CASE STUDY ON THE USE OF IMPORTANCE SAMPLING IN THE PROBABILISTIC RISK ASSESSMENT OF UNDERGROUND DISPOSAL OF RADIOACTIVE WASTES

K. Johnson
CAP Scientific
20-26 Lambs Conduit Street
London
WC1
ENGLAND

ABSTRACT

In developing a methodology for assessing potential sites for the disposal of radioactive wastes, the UK Department of the Environment has conducted a series of trial assessment exercises. In order to produce converged estimates of radiological risk using the SYVAC A/C simulation system an efficient sampling procedure is required. Previous work has demonstrated that importance sampling can substantially increase sampling efficiency. This study used importance sampling to produce converged estimates of total risk for the first DoE trial assessment.

Four nuclide chains were analysed. In each case importance sampling was more efficient than random sampling. Importance sampling produced converged risk estimates with between 10 and 170 times fewer runs of the SYVAC A/C model.

This increase in sampling efficiency can reduce the total elapsed time required to obtain a converged estimate of risk from one nuclide chain by a factor of 20.

The results of this study suggest that the use of importance sampling could reduce the total elapsed time required to perform a risk assessment of a potential site by a factor of ten.

INTRODUCTION

Background

As part of its development of a methodology for assessing potential sites for the disposal of radioactive wastes, the UK Department of the Environment is conducting a series of trial assessments. These rehearsals are designed to develop and test a methodology for probabilistic risk assessment (pra) against a principal target of individual risk of 10^{-6} per annum during the post-closure period, as specified by the UK regulatory authorities (1).

The first trial assessment (DRY RUN 1) (2) was of a hypothetical disposal facility for intermediate level wastes, at a depth of 150 metres below the AERE Harwell site in Oxfordshire, UK. A single groundwater transport scenario was considered with parameter values and uncertainties that were assumed invariant with time. This was followed by DRY RUN 2 (3), which considered several possible groundwater transport scenarios, employed more sophisticated models and used a revised inventory.

Both dry runs had produced estimates of radiological risk using the UK SYVAC A/C framework of Monte Carlo simulation with simple random sampling. However, in neither case were the risk estimates shown to be converged.

Previous work (4) has demonstrated a method of importance sampling that can significantly increase the sampling efficiency of the SYVAC simulation. To demonstrate the use of importance sampling in the pra of a potential site a study was conducted to reanalyse DRY RUN 1 and one scenario from DRY RUN 2 to obtain converged estimates of risk.

Objectives

The main objectives of the study were:

a) To produce converged estimates of radiological risk.

b) To demonstrate how importance sampling could be used during a site assessment.

c) To compare the resources and elapsed time required to produce converged estimates of risk using importance sampling and random sampling.

IMPORTANCE SAMPLING

Background

Annual individual risk at time t, R_t, is defined from the expectation of dose (committed effective dose equivalent) at time t, $H_t(\underline{x})$, resulting from a defined parameter space $\underline{x}$ by:

$$R_t = \gamma \int_{\underline{x}} H_t(x).p(x)dx \qquad (1)$$

where γ is the ICRP risk factor.

The use of Monte Carlo simulation with random sampling enables risk to be estimated from N doses (H_{ti}, i=1,...,N) resulting from N simulations by:

$$\hat{R}_t = \frac{\gamma}{N} \sum_{i=1}^{N} H_{ti} \qquad (2)$$

With the SYVAC simulation the dose distribution $H_t(\underline{x})$ is generally highly skewed, with the majority of runs producing either zero or very low doses. This results in the risk estimate being dominated by a few high doses. Consequently, with random sampling very large numbers of simulations may be required to obtain a converged estimate of risk.

A method of importance sampling (IS) has been demonstrated which has greater sampling efficiency than random sampling. With IS, a sampling distribution $s(\underline{x})$ is used which is different from the subjective distribution of the input parameters $p(\underline{x})$. Risk is then estimated by:

$$\hat{R}_t = \frac{\gamma}{N} \sum_{i=1}^{N} H_{ti} \frac{p_i}{s_i} \qquad (3)$$

Where p_i and s_i are the density values of $p(\underline{x})$ and $s(\underline{x})$ corresponding to the i^{th} sample parameter set $\underline{x}_i$. Importance sampling becomes highly efficient if $s(\underline{x})$ can be defined so that:

$$s(\underline{x}) \approx H_t(\underline{x})p(\underline{x}) \qquad (4)$$

Sensitivity Analysis

Equation (4) defines the form of $s(\underline{x})$ that will maximise the efficiency of IS at time t. In practice, $s(\underline{x})$ is defined by its marginal distribution for each parameter from:

$$S(\underline{x}) = \prod_{j=1}^{k} s(x_j)$$

$$s(x_j) \approx H_t(x_j)\, p(x_j) \qquad (5)$$

This enables each parameter x_j to be sampled *independently* from $s(x_j)$ (j=1,...,k). But, equation (5) shows that only the independent effects of each parameter on dose $H_t(x_j)$ can be taken into account.

To define the IS distribution for each parameter $s(x_j)$ plots of cumulative risk against parameter value are examined from an initial random sampled case. Figure 1 is an example of these plots which are constructed by sorting the output of the case by the value of the parameter of interest. The contribution to risk from each run is calculated by:

$$r_{ti} = \frac{\gamma}{N} H_{ti} \qquad (6)$$

The plot shows the cumulative sum of these contributions.

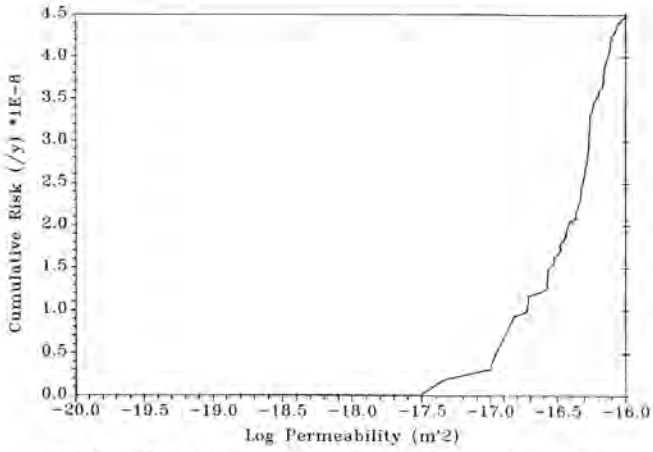

Fig 1. Typical plot of cumulative risk against parameter value from random sampling case

For a random sampling case the plot is an approximation to the cumulative distribution function (cdf) of $s(x_j)$. To enable $s(x_j)$ to be sampled in the subsequent IS case a Beta cdf is fitted to the curve in figure 1. The Beta family of distributions was found to offer sufficient flexibility to fit SYVAC cumulative risk curves.

Efficiency

In order to compare the relative efficiencies of IS and random sampling an efficiency statistic is defined as the ratio of the variances of the risk estimates from equal size cases of random and importance sampling. The objective of a case study is to produce a converged estimate of risk; that is one about which confidence limits can be calculated. The efficiency of IS gives an approximation of the number of random samples that would be required to obtain convergence. If an IS case of 2000 runs is converged with an efficiency of 10, then approximately 20,000 runs of random sampling would be required to obtain a converged risk estimate of equal accuracy. Previous work (5) has demonstrated efficiencies of over 100 for IS.

METHOD OF ANALYSIS

The application of IS is dependent on sensitivity analysis information (from the cumulative risk plots) from which the IS sampling distributions are defined. Figure 2 shows the main processes involved in performing an IS study. A preliminary random sampling case is run to provide sensitivity analysis graphs from which the IS distributions are defined. If this preliminary case does not generate sufficient non-zero doses a small IS case may be run with heuristically defined distributions.

The main IS case is then run using sampling distributions defined from the preliminary case. The results are analysed and tested for convergence. If the risk is not converged, the case may simply be extended or the sampling distributions refined using the extra information from the IS case.

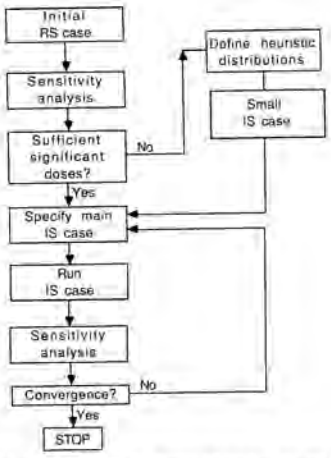

Fig 2. Flowchart of Importance sampling procedure

CASE STUDY

In DRY RUN 1, four nuclide chains made significant contributions to risk; the I-129 and Tc-99 nuclides and the Am-241 and U-234 chains. All four were reanalysed using importance sampling. Only one nuclide, Tc-99, from the mild artesian scenario, was analysed from DRY RUN 2.

In all cases, the waste is assumed to be placed in steel drums in a concrete repository in the Gault clay about 150 metres beneath the Harwell site. The clay is underlain by the Corallian aquifer. Transport of activity is modelled down through the clay and back to the biosphere along the aquifer.

Tables I and II show the sampled inputs to the SYVAC "A/C" model for each dry run.

Table I. Sampled input parameters for DRY RUN 1

PARAMETER		DIST'N	RANGE	
VAULT				
Sorption multiplier	(0)	UNIFM	0.2	5.0
Leach rate multiplier	(0)	CONST	1.0	
Buffer hydraulic cond.	m/s	LGUNIFM	1E-12	1E-8
Liner hydraulic cond.	m/s	LGUNIFM	1E-12	1E-8
Damage zone multiplier	(0)	CONST	2.0	
Buffer diffusivity	m^2/y	CONST	2.0E-2	
Liner diffusivity	m^2/y	CONST	2.0E-2	
Damage zone diffusivity	m^2/y	CONST	2.0E-2	
Buffer effective porosity	(0)	CONST	1.0E-1	
Time to buffer failure	yr	CONST	1.0E+10	
Velocity multiplier	(0)	CONST	4.0	
Barrier duration	yr	CONST	1.0E+3	
Liner porosity	(0)	CONST	1.0E-1	
1st GEOSPHERE LAYER				
Hydraulic gradient	(0)	UNIFM	0.1	1.0
Permeability	m^2	LGUNIFM	1E-20	1E-16
Porosity	(0)	UNIFM	0.01	0.3
Path length	m	UNIFM	50	10
Sorption multiplier	(0)	UNIFM	0.2	5.0
2nd GEOSPHERE LAYER				
Hydraulic gradient	(0)	UNIFM	0.0001	0.01
Permeability	m^2	LGUNIFM	1E-15	1E-11
Porosity	(0)	UNIFM	0.1	0.3
Path length	m	UNIFM	9600	19200
Dispersivity	m	UNIFM	1.0	100
Sorption multiplier	(0)	UNIFM	0.1	1.0

Table II. Sampled input parameters for DRY RUN 2

PARAMETER		DIST'N	RANGE	
VAULT				
Sorption multiplier	(0)	UNIFM	0.1	10
Leach rate multiplier	(0)	CONST	1.0	
Buffer hydraulic cond.	m/s	LGUNIFM	1E-11	1E-8
Liner hydraulic cond.	m/s	LGUNIFM	1E-11	1E-8
Damage zone multiplier	(0)	CONST	2.0	
Buffer diffusivity	m^2/y	CONST	2.0E-3	
Liner diffusivity	m^2/y	CONST	2.0E-3	
Damage zone diffusivity	m^2/y	CONST	2.0E-3	
Buffer effective porosity	(0)	CONST	1.0E-1	
Time to buffer failure	yr	CONST	1.0E+10	
Velocity multiplier	(0)	CONST	4.0	
Barrier duration	yr	CONST	1.0E+3	
Liner porosity	(0)	CONST	1.0E-1	
1st GEOSPHERE LAYER				
Hydraulic gradient	(0)	UNIFM	0.2	0.5
Permeability	m^2	LGUNIFM	1E-19	1E-18
Porosity	(0)	UNIFM	0.1	0.3
Path length	m	UNIFM	40	70
Dispersivity	m	UNIFM	0.01	1.0
Sorption multiplier	(0)	UNIFM	0.1	10
2nd GEOSPHERE LAYER				
Hydraulic gradient	(0)	UNIFM	0.001	0.0025
Permeability	m^2	LGUNIFM	1E-14	1E-12
Porosity	(0)	UNIFM	0.1	0.3
Path length	m	CONST	5000	
Dispersivity	m	UNIFM	30	60
Sorption multiplier	(0)	UNIFM	0.1	10
3rd GEOSPHERE LAYER				
Hydraulic gradient	(0)	CONST	0.01	
Permeability	m^2	LGUNIFM	1E-18	1E-16
Porosity	(0)	UNIFM	0.1	0.3
Path length	m	UNIFM	30	70
Dispersivity	m	UNIFM	0.01	1.0
Sorption multiplier	(0)	UNIFM	0.1	10

RESULTS

Risk at time

The use of SYVAC "A/C" with random sampling produces a risk estimate over the entire assessment time. However, the application of IS is based on the relationship between parameter values and dose (or risk) at a particular time after closure. This section gives results for risk at specific times. The next section describes how risk over the entire assessment time is obtained using IS.

Each of the five nuclide chains was analysed according to the procedure in Fig 2. From the results of the initial random sampling study risk over the entire assessment period was calculated and the time of maximum risk estimated. Optimal IS distributions were then defined using doses at the time of maximum risk.

Dry Run 2

For Tc-99 from DRY RUN 2, an initial random sampling case of 2000 runs was performed. Of these, 1930 runs generated zero doses, with only 4 doses contributing significantly to the total risk. The time of maximum risk was approximately 800,000 years after closure.

Figure 3 shows the cumulative risk graph for permeability in the second clay layer of the geosphere using doses from Tc-99 at 800,000 years from this case. The cumulative risk line is very jagged clearly showing the 4 main contributing doses. The dotted line is the cdf of the sampled distribution $p(x_j)$; in this case log uniform. If the parameter had little effect on risk, the cumulative risk line would approximate this cdf ie

$$H(x_j)\ P(x_j) \approx p(x_j) \qquad (7)$$

This parameter has a significant effect on risk. The solid line is a Beta cdf fitted to the cumulative risk line. Figure 4 shows the pdf of this Beta distribution, which was used as the sampling distribution in the subsequent IS study.

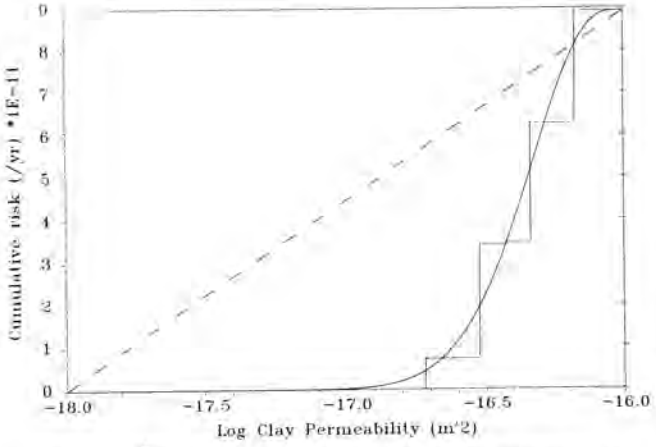

Fig 3. Cumulative risk graph for clay permeability

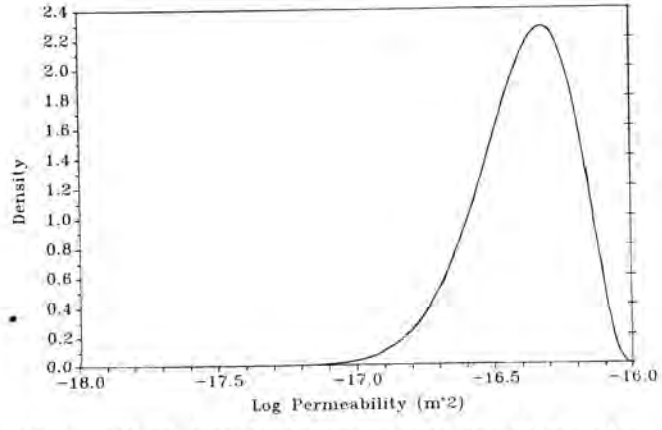

Fig 4. pdf of Beta distribution used for IS for clay permeability

This procedure was followed for each SYVAC input parameter to define IS sampling distributions. As the initial random sampling case generated so few significant doses, a small IS case was run to confirm the sampling distributions. A 500 run case was performed using the distributions defined initially. This generated many more significant doses. Cumulative risk graphs were plotted for this case and new distributions defined. These were then used for a 2000 run IS case. This case was successful in concentrating the sampling in the important (ie high risk) region. Of the 2000 runs, 500 made a significant contribution to total risk. Figure 5 shows the cumulative risk plot for permeability in the clay layer of the geosphere from this case. The graph clearly shows the large number of significant doses. The shape of the cumulative curve shows a strong resemblence to Fig. 3. This shows that although the initial random sampling case had few significant doses and a very poor risk estimate the sensitivity information was still very useful. Figure 5 also shows the cdf of the Beta distribution used to sample the parameter in the 2000 run case. This is very similar to the cumulative risk line and this indicates that the sampling was close to optimal.

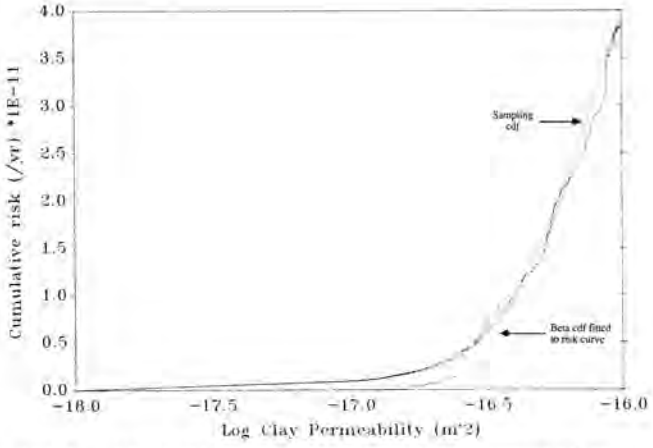

Fig 5. Cumulative risk plot for clay permeability for 2000 runs IS case

To confirm the convergence of this case, the sample was split into ten equal size sub-samples. A risk estimate was calculated for each of these subsamples. These sub-sample risks were plotted on normal probability paper and are shown in Fig. 6. The points on the graph approximate a straight line which indicates the normality of the risk estimates and confirms the convergence of the case.

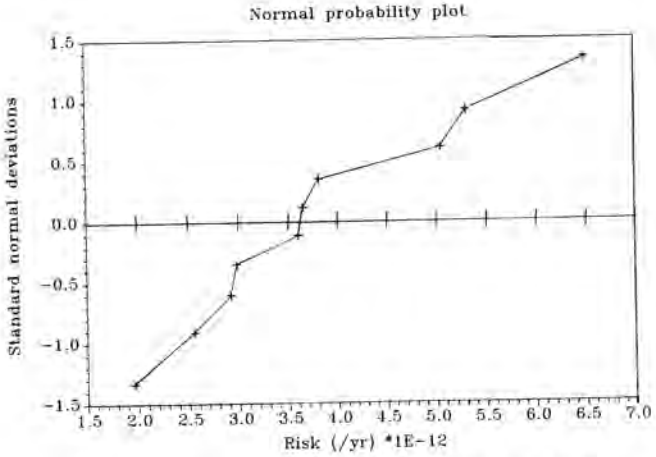

Fig 6. Normal probability plot for risk from Tc-99 from DRY RUN 2

For comparison a 50,000 run case using random sampling was performed for this nuclide. Figure 7 shows how the risk estimates from random and importance sampling converge with increasing sample size. The IS case was converged by 2000 runs and this was confirmed by extending the case to 4000 runs. The graph shows how slowly random sampling converges, with confidence limits only being valid after 50,000 runs.

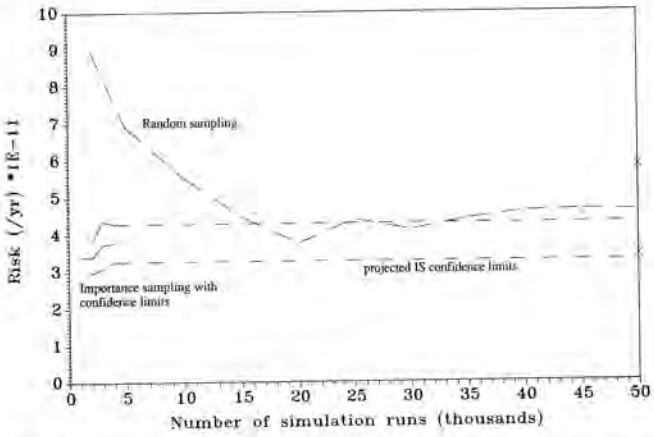

Fig 7. Variation of risk with sample size for Tc-99 from DRY RUN 2; comparison of random and importance sampling

Dry Run 1

A similar procedure was followed for each of the four nuclides for DRY RUN 1. In each case the final IS results generated many more significant doses than random sampling; typically 300-500 from a 2000 run case. Table III gives a summary of results for each of the five nuclides analysed. For the single member chains in DRY RUN 1 (I-129, Tc-99) the efficiency of IS is estimated to be 10. So, for I-129, IS had converged in 1000 runs implying that 10,000 runs of random sampling would be required. However for the U-234 chain the efficiency of IS is estimated at 177. The IS case reached convergence after 1000 runs implying that approximately 177,000 runs of random sampling would be required to obtain a converged risk estimate of equal accuracy. With the current DEC VAX based SYVAC computing facilities that would require 62 days of continuous computer time. In this case IS is essential if a converged risk estimate is to be produced in a realistic timescale.

Table III. Summary of converged risk results for each nuclide analysed

DRY RUN 1

Case		Time of analysis (years)	No. runs	Sig. risks	Risk (/year)	Efficiency
I-129	RS	200,000	2000	60	6.00E-8	
I-129	IS	200,000	1000	300	7.76E-8	10
Tc-99	RS	800,000	2000	60	1.91E-8	equiv. to I-129
U-234	RS	500,000	2000	7	4.15E-8	
U-234	IS	500,000	1000	617	6.51E-8	177
Am-241	RS	500,000	2000	7	4.28E-8	
Am-241	IS	500,000	1000	280	4.47E-8	80

DRY RUN 2

Case		Time of analysis (years)	No. runs	Sig. risks	Risk (/year)	Efficiency
Tc-99	RS	800,000	2000	4	8.95E-11	
Tc-99	IS	800,000	2000	475	3.84E-11	38

Risk over time

A full analysis of a case requires a risk profile to be calculated over the whole assessment time period. Random sampling generates a risk estimate at each time point automatically although as shown in the previous section very large samples may be required to obtain converged risk estimates. With IS the sampling distributions are only optimal at one time, the time of the original sensitivity analysis. However, these distributions can be used to calculate risk at other times. Although they will not be optimal at these times they may still result in a substantially improved efficiency over random sampling.

This technique was successfully used to produce converged risk estimates for DRY RUN 1 over the assessment time period. Figure 8 shows risk over time for the four nuclides using IS. Figure 9 compares the total risk estimate from IS with the original DRY RUN 1 risk estimate. The original risk profile correctly estimated the peak risk from the single number chains at about 200,000 years, but underestimated the main risk peak at 500,000 years.

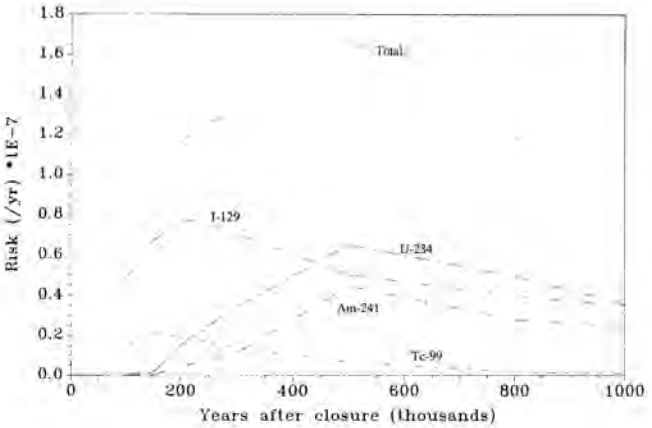

Fig 8. Final converged risk profiles for major contributing nuclides for DRY RUN 1

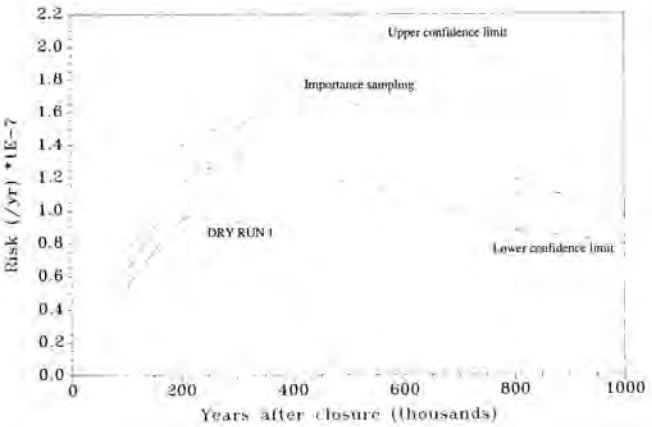

Fig 9. Total risk profile for DRY RUN 1: comparison of original estimate with converged IS estimate

Resources

Table III shows IS producing increased sampling efficiencies of between 10 and 177 compared with random sampling. However, this is potentially misleading as an IS study requires additional analysis effort to perform sensitivity analysis and select IS sampling distributions. A more realistic comparison is of the total elapsed time required to produce converged risk estimates. As an example, the IS study of risk from U-234 from DRY RUN 1 required 7 elapsed days. As explained a random sampling study would require approximately 62 days. Thus, although the sampling efficiency of IS is 177, the elapsed time efficiency is about 9 (ie 62/7). Experience has shown that IS produces elapsed time efficiencies of between 1 and 20.

A complete risk assessment of a potential size will involve analysing several nuclides, for several scenarios. Many factors will affect the overall resource requirements. However, experience to date suggests that the use of importance sampling could reduce the elapsed time required for a typical risk assessment by a factor of ten.

CONCLUSIONS

Although random sampling is attractive for use in Monte Carlo simulations due to its simplicity it can be very inefficient. This is particularly true if the dose function $H(\underline{x})$ is highly non-linear.

The use of importance sampling can greatly increase the efficiency of Monte Carlo simulation, reducing the sample size required to obtain converged results by factors of over 100.

The use of importance sampling in the risk assessment of a potential disposal site is essential if the work is to be completed in a realistic timescale. This is particularly true if multi-member nuclide chains need to be analysed.

ACKNOWLEDGEMENTS

The work presented in this paper was funded by the UK Department of the Environment under contract number PECD 7/9/226-70/84.

The author would like to acknowledge Dr G Dalyrmple and Mr R Lucas of CAP Scientific for their contributions to this work and Dr B Thompson of UK Department of the Environment for helpful comments on earlier drafts of this paper.

Although the work described in this paper may be used in the formulation of UK Government policy it does not at this stage necessarily represent Government policy.

REFERENCES

1. Department of the Environment, "Disposal Facilities on Land for Low and Intermediate Level Radioactive Wastes: Principles for the Protection of the Human Environment" (1984).

2. B. G. J. THOMPSON and T. W. BROYD, "DRY RUN 1: An Initial Examination of a Procedure for Post-Closure Radiological Risk Assessment of an Underground Disposal Facility for Radioactive Waste", DoE Disposal Assessments Report TR-DoE-4, Edition 1, February 1986.

3. Z. A. GRALEWSKI et al, "Development of a Methodology for Post-Closure Radiological Risk Analysis of Underground Waste Repositories. Illustrative Assessment of the Harwell Site", (DRY RUN 2), DoE Report DoE/RW/1986, July 1986.

4. G. J. DALRYMPLE et al, "Testing a Method of Importance Sampling for use with SYVAC", DoE Report TR-STH-15, October 1985.

5. K. JOHNSON and R. LUCAS, "The use of Importance Sampling in a Trial Assessment to Obtain Converged Estimates of Radiological Risk", DoE Report TR-STH-23, November 1986.