ABSTRACT

In March 1994, the Department of Energy Carlsbad Area Office (DOE/CAO) implemented a performance-based planning method to assist in programmatic prioritization within the Waste Isolation Pilot Plant (WIPP) project with respect to applicable Environmental Protection Agency (EPA) long-term performance requirements stated in 40 CFR 191.13(a) and 40 CFR 268.6. This method, the Systems Prioritization Method (SPM), was designed by Sandia National Laboratories (SNL) to: (1) identify programmatic options (activities) and their costs and durations; (2) analyze potential combinations of activities in terms of predicted contribution to long-term performance; and (3) analyze cost, duration, and performance tradeoffs. SPM results were the basis for recommendations to DOE/CAO in May 1995 for prioritization within the WIPP project. This paper presents a summary of the SPM implementation, key results, and lessons learned.

THE SPM APPROACH

The goal of SPM was to provide information about how potential activities—twenty-one scientific investigations, three engineered alternatives, and two waste acceptance criteria—when viewed singly or in combination, could contribute to a demonstration of compliance with EPA long-term performance requirements for the WIPP disposal system. For each activity set (a combination of activities), SPM calculated the probability of demonstrating compliance (PDC) if the activity set was implemented, along with the activity set’s projected cost and duration. These performance measures (PDC, cost and duration) were contained in a decision matrix that was analyzed to find programmatic options that maximized incremental PDC while minimizing activity set cost and duration. SNL performance assessment models were used to estimate how the disposal system might perform if activities were implemented, and this evaluation was the basis for calculating an activity set’s PDC. SPM analyzed roughly 46,700 activity sets. Probabilistic performance calculations for these activity sets resulted in over 1.3 million complementary cumulative distribution functions (CCDFs).

As applied to the WIPP, SPM can be described in terms of eleven key steps (Figure 1):

1. definition of the performance objective (i.e., long-term performance in 40 CFR 191.13(a) and 40 CFR 268.6);
2. development of a technical baseline for SPM calculations;
3. performance modeling of the baseline;
4. determination of whether the baseline was predicted to succeed or fail in meeting the performance objectives;
5. (if the baseline failed to meet performance objectives), identification of activities that, if implemented, could improve a predicted ability to meet the performance objectives;

6. evaluation of the baseline combined with potential outcomes of activities (i.e., calculation of the probability of demonstrating compliance);

7. creation of a decision matrix containing the PDC, cost, and duration for all activities and subsequent decision analysis to develop final recommendations;

8. DOE/CAO programmatic decisions about which activities to implement, if any;

9. implementation of activities;

10. re-definition of the technical baseline with actual results from the activities, iterating the overall process as necessary until the baseline is predicted to meet performance objectives; and,

11. when the baseline is predicted to comply, final compliance calculations with approved data and models, etc.

---

![Diagram](image-url)  

**Figure 1. Key steps of SPM as applied to the Waste Isolation Pilot Plant.**

A key to understanding how SPM works is in the relation between the output of the performance assessment models, the regulatory performance requirements, and decision analysis methods used to analyze results. It is also important to understand the role of expert judgment in performance assessment calculations.

Performance assessment models are used by the WIPP project to produce performance measures that can be compared to regulatory requirements (see reference 4). One such measure is a CCDF, which represents the probability distribution of summed normalized releases from the disposal system to the accessible environment. The WIPP disposal system is predicted to be in compliance with 40 CFR 191.13(a) if no point on the CCDF submitted in the compliance certification application to the EPA exceeds the summed normalized release limits.
While the regulatory release limits are fixed, estimates of predicted performance of the WIPP disposal system is not; they are determined by a state of knowledge that changes over time. Changing the state of knowledge through scientific investigations, implementing engineered alternatives, or modifying waste acceptance criteria can alter the position of the CCDF with respect to the release limits. Our state of knowledge can be expressed, in part, through probability distributions. For example, while the solubility of plutonium in WIPP brines is not known accurately at the present time, a range of solubilities under various chemical conditions and based on many types of existing information can be postulated, thus defining a portion of the WIPP disposal system technical baseline.

Envision the design of scientific experiments to more accurately determine the solubility of plutonium in brine. The experimental design anticipates a range of possible outcomes based on both published information and expert judgment. For simplicity, suppose that the experimental outcomes can be classified into five ranges (really probability distributions), from lowest to highest solubility. Denote the event that the experimental outcomes are in the first range by \( x_1 \), in the second range by \( x_2 \), etc. Denote the five possible probability distributions corresponding to the five experimental outcomes by \( f_1, f_2, \ldots \). After the experiment is complete, our knowledge of plutonium solubility changes to reflect new information produced by the study. All uncertainty, however, will not be resolved by the experiments. Uncertain repository conditions make it impossible to know with certainty what the solubility will be. Therefore, after the experiments are completed, there is still residual uncertainty about the solubility which can, again, be expressed through a probability distribution that reflects the new evidence and that incorporates new expert judgments.

Now, suppose that we use expert judgment to specify potential experimental outcomes \( x \) and associated probability distributions \( f \), before conducting the experiment and use these distributions in performance assessment models to estimate the corresponding CCDFs: \( \text{CCDF}_1, \text{CCDF}_2, \ldots \). In addition to providing the \( x \) and \( f \), we also use expert judgment to specify the relative likelihood or probabilities of the various events \( \{x_i\} \), denoted by \( p_i \). Suppose that performance calculations predict events \( x_1, x_2, x_3, x_4 \), and \( x_5 \) will indicate compliance with long-term performance requirements but that the event \( x_5 \) will indicate non-compliance. The predicted probability of successfully demonstrating compliance for the five events \( x_1, x_2, x_3, x_4, \) and \( x_5 \)—viewed prior to conducting the experiment—is then \( p_1 + p_2 + p_3 + p_4 \). This process is the fundamental basis for calculating the PDC of an activity set, the key measure of programmatic value in the SPM method. For the WIPP, this technique applies to any activity that can be expressed in terms of effects on WIPP performance assessment components.

Although this discussion has been restricted to compliance with 40 CFR Part 191.13(a), WIPP must also comply with the RCRA regulatory requirements in 40 CFR 268.6. The SPM criterion for success was that the CCDF is at all points less than the release limits and that the RCRA soil concentration limits are not exceeded. The compliance indicator (CI) for each activity set outcome indicates whether the 40 CFR Part 191.13(a) regulatory release limits and the 40 CFR 268.6 soil concentration requirements are met by the activity set. If both requirements are met, the compliance indicator is equal to one; otherwise it is zero.

For example, suppose an activity set composed of activities \( x_1 \) and \( x_2 \), each with two possible outcomes, and suppose that performance results show that quantitative performance requirements are satisfied only if activity \( x_1 \) has outcome \( O_{A1} \) and activity \( x_2 \) has outcome \( O_{B2} \). The compliance indices for each of the four possible activity set outcomes would then be equal to zero for all but the outcome consisting of both \( O_{A1} \) and \( O_{B2} \), which would have a compliance indicator equal to one. Once the compliance index (CI) values have been determined for each activity set.
outcome, PDC is calculated by summing the probabilities for all activity set outcomes where CI is 1. The PDC for the activity set consisting of \( x_1 \) and \( x_2 \) would then be calculated as follows:

\[
PDC = 0 \times (P_{A1 \times P_{B1}}) + 0 \times (P_{A2 \times P_{B2}}) + 0 \times (P_{A3 \times P_{B3}}) + 1 \times (P_{A4 \times P_{B4}})
\]  

(1)

Thus, because all terms (outcomes) with a compliance indicator not equal to one would drop out of the PDC calculation, the PDC would equal \( P_{A2 \times P_{B2}} \).

Because of the multiple possible outcomes of SPM activities, activity sets can have anywhere between two and nearly 60,000 possible outcome combinations, each of which corresponds to a CCDF and a compliance indicator. Thus, the PDC for an activity set represents a logically straightforward but very computationally intense set of calculations.

**SPM-2 RESULTS**

The first iteration of SPM (SPM-1), the prototype of SPM, was completed September 1994. It served to develop the tools needed for the second iteration (SPM-2), which was completed in March 1995 for programmatic decision making. SPM-2 used technical positions derived from WIPP project technical staff, stakeholders, and oversight groups as a starting point for establishing a baseline. Technical teams also defined proposed activities and were elicited on the predicted outcomes of those activities. Trained elicitors external to the WIPP project worked with the technical teams in a formal, structured process to elicit the parameters and models to describe the activity outcomes and the probabilities of those outcomes. Activity cost and duration estimates completed the activity descriptions. DOE/CAO and the Westinghouse Waste Isolation Division provided information with regard to engineered alternatives, potential changes to waste acceptance criteria, and other programmatic guidance.

Potential outcomes were initially elicited for thirty-seven scientific investigations, eighteen engineered alternatives, and three waste acceptance criteria. These were screened to twenty-six discrete activities for the final SPM-2 analysis—twenty-one scientific investigations, three engineered alternatives, and two waste acceptance criteria. SPM-2 used existing WIPP performance assessment computer codes, with modifications required to model the baseline and activity sets, to calculate CCDFs of potential radionuclide releases. SPM-2 evaluated more than 600,000 possible activity sets. Activities that had no performance impact were removed from the decision matrix, reducing the number of activity sets in the decision matrix to roughly 46,700. Because each activity set had multiple outcomes, approximately 1.3 million CCDFs were needed to complete the SPM-2 analysis.

For activities in the decision matrix, SPM-2 showed that the probability of demonstrating compliance generally increased, as expected, with increasing activity set cost and duration. Figure 2 shows the overall structure of the results in terms of the probability of demonstrating compliance versus activity set cost. The large cluster of diamond-shaped points (each one corresponding to an activity set) on the far left includes only scientific activities. Activity sets near the top of Figure 2 all include one or more engineered alternative. Activity sets with a PDC of zero are not shown in Figure 2 for reasons of clarity, but are contained in the SPM-2 CD-ROM, an information management tool produced as part of the SPM project. Programmatic dependencies were also apparent from general trends in the data and are discussed in the statistical regression portion of this paper.
The SPM-2 baseline calculation predicted release of radionuclides in violation of 40 CFR 191.13(a) but compliance with respect to 40 CFR 268.6. About 40% of the SPM-2 activity sets also had a probability of demonstrating compliance of zero i.e., with no predicted value in supporting a demonstration of compliance. Of the remaining 60% of the SPM-2 activity sets, one half had a probability of demonstrating compliance equal to one. When conducted alone, no single activity—whether a scientific investigation, an engineered alternative, or a waste acceptance criterion—had a non-zero PDC.

Activity sets with a PDC of 1.0 included the scientific activity for colloids investigation and one of two engineered alternatives: either, (1) backfill and a pH buffer to control actinide solubility (EA1), or (2) an engineered backfill (such as clay) in combination with waste form modification (EA2). (Note that engineered alternatives and waste acceptance criteria were assumed to be optimally effective and were assigned a 100% probability of yielding the predicted performance. Subsequent sensitivity studies investigated the impact of this assumption on the final decision.) Two waste acceptance criteria (WAC) were analyzed by SPM-2. In the WAC-1 activity, steel drums used to store the waste were replaced with noncorrodible materials. WAC-1 added costs to the program and slightly reduced the probability of demonstrating compliance. WAC-2, elimination of all high-molecular weight organic compounds (such as soils) from the waste, had no discernible impact on the probability of demonstrating compliance.

The sensitivity of SPM-2 results to the probability of engineered alternative performance was straightforward to evaluate (see reference 3 for details). The DOE/CAO had a preliminary decision to make, either:

1. depend on a program consisting of engineered alternatives and minimal scientific investigations to provide a basis for the final compliance calculations; or

2. reserve engineered alternatives for possible use in providing assurance and depend on the scientific investigation to demonstrate compliance.

In May 1995, DOE/CAO chose the second option. Additional work has been conducted on engineered alternatives since the completion of SPM and the final balance between predicted per-
formance of the geologic system, engineered alternatives, and waste acceptance criteria will be described in the compliance certification application to the EPA.7

The final programmatic recommendations made to DOE/CAO in May 1995 considered the SPM-2 results along with existing information such as the 1992 WIPP PA Sensitivity Analysis and some sensitivity and uncertainty analyses. The sensitivity and uncertainty analyses did not alter the recommended series of activities. Other issues that were considered in using the SPM-2 results for decision-making were:

1. The technical baseline was for SPM use only. The final project technical baseline that will be used for preparing the WIPP compliance certification application will incorporate information from the activities completed subsequent to the SPM-2 effort.

2. The results were based on calculations using mean values, and were therefore valid for discriminating between activities intended to shift a mean value for a parameter but not for discriminating between activities intended to reduce uncertainty about a mean.

ANALYSIS OF RESULTS

The SPM-2 project generated roughly 46,700 unique activity sets. In order to understand the structure of the probabilities of demonstrating compliance among these activity sets, a statistical regression analysis was conducted. This analysis employed a logit regression methodology. A logit regression assumes that a probability, \( p \), (or other number bounded by zero and one) is related to several independent variables through equation 2:

\[
\log \left( \frac{p}{1-p} \right) = \sum b_i x_i \tag{2}
\]

In equation 2, \( x_i \) are indicator variables \((0,1)\) and \( b_i \) are regression coefficients to be estimated. Here, \( p \) is the probability of demonstrating compliance. Because the left side of the equation is unbounded at \( p=0 \) and \( p=1 \), the probability of demonstrating compliance values were decreased slightly towards 0.5 as shown in equation 3:

\[
p = (p - 0.5)(1 - \varepsilon) + 0.5, \text{ where } \varepsilon \text{ is a small number such as 0.01} \tag{3}
\]

An initial inspection of activity sets in the decision matrix revealed two very strong relations. First, if neither colloid activity (NS 8.1 nor NS 8.2) was included in an activity set, the probability of demonstrating compliance was zero. Second, if either NS 8.1 or NS 8.2 was in an activity set, the probability of demonstrating compliance was one as long as an engineered alternative (EA 1 or EA 2) was also in that activity set, and less than one otherwise. Both of these relations were always true and thus the first relation provided a sufficient condition for creating a probability of demonstrating compliance to equal zero. The second relation provided a condition that was both necessary and sufficient for probability of demonstrating compliance to equal one. These two relations logically limited the PDC of activity sets without EA 1 or EA 2 to \( 0 \leq \text{PDC} < 1 \).
In the absence of EA 1 and EA 2, what scientific programs should be undertaken to achieve a high probability of demonstrating compliance? This question was important because the predicted performance of EA 1 and EA 2 did not account for the possibility that an EA might prove less effective than assumed. Moreover, there were reasons to believe that the system-wide costs of EA 1 and EA 2 might ultimately be larger than initially estimated. For these reasons and to better understand the cost/benefit tradeoffs for the scientific program, a statistical analysis was limited to those activity sets where: both, (1) NS 8.1 or NS 8.2 was present, and (2) neither EA 1 nor EA 2 was present.

Using the logit model and excluding from the data set those activity sets without either NS 8.1 or NS 8.2 and excluding those having some combination of colloid activity with EA 1 or EA 2, regression coefficients were obtained. Based on regression results, activities are ordered from those with the greatest impact to those with the least impact, creating a series of activities such that as activities are added to the series, the PDC continues to increase but at a decreasing rate (see Figure 3). If the costs of the activities are similar, it is, in principle, possible to build a concave, monotonically increasing function that maximizes incremental PDC gained while minimizing incremental costs as more activities are added to the series. Two such activity series are shown in Figure 3 (the two curves on the left-most side of the graph) but they are not fully concave. The far-left curve is unconstrained by duration while the middle curve is constrained by a 19-month duration. The reason that these curves are not fully concave is that there are both thresholds and interactions (synergies) among some activities. The right-most curve in Figure 3 is a sub-optimal activity series that ultimately reaches nearly the same PDC as the pareto-optimal series but without the same ability to maximize incremental PDC per dollar at every point in the series.

For both the duration-constrained and unconstrained activity series in Figure 3, no improvement in the probability of demonstrating compliance was obtained by performing NS 8.1 by itself.
(Here NS 8.1 was chosen over NS 8.2 because of equal impact on the probability of demonstrating compliance and lower cost for NS 8.1.) However, for the duration-constrained series, the addition of NS 2 and NS 4 increased the probability of demonstrating compliance to 0.56. Addition of NS 7 further increased the PDC to 0.82. As Figure 3 shows, the addition of AST 1.2 did not increase the probability of demonstrating compliance. However, AST 1.2 was necessary to gain the PDC improvement provided by the combination of RM 1, SL 4, and DR 2. In fact, without first performing AST 1.2, the addition of RM 1, SL 4, and DR 2 produced a decrement in PDC. The same unexpected behavior occurred when one switched the order of the activities. One therefore concludes, that some interaction is taking place between AST 1.2 and the collection of three activities. Addition of any other activity to the series only brings minuscule improvements. A probability of demonstrating compliance of 0.96 is achieved from the duration-constrained pareto-optimal series.

The two left-most series are both considered pareto-optimal, that is, the series that cannot be bettered simultaneously in both cost and probability of demonstrating compliance. Faced with programmatic options limited to the scientific program—without engineered alternatives or waste acceptance modifications—both the duration-constrained and unconstrained activity series appear to be logical programmatic choices. However, the duration-constrained series, which eliminated two scientific activities (NS 3 and NS 5), resulted in virtually the same probability of demonstrating compliance as the unconstrained set and with lesser cost. The duration-constrained series was selected for implementation by the DOE/CAO.

DISCUSSION

The SPM-2 decision matrix yielded valuable information for identifying: (1) activity sets necessary to achieve a given PDC; (2) activity sets that give the maximum PDC; and (3) activities that have minimal impact on the PDC. Moreover, the use of quantitative analyses balanced with expert judgment was essential in developing insights about decision options in a highly nonlinear system. The SPM project required a significant commitment of human and computational resources but numerous improvements could be made to increase efficiency.

Information needed for the SPM analysis was acquired as expert elicitation from individuals directing the various activities and those proposing new activities. Adequate time for training participants in an expert elicitation process is essential.

Concerns were raised that the SPM baseline was excessively conservative and would not produce a useful basis for evaluation of activities. A management review was held to assure that the baseline was, in fact, appropriately balanced and integrated and that it was acceptable as the basis for performance calculations. Review is recommended of both the baseline and activities prior to performing calculations to assure appropriate consistency and integration of information elicited from many different individuals.

Side efforts (also known as side bar calculations, or side calculations) were also important in being able to keep the probabilistic calculations tractable and in preventing unnecessary conservatism in the baseline. Side efforts refer to confirmatory evaluations required to address certain technical positions embedded in the SPM-2 baseline. These confirmatory evaluations included scenario screening work, literature searches, bounding calculations, and some computer modeling.

There are computational limitations to probabilistic calculations underlying SPM. Suppose, for example, that m activities are to be considered and each activity has k potential outcomes. The number of endpoints to be evaluated is \( \sum_{i=1}^{m} k^m / [m! (m-i)! i!] \), which becomes very large,
very quickly. Clearly, not all combinations of activities can be evaluated. But this is where judgment and an understanding of disposal system performance can be used to create reasonable sets of activities for evaluation. Other computational schemes, such as sampling certain computational intensive parts of the performance assessment model, should be explored. In addition, multi-attribute utility analysis techniques (see reference 5) could be useful for up-front screening and focusing an initial large set of potential activities into a smaller set that require quantitative evaluation.

The usefulness of an SPM-like method depends upon the quality of the elicited information about activity outcomes, their probabilities, and the state of knowledge about system parameters and conceptual models. Retrospective analyses of SPM results can assess the degree to which actual outcomes were consistent with elicited predictions. Bayesian updating methods could be used if SPM were applied on an iterative basis.

SUMMARY

SPM identified viable combinations of programmatic options (activities) that, if implemented, were predicted to lead to a positive demonstration of compliance with long-term performance requirements. Moreover, analysis of the results also indicated that optimal programmatic pathways existed and that these activity series could provide useful insights into which activities to cut or add if budgets changed. Indeed, the analysis indicated that a positive demonstration of compliance with the long-term performance requirements could be anticipated within the DOE/CAO schedule.

SPM focused on work to achieve compliance with long-term disposal system performance requirements and helped eliminate concerns that activities would merely contribute to scientific knowledge. SPM utilized the existing performance assessment codes to calculate the expected results of various programmatic options. Use of quantitative performance assessment tools for prioritization was essential in gaining insights into the behavior of a highly coupled, nonlinear disposal system. SPM built upon the power of both performance assessment and decision analysis techniques, providing insights for decision making.

The general method could be applied to other complex issues in the environmental and waste management arena that need to clearly focus scientific and engineering activities on specific (and measurable) objectives within cost and schedule constraints. Because SPM combines decision analysis methods with quantitative analyses, it is conceptually applicable to any complex problem for which performance objectives, performance measures, and options to achieve the performance objectives can be defined. Projects that would likely benefit most are those with a complex set of technical issues and decision options that would benefit from planning based on calculated performance, rather than expert prediction alone. Projects with significant stakeholder involvement or with multiple participants might also benefit. Finally, probabilistic techniques used to treat uncertainties in the physical system could also be used to treat uncertainties in the cost or duration of programmatic alternatives.

ACKNOWLEDGMENTS

This paper is a summary of work done by many people who are specifically acknowledged in a number of reports now in preparation. The authors gratefully acknowledge their participation and important contributions to this project. We would like to specially acknowledge Walt Beyeler for his significant contributions to many aspects of SPM.
REFERENCES


This work was supported by the United States Department of Energy under Contract DE-AC04-9AL85000.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.