THE ACQUISITION OF DATA FOR USE IN THE PROBABILISTIC RISK ASSESSMENT OF UNDERGROUND DISPOSAL OF RADIOACTIVE WASTE

G. J. Dalrymple

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 assessments. For the trial assessments there are few measured data available and the input probability distributions to the SYVAC A/C framework of Monte Carlo simulations have to be assessed from expert opinion. A feasibility study was conducted on the use of a formal method for encoding probability distributions. Three parameters were successfully encoded as part of the study and helpful conclusions drawn. At the stage of a more detailed site assessment measured data will be available and these need to be incorporated into the subjective distributions obtained from experts. Two illustrative case studies were conducted in which a Bayesian approach was used to incorporate the measured data. The first case study illustrated how the measured data can substantially reduce the uncertainty about a parameters value whereas the second case study contained a large amount of measurement uncertainty.

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 annual 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. The DRY RUN 1 rehearsal used the UK SYVAC A/C framework of Monte Carlo simulations to obtain estimates of radiological risk. The uncertainty in the value of the parameters used as input to the SYVAC A/C models is represented in the form of probability distributions. For the preliminary site assessments and the trial assessment in particular there are few measured data available and therefore the parameter values have a large degree of uncertainty. Under these circumstances probability distributions have to be assessed by relying on the judgement of experts. In the DRY RUN 1 rehearsal formal methods were not used to elicit the subjective input probability distributions.

A feasibility study was conducted on the use of a formal method to acquire input probability distributions for SYVAC A/C.

At the later stages of a site investigation measured data will become available. The data will reduce the uncertainty in the parameter values and the subjective probability distributions will need to be updated with the new information available. The feasibility of using a Bayesian approach for incorporating measured data into the subjective probability distributions was conducted through two illustrative case studies.

Objectives

The main objectives of the two studies were:

- to demonstrate the feasibility of using a formal method to elicit subjective probability distributions from expert opinion;
- to demonstrate the feasibility of using a Bayesian approach for incorporating

measured data into subjective probability distributions.

METHOD

Data Acquisition - background

Several formal methods have been proposed for assessing subjective probability distributions. An earlier review of the quality of assessed distributions (3) found that, in general, people tend to be too sure of themselves in assessing these distributions. Practical procedures have evolved and these are described by Stael von Holstein and Matheson (4). The procedures for eliciting subjective probability distributions from expert opinion described in Ref 4 was used in this study and is referred to as the SRI approach having been developed by SRI International. A further refinement to the technique which was used in this study was to elicit the subjective distributions from a group of experts rather than individuals. This is because when several people get together they share their expertise thus providing some correction to biases that are inherent in the limited experience that any individual necessarily holds.

Data Acquisition - method

The data acquisition method contains five stages:

- motivating
- structuring
- conditioning
- encoding
- verifying

Each of these stages is described in Ref 4 and summarised here. The motivating stage introduces the expert participants to the encoding task. In the structuring stage the uncertain quantity is clearly defined and an exploration is conducted of the way the participants think about the quantity. The conditioning stage is designed to draw out the participants' knowledge of the uncertain quantity so that it can be shared by all participants. The participants are encouraged to think over a broad range of possible events and scenarios. The last two stages encode the probability distribution and verify them for consistency against other information provided by the experts.

Bayesian updating

Bayes theorem results from the laws of probability and is expressed for a single parameter θ s:

$$p\ (\theta/D)\ \alpha\ p(\theta)\ p(D/\theta) \qquad (1)$$

where

$p(\theta)$ is the initial subjective distribution for θ generally referred to as the prior distribution.

D is a set of measured data.

$p(D/\theta)$ is the conditional probability of D occurring for any given value of θ. This is generally referred to as the likelihood function.

$P(\theta/D)$ is the distribution of θ having collected the data D; this is known as the posterior distribution.

The prior distribution is obtained from the data acquisition exercise using the method described above. The measured data, D, will be provided from a site measurement program. The main difficulty in applying the Bayesian approach is deriving the likelihood function and combining it with the prior. This is complicated by correlated parameters and when the parameter used in the models is not directly measurable. The feasibility study examined two illustrative cases of deriving the likelihood function and determining the posterior distribution.

CASE STUDIES

Data Acquisition

The feasibility of using the SRI technique was studied for three parameters, one from each of the sub-models used in SYVAC A/C ie the vault, geosphere and biosphere sub-models. The parameters selected and their definitions, agreed in the data acquisition exercises, are given in Table I.

TABLE I

Definitions of Uncertain Quantities Agreed in the Data Acquisition Exercise

Sub-model	Parameter definition
Vault	The hydraulic conductivity of cement based backfill material after curing for 50 years for a repository for the disposal of intermediate level radioactives wastes.
Geosphere	The average hydraulic conductivity for horizontal flow along the Corallian aquifer of the South Oxfordshire region.
Biosphere	The ratio of the average dry weight concentration of I129 in the upper parts of plants to the average dry weight concentration of I129 in soil.

Information on each of the uncertain quantities was provided to the experts prior to the meeting. Either two or three experts were involved in each exercise.

Bayesian Updating

The feasibility of using a Bayesian approach to incorporate measured data into subjective probability distributions was illustrated through two case studies of different complexity. The simpler case study concerned the effective hydraulic conductivity of the cement-based backfill material of the vault. This parameter is a spatial average of the conductivity of the entire backfill material. In this case the prior distribution was obtained from experts and simple assumptions were made about the nature of the likelihood function.

The second case study concerned the local measured value of the hydraulic conductivity of the Corallian aquifer which underlies the Harwell site in Oxfordshire. In this case both the prior distribution and the likelihood function were obtained from experts.

RESULTS

Data Acquisition

As explained earlier the data acquisition exercises pass through a number of stages before encoding begins. The definitions of the parameters were revised during the course of the exercises and the agreed definitions have been given in Table I. The experts' judgements of the factors that influence the parameters value and the assumptions that are made in providing information about the parameters value are listed in Table II. In the encoding stage experts were asked to estimate the probability of parameter values being less than a given value. The experts judgements for the hydraulic conductivity of cement-based backfill material are shown in Fig. 1 together with a beta cumulative distribution function (cdf) fitted to the estimates. The beta probability density function derived from this cdf is shown in Fig. 2. The resulting pdfs for the other two parameters are given in Figs. 3 and 4.

The parameter distributions used in the DRY RUN 1 rehearsal are also given in Figs. 2-4 for comparison. It should be stressed that this was a feasibility study and that in a full data acquisition exercise more time would have been spent verifying the resultant pdfs. For example, in Fig. 2, the steep rise in probability density at 10^{-14} m/s would have been verified with the experts. The results should therefore be regarded as preliminary.

Bayesian Updating

In the first case study, the prior distribution of the hydraulic conductivity of the cement based backfill material was obtained from experts in the data acquisition exercise described above and shown in Fig. 2. As no measured data was available, an expert provided twelve typical values of hydraulic conductivity for this type of material which are given in Table III.

Table III. Typical measured values of hydraulic conductivity of cement-based backfill material supplied by an expert

Hydraulic conductivity (m/s)	
1.0×10^{-11}	5.0×10^{-13}
5.0×10^{-12}	5.0×10^{-13}
1.0×10^{-12}	1.0×10^{-13}
1.0×10^{-12}	1.0×10^{-13}
1.0×10^{-12}	5.0×10^{-14}
5.0×10^{-13}	1.0×10^{-14}

No explicit information on the likelihood function was available and therefore it was assumed that the global average value of this parameter is the geometric mean of locally measured values. It was also assumed that both the measurement and spatial uncertainty are distributed log normally. Hence

If x_i ($i = 1, 2 \ldots 12$) are the measured values and

$$y_i = \log_{10} x_i$$

the measurement variability is modelled by:

$$y_i \sim N(\theta_i, \sigma^2) \qquad (2)$$

where θ_i is the true local value and σ^2 is the measurement error variance. Similarly the spatial variability is modelled by:

$$\theta_i \sim N(\theta, r^2) \qquad (3)$$

where θ is the true global value and r^2 is a measure of the spatial variability. Equations (2) and (3) can be combined;

$$y_i \sim N(\theta, \pi^2) \qquad (4)$$

where $\pi^2 = \sigma^2 + r^2$

As no information was available on the likelihood variance, π^2 becomes a second unknown parameter. The posterior distribution for θ, $P(\theta/y, \ldots y_{12})$ is then defined as:

$$P(\theta/y_1, \ldots y_{12}) = \int P(\theta, \pi/y_1, \ldots y_{12}) d\pi \qquad (5)$$

where

$$P(\theta, \pi/y_1, \ldots y_{12}) = \prod_{i=1}^{12} P(y_i/\theta, \pi) P(\theta) P(\pi)$$

$P(\pi)$ is as yet unspecified and a convenient specification, representing relative ignorance about the likelihood variance, π^2, is to take $P(\log \pi)$ = constant. The resulting posterior distribution was calculated using the methodology set out in ref. 5 and is shown in Fig. 5.

The posterior density distribution shows a substantial decrease in variance compared with the prior. This reflects the fact that the posterior is dominated by the measured data and that the beta prior was very weak (ie uninformative). The sensitivity of the results to the form of the prior distribution, the number of measured data points and the form of the likelihood function were investigated.

The results are insensitive to chosing a uniform prior distribution or defining the likelihood function with a t-distribution instead of a normal. The analysis was re-run using only half the data (ie 6 data points). The resulting posterior had a significantly larger variance than with the full data set. This confirms the influence of the data on the posterior.

In the second case study measurement uncertainty was treated in detail and the Corallian aquifer was assumed to be a homogenous medium (ie there is no spatial variability). The prior distribution was again obtained from expert opinion using the approach described above and is shown in Fig. 6.

The likelihood function was also obtained from expert opinion, such that, given any set of data a posterior distribution could be determined.

Firstly the uncertainty in measuring hydraulic conductivities in strata using pulse tests was assessed. An expert was asked to assume that she knew that the true value of hydraulic conductivity of the aquifer was 10^{-7} metres per second and to assess the probability that test results would be less than selected values. The exercise was repeated for true values of hydraulic conductivity of 10^{-5} and 10^{-9} metres per second. The resultant pdfs are shown in Fig. 7.

Table II. The assumptions and factors that influence the parameters which were the subject of the data acquisition exercise

Exercise	Uncertain Quantity	Factors which influence the parameter	Assumption
1.	Hydraulic conductivity of cement based backfill material	- composition of materials making up the concrete - compaction - curing - cracking - thermal stresses from concrete and waste - quality of workmanship	- microcracking of concrete but no major structural cracking - concrete made on site and made to minimise the hydraulic conductivity - no major inflow of water during construction and curing
2.	Hydraulic conductivity of the Corallian aquifer	- measurement error - fissures in the strata - bias in the emplacement of drill holes - the stress and history of stress - the quantity of water pumped out and the water input to the aquifer	- a saturated aquifer - isotropic flow horizontally - pore water has the viscosity of pure water - homogenous layer but including fissures and faults - spatial average thickness of the layer - spatial average velocity of flow through the layer
3.	The ratio of dry weight concentration of I129 in the upper part of plants to the dry weight concentration in soil	- the time of year at which measurements are made - the stable iodine concentration in soil - the type of pasture assumed - the depth and type of soil	- the concentration includes root uptake and uptake from soil via other mechanisms - current climatic conditions and agricultural practice - brown earths and glayed brown earths - soil is homogenous over the rooting zone - a homogenous mix of iodine - the main land use is pasture and a 3 km² area is considered - the grass is cut twice a year for hay and the iodine concentration is measured at the time of cutting

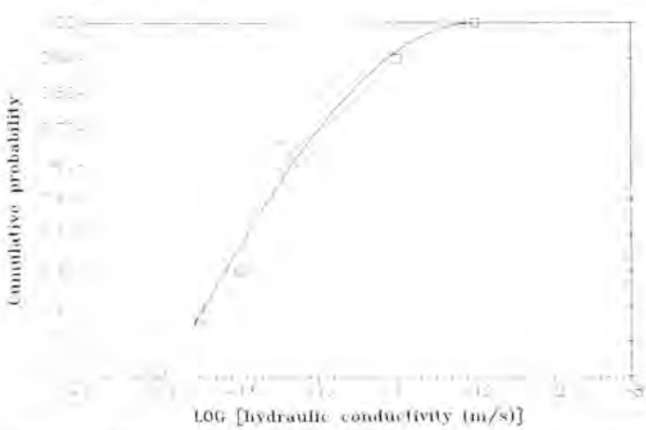

Fig. 1. Experts cumulative probability judgements and the fitted beta cdf of the hydraulic conductivity of cement-based backfill.

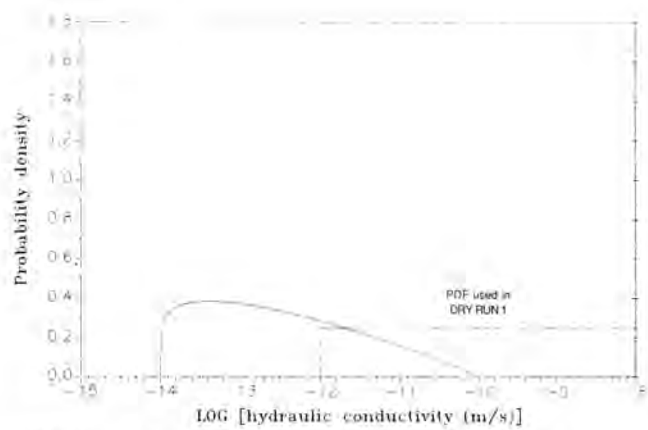

Fig. 2. Probability density function for the hydraulic conductivity of cement-based backfill.

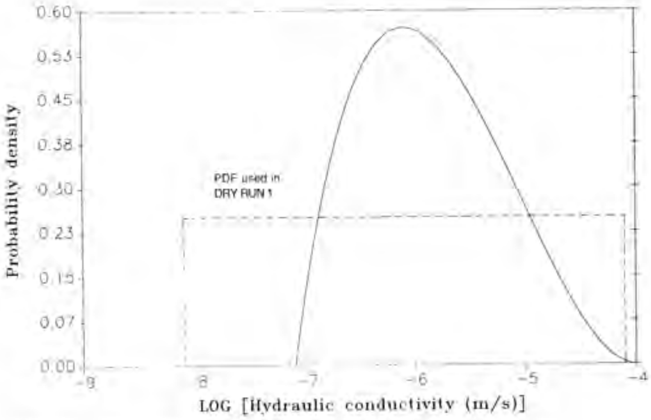

Fig. 3. Probability density function for the average hydraulic conductivity of the Corallian aquifer underlying the Harwell site.

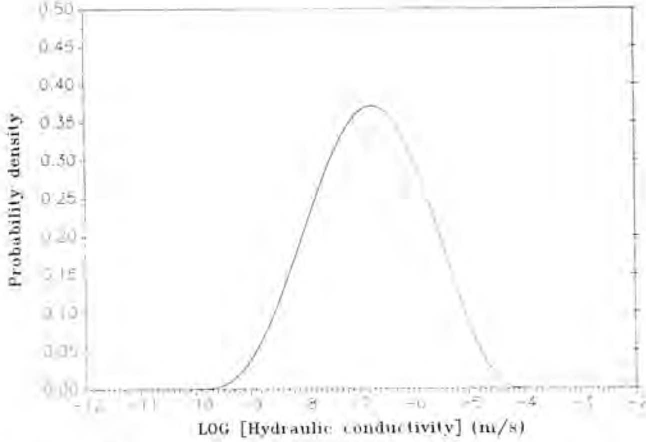

Fig. 6. Probability density function of the local hydraulic conductivity of the Corallian aquifer assuming a homogenous medium.

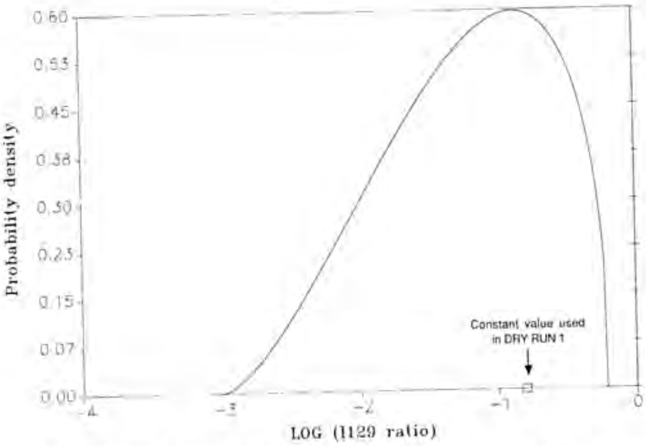

Fig. 4. Probability density function for the ratio of I129 in the upper parts of plants to I129 in soil.

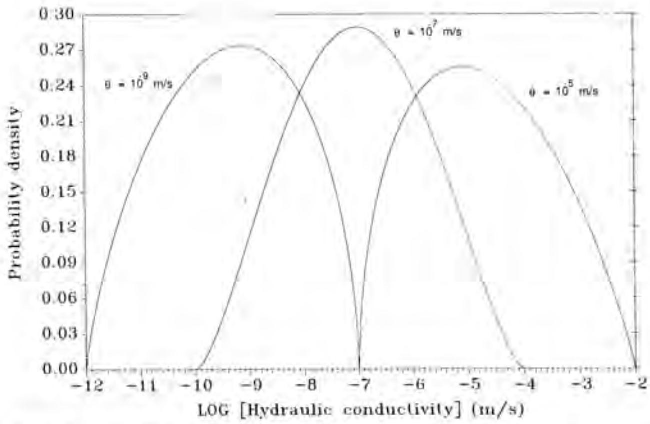

Fig. 7. Probability density functions representing the measurement uncertainty of the hydraulic conductivity of the Corallian aquifer given true values of conductivity of 10^{-5}, 10^{-7} and 10^{-9} m/s.

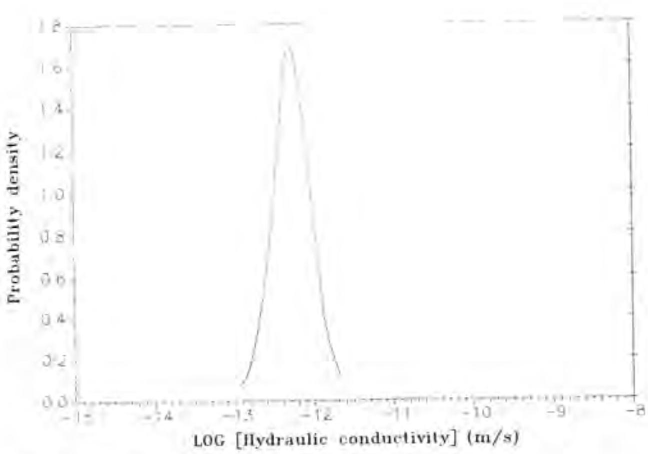

Fig. 5. Posterior density distribution for the hydraulic conductivity of cement-based backfill material.

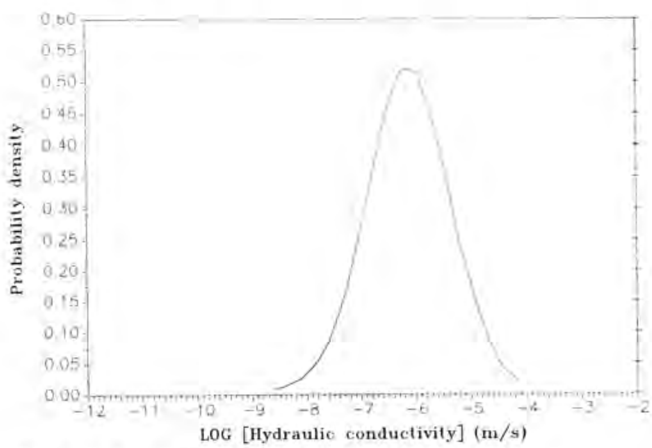

Fig. 8. Posterior density distribution for the hydraulic conductivity of the Corallian aquifer assuming a homogenous medium.

To perform the updating procedure a general likelihood function is required defining the measurement error associated with any value of true conductivity. The parameters which describe the beta distributions in Fig. 7 are roughly linearly related to the true value of hydraulic conductivity, θ, by:

$$(B-A)\frac{p}{p+q} + A = -5.6 + 0.2\theta \qquad (6)$$

where A and B are the lower and upper limits of the parameter range

p and q are the parameters of the beta distribution which describe its shape

This relationship enables a general Beta likelihood function to be defined, giving the likelihood of <u>any</u> measured value for <u>any</u> true value of θ. A data set was created from the measurement uncertainty distribution provided by the expert for a true value of hydraulic conductivity of 10^{-7} metres per second. The data set is given in Table IV.

Table IV. Measured value of the hydraulic conductivity of the Corallian aquifer for a true value of 10^{-7} m/s

Measured value (m/s)	
6.52×10^{-8}	8.09×10^{-9}
1.36×10^{-6}	4.81×10^{-8}
9.20×10^{-8}	7.83×10^{-7}
6.49×10^{-9}	1.90×10^{-6}
1.56×10^{-6}	7.24×10^{-6}

Using this simulated data and the general Beta likelihood function the posterior distribution was calculated using the methodology set out in ref. (5) and is shown in Fig. 8.

The posterior distribution shows a reduced uncertainty compared with the prior distribution. However because of the high level of measurement uncertainty the data does not have as marked an influence on the posterior distribution as in the first case.

DISCUSSION AND CONCLUSIONS

Data Acquisition

The main conclusion of the study is that it appears feasible to obtain probabilistic information from groups of experts. A computer program to fit beta cdfs to the elicited data and display the resultant pdf is particularly useful in the data acquisition exercise. This allows the experts to immediately examine their judgements and assists in the verification stage. It was clear from the feasibility study that four hours should be allocated to an encoding session to allow sufficient time for familiarisation and verification. We also conclude that the participants should be selected carefully to cover a wide mix of expertise. At least one participant should hold practical experience of relevance to the quantity and the site conditions.

The SRI method requires modification so that it will apply more effectively in a group setting. A more efficient mechanism for reaching a consensus among experts needs investigating.

Bayesian Updating

The Bayesian updating study considered two cases. The <u>first case</u> demonstrated the mechanism of incorporating hypothetical measured data into subjective distributions by making reasonable assumptions about the measurement and spatial uncertainties. A posterior distribution was successfully calculated and showed how, in this case, the measured data substantially reduced the uncertainty about the parameters value. The sensitivity analysis examined the effects of changing the distributions of the components of the equation describing Bayes theorem ie the prior distribution, the likelihood function and the data. The analysis demonstrated, for this case, the dominant effect that the data had on the form of the posterior distribution.

The <u>second case</u> extended the analysis to consider measurement uncertainty in more detail assuming a homogenous Corallian aquifer. This case demonstrated that it is feasible to determine a likelihood function by consulting experts. The prior distribution and the likelihood function have comparable levels of uncertainty and therefore comparable influences on the posterior distribution. This contrasts with the first case where the data was clearly the dominant influence on the posterior.

The two case studies have shown clearly how the posterior distribution is able to combine the information contained in the prior distribution and the data. The posterior distribution is thus more informative than either the prior distribution or the data alone.

Further work

The analysis described above assumed a homogenous aquifer. The aquifer is known to consist of three distinct rock types ie dense, sandstone, mudstone and unconsolidated sandstone and limestone laid down non-uniformly. The assumption of homogeneity cannot therefore be justified even if statistical averaging is assumed. The second case study has been extended to include a more representative model of the aquifer whereby three distinct parallel rock types are assumed together with a variation of hydraulic conductivity with the depth of the aquifer. The aquifer model assumes a parallel resistive analogy for the layers. This analysis is in progress.

ACKNOWLEDGEMENTS

The author would like to thank the UK Department of the Environment for sponsoring the work described in this paper and the author thanks the DoE for permission to publish this paper. The author would also like to thank the DoE project manager, Dr B.G.J. Thompson for his helpful comments, Dr L. Phillips, Decision Analysis Unit, London School of Economics, Professor A.F.M. Smith, Nottingham University and Mr K.B. Johnson, CAP Scientific for their contributions to the work.

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, HMSO (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. G.J. DALRYMPLE et al, Final Report on the Acquisition of Data for use in the Probabilistic Risk Assessment of Underground Disposal of Radioactive Wastes, CAP Scientific, DoE report DOE/RW/86.071, January 1986.

4. C-A.S. STAEL VON HOLSTEIN and J.E. MATHESON. A Manual for Encoding Probability Distributions, SRI International, August 1979.

5. J.C. NAYLOR and A.F.M. SMITH, Applications of a Method for the Efficient Computation of Posterior Distributions, Applied Statistics, 31, 214-225, 1982.