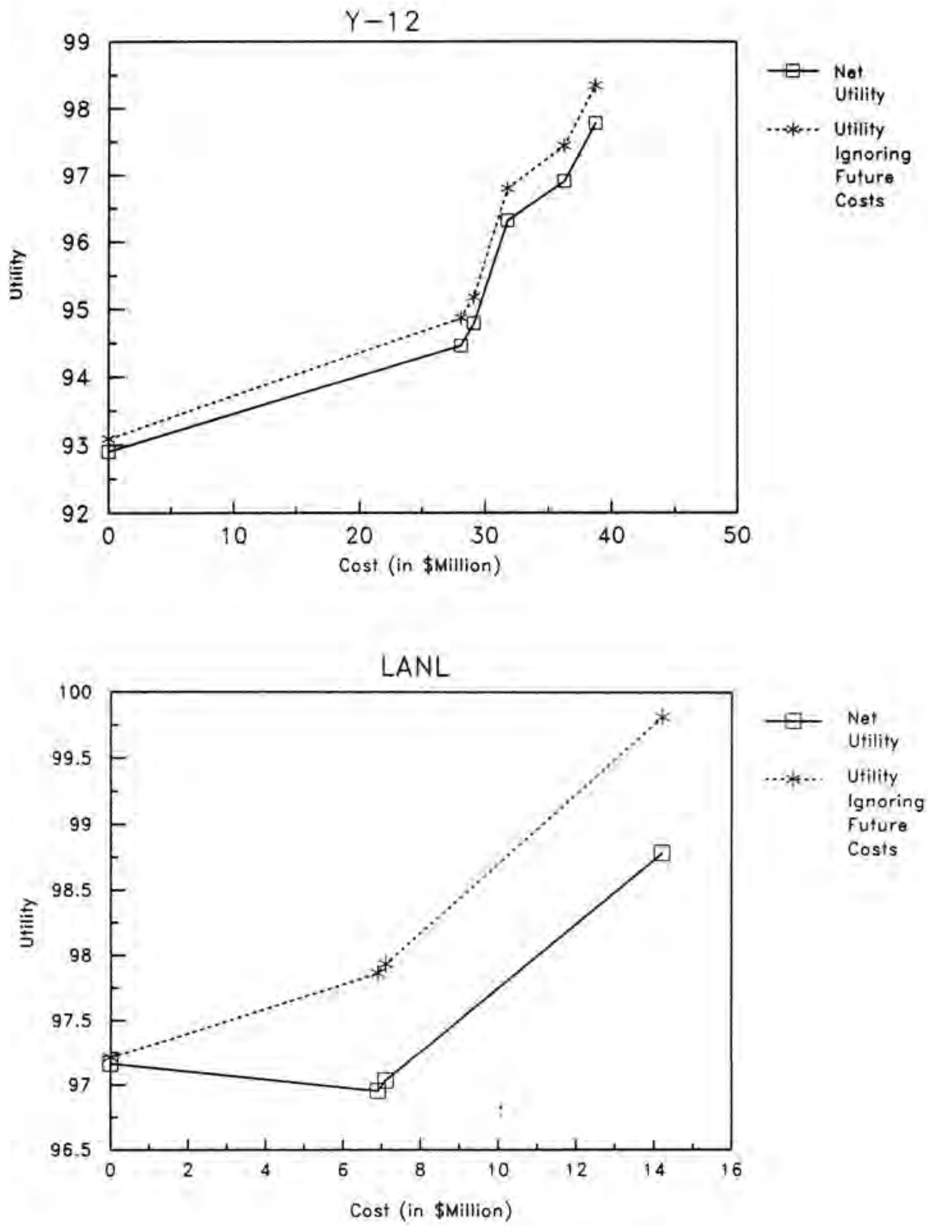

Fig. 2. Multiattribute Utility of Alternative Program Funding Levels at Two Facilities (Y-12 and Los Alamos National Laboratory) -- Base Case.

proposed targets -- were excluded unless such distributions produce a significantly higher aggregate utility.

To implement this last rule, a penalty function approximately representing the inequity of the allocation was devised. The penalty function was defined as the variance of the allocation times a scaling factor, where the variance between two facilities was defined to be 0 if both facilities receive the same program (e.g., both target), 1 if they receive adjacent programs (e.g., one receives its target and the other receives its enhanced program), 4 if they receive programs that are two steps apart (e.g., one receives its target and the other its maximum program), and so forth.

A computer program for use on a personal computer was developed to calculate the aggregate utility of each funding combination not excluded by the above rules by adding the utilities associated with the specified funding levels at the facilities. That combination that produced the greatest aggregate utility (minus the inequity penalty) for each given total (summed) cost was then identified as an "optimal" allocation. The scaling factor for the penalty function was set empirically so as to eliminate about half the optimal allocations (400 rather than 800). With this scaling factor, the utility of eliminated combinations was at most only a few hundredths of a utile (less than a million dollars in benefits) greater than the more equitable combinations that were not eliminated.

Figure 3 shows the aggregate utilities and equivalent monetary benefits of the optimal allocations plotted versus total HQ cost. To aid interpretation, equivalent monetary benefits (consistent with the tradeoff judgments) are also given. Straight lines have been drawn through the points to help identify "knees" in the curve total HQ funding levels where the average slope (incremental benefit obtained for each additional dollar spent) decreases abruptly. The curve provides a useful aid for identifying cost-effective overall budget levels. For example, the curve suggests that funding levels beyond about $317 million are not cost-effective because above this level each dollar invested returns only about 40¢ in additional benefit.

Tabular summaries, listing the recommended allocation for each total cost, were also provided. The results suggest that some facilities should be funded at or below their target levels, while others should be funded at levels far above target levels.

Figure 4 shows a graphical summary of the base-case results. It shows the allocation of funds among field offices implied by the optimal combinations. For any given total HQ funding along the x-axis, the y-axis shows how those funds are divided among the field offices.

Step 6: Sensitivity Studies

The sensitivity of POS recommendations to assumptions and variations in inputs was explored through numerous sensitivity studies. First, variations in scores were explored to determine the extent to which scoring errors might alter results. Second, the scaling factors were varied to reflect a range of tradeoffs viewed as reasonable. Third, the magnitude of the inequity penalty was varied to determine its influence. Finally, the effect of allowing intermediate funding levels (e.g., a level between enhanced and maximum programs at Savannah River) was explored to determine whether facilities with larger funding needs might be subject to a bias resulting from the large funding demands of their programs.

In summary, the sensitivity studies showed that the results of the POS are relatively stable. Small changes in scores, costs, or weights produce small and predictable changes in POS results. This enhanced confidence that the results provide a useful input to funding decisions.

CONCLUSIONS

The POS has several attractive features. It makes use of the field offices' detailed knowledge of conditions at their sites and allows field office flexibility in the design of programs to deal with these conditions. Judgments regarding the consequences of alternative funding levels are clearly articulated and communicated between the field offices and HQ. Funding allocations are based on an evaluation of those consequences using explicit value judgments specified by individuals with appropriate policy responsibility. Unlike a priority system -- which would require HQ to rank all of the individual projects proposed by the field components and then simply drop off the lowest-ranked projects that exceed the available budget -- the POS does not require "all or nothing" decisions with respect to individual projects. Instead, the POS allows each field office to propose different levels of effort for a project or activity, depending on the budgets available. Explicit consideration of such partial funding options allows field offices to propose the most cost-effective programs achievable for each budget level.

Due to its approximations and simplifications, the POS produces results that can be considered only guides for decision making. Nevertheless, the results suggest several important insights:

1. Even at relatively high HQ budgets, additional funding increments could be used by field offices to conduct activities estimated to produce significant benefits to health and safety, regulatory responsiveness, or reductions in public concern.
2. Optimal funding allocations depend dramatically on the total HQ budget. At higher HQ budgets, it appears to be efficient to fund some facilities at their target levels and others at levels that are well above their target levels (e.g., in some cases, more than 100% above target).
3. The allocation of funds to facilities fundamentally requires policy judgments regarding the relative importance of health and safety, regulatory responsiveness, and public concern.

It is clear that the process of developing and applying the POS has significantly improved the Department's understanding of the difficult problem of budgeting for environmental restoration at its Defense Programs facilities. Some facility-funding levels were substantially changed, largely in response to the insights generated. The experience of applying the POS to the FY 1990 ER budget

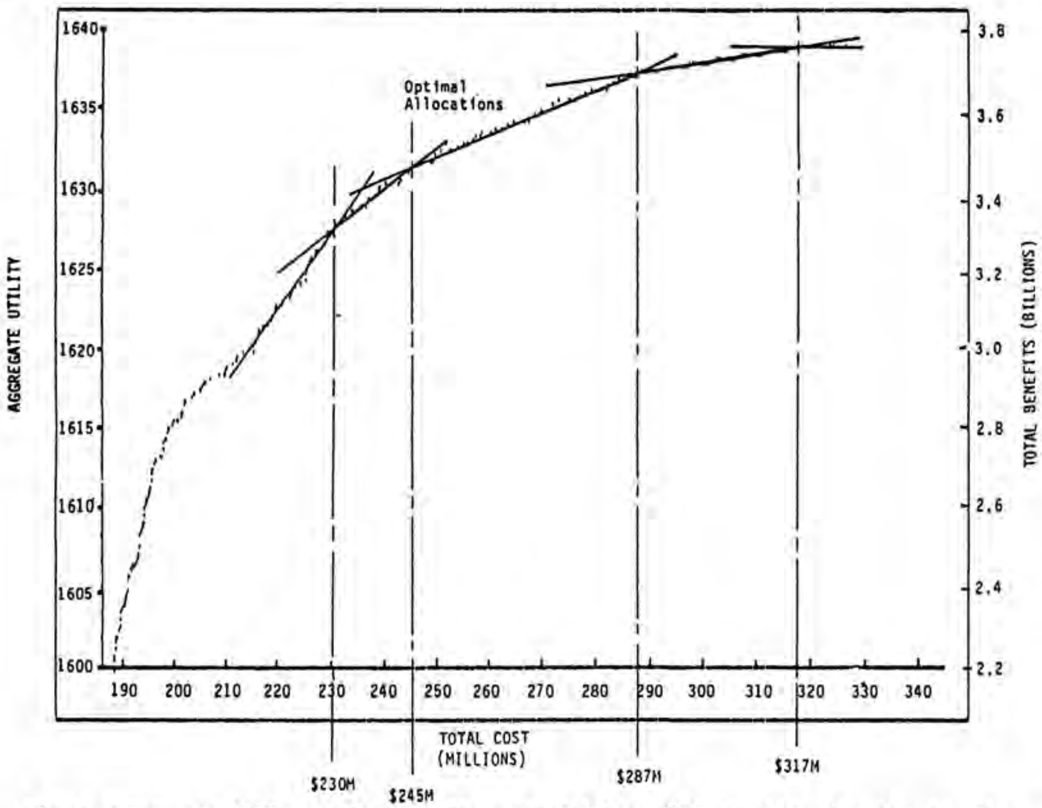

Fig. 3. Aggregate Utility and Benefits Versus Total Cost of Optimal Funding Combinations -- Straight Lines Indicate Levels of Total

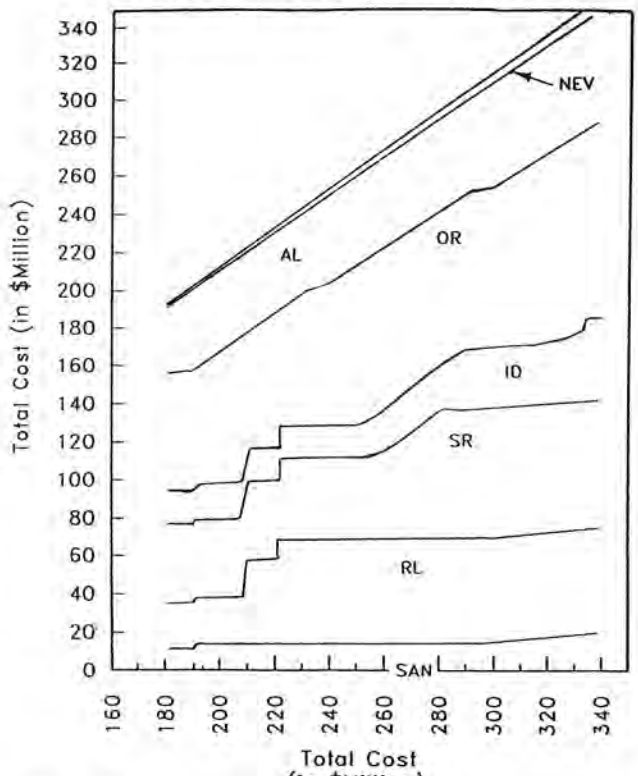

Fig. 4. Optimal Allocations to Field Offices for Various Total HQ Budgets -- Base Case.

has clearly demonstrated benefits for both DOE HQs and field offices. According to the Department's report on the POS to Congress, the

> "POS provides a logical, defensible, auditable, and flexible system for aiding budget decisions. It provides a valuable information base for supporting budget requests and for evaluating the impacts of budget changes."

Furthermore, "By clearly identifying benefit measures to be applied uniformly and consistently in evaluating programs, the POS promotes development of programs that are aligned with DOE HQ objectives for the ER budget. It also creates incentives for the field offices to plan the most cost-effective mix of activities at each site for each possible budget level."

ACKNOWLEDGEMENTS

The application summarized by this paper was funded by the Hazardous Waste and Remedial Actions Division (HWRAD) of the U.S. Department of Energy. Many individuals participated in the analysis. Dr. Critz George, Director of HWRAD at the time of the study, and Mr. Jack Petty of Martin Marietta Energy Systems for Waste Management Research and Development were, in addition to the authors, major contributors.

In preparing this paper, Dr. Merkhofer was partially supported by NSF Grant No. 8606906. The opinions expressed by the authors are solely their own and do not necessarily represent the views of the National Science Foundation or the Department of Energy.

REFERENCES

1. M. W. Merkhofer, T. A. Cotton, and T. P. Longo, "A Program Optimization System for Aiding Decisions to Fund the Cleanup of Hazardous Waste Sites at Department of Energy Defense Facilities," pp. 39-43 in Superfund '88, Proceedings of the 9th National Conference sponsored by the Hazardous Materials Control Research Institute, held November 28-30, 1988, Washington, D.C.
2. G. Whelan, D. L. Strenge, J. G. Droppo, JR., B. L. Steelman, and J. W. Buck, The Remedial Action Priority System (RAPS): Mathematical Formulations, prepared by Pacific Northwest Laboratory under Contract DE-AC06-76RLO 1830 for the U.S. Department of Energy, Office of Environment, Safety, and Health, DOE/RL/87-09 (PNL-6200 UC-11), August 1987.
3. R. W. Keeney and H. Raiffa, Decisions with Multiple Objectives, Preferences and Value Tradeoffs, Wiley, New York, 1976.
4. M. W. Merkhofer and R.W. Keeney, "A Multiattribute Utility Analysis of Alternative Sites for the Disposal of Nuclear Waste," Risk Analysis 7 (2), pp. 173-194, 1987.