rec'd with letter dtd
7/7/94
9408150350

- **DRAFT** -
Printed 14/6/94

The Probabilistic System Assessment Group – History and Achievements 1985 – 1994

NUCLEAR ENERGY AGENCY
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

426.1

Preface

The NEA Radioactive Waste Management Committee (RWMC), established in 1975, is an international committee of senior governmental experts familiar with the scientific, policy and regulatory issues involved in radioactive waste management. A primary objective of the RWMC is to improve the general level of understanding of waste management issues and strategies, particularly with regard to waste disposal, and to disseminate relevant information. Current NEA programmes under the RWMC focus on methodologies for the long-term safety assessment of waste disposal, and on site evaluation and design of experiments for radioactive waste disposal.

The Probabilistic System Assessment Group (PSAG) was established by the RWMC in January 1985 (originally as PSAC, the Probabilistic System Assessment Code User Group) to help coordinate the development of probabilistic safety assessment computer codes in OECD member countries. PSAG has met approximately twice per year between its founding and its final meeting in June 1994. It has discussed topical issues, exchanged information and conducted code intercomparison exercises.

The PSAG has been assisted by the NEA Data Bank, which undertakes the collection, validation, and dissemination of computer programs and scientific data within the NEA's field of interest. Its support for radioactive waste management activities includes code exchange and analysis of code intercomparisons.

Acknowledgement

This report was prepared by J E Sinclair, Chairman of PSAG from 1991 to 1994. Suggestions for the scope and content of the report were received from a number of past and present PSAG members.

Contents

1. Introduction

The Probabilistic System Assessment Code User Group (PSAC) was set up by the NEA in 1985. Later, the name was changed to simply the Probabilistic System Assessment Group (PSAG). The founding of the Group followed the coming together of representatives of several national groups who were using the Canadian computer program SYVAC (Systems Variability Assessment Code). It became clear that there was interest in many of the OECD member countries in developing codes for assessment of radioactive waste disposal projects, and that this code development would not necessarily be based, in many cases, directly on SYVAC. In setting up the PSAC Group under NEA sponsorship, terms of reference were proposed, and agreed by the Radioactive Waste management Committee of the NEA, that included exchange of codes, information and experience, conducting mutual peer reviews, contributing to code justification by code comparison or other exercises, and discussing technical issues identified as being of concern for the further development of the PSA approach.

Since its founding, the PSAC/PSAG met approximately twice per year, until 1993, with typically 15 to 30 people attending each meeting, representing 20 or more groups from at least 10 counties. The final meeting of the Group took place in June 1994. The Group has been assisted by a secretariat provided by the NEA, and the NEA Data Bank gave assistance in the exchange of computer codes and analysis of some of the exercises.

In keeping with the aims of the Group, exchange of information and experience has been taken place through presentation at the meetings of progress reports by members, and reports on related NEA-sponsored and other international activities of relevance. Detailed exploration of important technical issues has been carried out by holding Topical Sessions at most PSAG meetings. These sessions have been contributed to by PSAG members and by invited speakers from outside the Group. A major part of the Group's time, however, has been devoted to conducting a series of code comparison exercises, known as PSACOIN Exercises. These were designed to fulfil the Group's purpose of contributing to justification of the computer codes that the participating national groups were preparing for use in safety assessments in their own national programmes.

Reports on the PSACOIN exercises have been published by the OECD. Records of the PSAG meetings, and collections of papers presented at Topical Sessions, have been distributed to Group members, and to other related NEA groups, especially the Performance Assessment Advisory Group (PAAG), and the RWMC.

In 1990 a booklet was published by the OECD explaining the background to the activities of the PSAG, and outlining the progress in its work up to that date. The present report has been prepared to provide a record and summary of the achievements of the PSAG over its nine years of operation. In Chapter 2, a brief technical account of the concepts and methodology of probabilistic system assessment (PSA) is presented. A summary of the development and application of PSA around the world is then given in Chapter 3. Some of the important technical issues addressed by PSAG in its discussions and Topical Sessions are summarised in Chapter 4. Chapter 5 is devoted to the PSACOIN Exercises, and summarises the content of each exercise, and the conclusions reached. A summary of the achievements of the PSAG is then given in Chapter 6. Although the activities of the PSAG have now finished, it could not be said that the development of PSA is fully accomplished, nor that all philosophical and practical questions about the basis of PSA and the practicalities of its application have been answered. Chapter 7 presents an outlook on some of the unresolved issues.

2. What is PSA?

In this Report, the letters PSA are taken to stand for Probabilistic System Assessment, although the acronym could equally be taken to represent Probabilistic Safety Assessment. A probabilistic assessment could in principle use any method for evaluating the performance of a safety-critical system, in which account is taken of the uncertainty that exists as to the way the system behaves, and in which that uncertainty is expressed in probabilistic terms. In the field of radioactive waste disposal, however, the name has come to be used for a particular computational approach to taking uncertainty into account. This approach, usually called Monte Carlo, involves repeated application of mathematical models, with at least some of the parameters of the models being given randomly selected values for each repetition.

Uncertainty enters into the long-term prediction of the performance of waste disposal systems for several reasons, including:

a) the limited characterisation that can be achieved of the present state of the facility and its surroundings;

b) the unpredictability of future environmental conditions and natural events;

c) the unpredictability of future behaviour of human beings, both as potential recipients of risks and as originators of influences on the behaviour of the system;

d) incomplete knowledge of the physical and chemical processes that combine to produce the overall behaviour of the system;

e) the existence of alternative defensible models for the behaviour of parts of the system.

Because of these uncertainties, any predictive modelling of the system behaviour must have uncertain inputs (choice of models, choice of data values), and consequently uncertain outputs (the measures of system performance). If the uncertainties in the outputs were small (for instance, just a few percent), then they could be ignored, or easily accommodated by the application of small safety factors. It is because the uncertainties in the outputs can be considerable (covering orders of magnitude) that elaborate approaches like PSA have been developed for quantifying the uncertainties. The desirable position of being able to give simple, unqualified assurances of safety is unattainable.

The Monte Carlo method is applied in several other fields. For some applications, e.g. the simulation of neutron fluxes in nuclear reactor shielding structures, the aim is to determine average patterns of behaviour in a system in which a very large number of events or processes occur, with random variation from one occurrence to another. In such cases, a Monte Carlo simulation will involve a randomly selected set of calculated cases that is small compared with the actual ensemble of events being modelled. In application of PSA to estimating long-term risks associated with waste disposal, the situation is rather different. There will only be one future course of the events and processes that constitute the performance of the system; the 'average' consequence is merely an agreed basis for assessing the future behaviour in the presence of uncertainty. No population of disposal systems in an ensemble of equally plausible worlds exists.

The mathematical concepts in PSA

The mathematical principles of Monte Carlo PSA are easy to state. A system is considered whose behaviour can be calculated by applying mathematical models. For an underground disposal facility for radioactive waste, the system will include the waste itself, the engineered components, such as waste packages and repository structures, and the local natural environment. The mathematical models allow the calculation of one or more measures of the performance of the system, such as the radiological dose rate at any time to an individual with given habits. Other measures of consequence could be the health risk corresponding to exposure to the radiological dose, the total release of activity to a specified physical boundary up to a given time, or the fraction of activity contained within a particular part of the system. It will suffice to consider a single measure of consequence, written as

$$q(x_1, \dots x_k) . \qquad (1)$$

This function may be quite time-consuming to compute numerically, using the mathematical models. The consequence, q, has here been shown as a function of several parameters, x_j, which may correspond to measurable properties of system components, boundary conditions applied to the modelled region by the world outside, and so on. All the uncertainty about the system is expressed in terms of uncertainty about the values of the parameters, x_j. In practice, consequences will often be calculated as a function of time, and possible for several locations, exposure pathways, etc., but these are not considered among the parameters.

The means provided by probability theory to express uncertainty about a set of quantities, $x_1, \dots x_k$, is the joint probability density function (PDF), $f(x_1, \dots x_k)$. The probability that the true value of x_1 lies between X_1 and X_1+dX_1, and that the value of x_2 lies between X_2 and X_2+dX_2, and so on, is $f(X_1, \dots X_k)\, dX_1\, dX_2 \dots dX_k$. In the case that all the parameters are independent (probabilistic statements about the value of one x_j do not depend on the values of any of the other parameters), the joint PDF becomes a simple product of independent PDFs, $f_j(x_j)$:

$$f(x_1, \dots x_k) = f_1(x_1)\, f_2(x_2) \dots f_k(x_k) . \qquad (2)$$

Because q depends on all the x_j via a fixed function, the value of q has an implied uncertainty that can also be characterised by a PDF. The aim of PSA is to estimate this PDF (or some other equivalent way of describing the probability distribution of q). A common requirement is to find the expectation-value (mean) of the distribution, and other simple properties, such as percentiles.

The procedure of PSA is to make many calculations of $q(x_1, \ldots x_k)$, using randomly sampled values of the parameters x_j. Suppose that N sample sets are taken, and that in the n-th set the parameter x_j takes the value $x_j^{(n)}$, and the consequence takes the value

$$q^{(n)} = q(x_1^{(n)}, \ldots x_k^{(n)}) . \tag{3}$$

If all the sampled parameter sets are equally probable according to the joint PDF $f(x_1, \ldots x_k)$, then the mean of q, which is formally defined by the integral

$$\bar{q} = \int dx_1 \ldots \int dx_k \, q(x_1, \cdots x_k) f(x_1, \cdots x_k) , \tag{4}$$

can be estimated from

$$\overline{q_e} = \frac{1}{N} \sum_{n=1}^{N} q^{(n)} , \tag{5}$$

that is, the simple mean of the sample calculated q values. Other simple statistics, such as the variance of q, or percentile values for the distribution of q, similarly have simple estimation formulae, usually closely related to the corresponding statistics for the finite population of sample parameter-value sets and corresponding consequence values.

The fact that probabilistic language is used to express the uncertainty in the parameter values does not imply that the physical quantities corresponding to the parameters should be considered as random. Thus, it is far preferable to talk about *uncertainty* in the parameters and in the measures of performance of the system than about *variability* in the parameters or consequences. Variability in time or in space of some observable quantity, such as a rock permeability, may be one reason for uncertainty about the appropriate value for a model parameter that represents the value of that quantity at a particular time or position, or for a parameter that represents an average over temporal or spatial variability. But even when this is the case, the degree of variability of the underlying quantity and the corresponding uncertainty in the parameter will have different measures. From this point of view, it is perhaps unfortunate that some PSA codes have been given names including the word 'variability', e.g. SYVAC (Systems Variability Analysis Code).

The random sampling process of the Monte Carlo approach is also not something that is made fundamentally necessary by the probabilistic description of the uncertainties. The mean value expressed by equation (4) above is simply an integral to be estimated, and systematic rather than randomly sampled values of the integration variables, the x_j, could serve equally well for estimating the integral. Indeed, the simple Monte Carlo procedure is often departed from in PSA for certain purposes. For instance, it is possible to use random sampling that is biased in various ways, to achieve improved accuracy in estimating statistics such as mean values. Systematic patterns of parameter-value selection can also be used in performing sensitivity analysis (that is, testing the sensitivity of the performance measures to individual parameters).

PSA can only be applied to uncertainties that can be expressed in terms of parameters. Some uncertainties with regard to choice of models may fall into this category, and some may not. Furthermore, PSA can only be applied when the uncertainties can be characterised in probabilistic terms.

Features of the models used in waste disposal applications of PSA

The system is broken down into subsystems: Repository and engineered barriers; geological barrier; biosphere; environment, human influences (this breakdown is not, of course, specifically related to the probabilistic approach).

Major pathways for radionuclide transport (groundwater, gas, human intrusion, natural disruptive events)

Indicate processes that typically have to be modelled

As a result of above, system model usually built up out of submodels

Typical approaches to modelling: effective continuum models (1/2/3-dimensional), network models, fracture-system models, compartment models, plain multipliers

In general, PSA is made more practicable if the computational models used execute very quickly on a computer, because many runs of the models must be made. To achieve this, simplifications are often introduced.

Typical PSA code structure: executive plus attached codes or called submodels; modules for input of specification of problem, for random number generation, for statistical analysis of outputs, graphing

Analysis of PSA results

The 'raw' output of a PSA calculation is a large body of numerical data, consisting of a set of output quantities calculated for each of the sample cases. The set of output for each sample case may include several different quantities, such as release rates of activity, cumulative releases, or dose rates; each quantity may be calculated for a number of different radionuclides, and may be evaluated for a range of times and for several different locations. The input data, that is the sampled values of the model parameters, should also be considered as part of the output, since it may be of interest to investigate how the outputs vary with individual parameter values.

This body of output data is of little practical use, simply as a collection of numbers. A variety of statistical analyses are therefore usually conducted. The simplest analysis is to calculate the mean value of some output quantity. Worthy of particular mention is the calculation of individual risk from the mean of the individual annual dose. An individual receiving a dose, D, in one year is reckoned to suffer a resultant risk, R, of some serious health effect (different risk measures can be defined with respect to different health effects). For small annual doses, the risk is taken to be proportional to the dose:

$$R = hD. \tag{6}$$

Values to be used for the constant of proportionality, h, are recommended by the ICRP. Each sample case of a PSA calculation may generate a different value for the annual dose. Each value may be associated with a conditional risk, calculated using the above formula. The overall risk must be calculated by taking the probability of each possible dose level, and multiplying by the corresponding conditional risk. If the PDF of dose, $f(D)$, were known, the overall risk could be calculated as

$$\overline{R} = \int hDf(D)dD\,, \tag{7}$$

that is, as h multiplied by the mean dose rate. The PSA calculation provides an estimate of the mean dose as the average of the dose values given by the set of equally probable sampled cases.

Other simple statistics can be calculated from PSA results. The variance of an output quantity can be calculated as a measure of the spread of its distribution. The median of the distribution (the value exceeded with 50% probability) is an alternative to the mean as an indicative central value. Calculating the mean and other properties of the distribution of the output quantities is part of Uncertainty Analysis (UA), the exploration of the uncertainties that apply to the outputs of the system model, given the uncertainties in the inputs. UA can include investigation of the whole distribution of any output quantity, not just simple attributes such as the mean. The distribution may be displayed graphically in a variety of ways, as explained in the next subsection.

A second aspect of the analysis of PSA results is Error Analysis. The statistics such as means calculated in the UA are only estimates of the properties of the distributions of the output quantities. The true values of these statistics could only be found exactly by the Monte Carlo PSA method if an infinite number of sample cases were taken. It is therefore important, for any given finite sample size, to have some estimate of the possible error in a calculated mean or other statistic. A widely used way to express this estimation error is to attribute Confidence Interval to an estimate of a statistic. For instance, an estimate of a mean value may be given as $M \pm E$, at a given level of confidence, for instance 95%. This indicates that a probability of 95% can be attributed to the true value of the mean lying within a distance E either side of the value M. In general, the interval size $\pm E$ depends on the variance of the distribution of the quantity whose mean is being estimated, the number of samples taken, and on the desired level of confidence. The relationship is given in most elementary texts on statistics for the case where the sampled quantity has a Normal distribution. However, this is far from being the case for the outputs of most waste disposal system models to which PSA is applied.

The question of obtaining confidence intervals, without assuming anything about the shape of the distribution, was considered by PSAG in its discussions and in conducting intercomparison exercises. The Group was grateful to have its attention drawn by Woo [ ] to a little-known analysis by Gutman [ ] on this subject. In general, a confidence interval size is found using a formula of the form

$$E = \alpha S_N / N^{1/2} \qquad (8)$$

where S_N is the standard deviation of N sampled values of the output quantity. For a confidence level of 95%, the constant of proportionality, α, is 1.96 if it is known that the quantity has a Normal distribution. A theorem due to Chebyshev leads to a value of 4.47, based on no restrictive assumptions on the distribution of the mean estimator. However, the analysis of Woo [ ] following Gutman justifies a value for α of 2.86, taking credit for the fact that the mean estimator is itself the sum of N samples from the same parent distribution. For higher levels of confidence than 95%, the advantage of the Gutman limit over the Chebyshev limit is even greater.

A third kind of analysis of PSA results is Sensitivity Analysis (SA). This means the characterisation of how sensitive the calculated results are to the values of the individual input parameters. The question of how SA should be performed was the subject of several discussions of the PSAG, and a whole intercomparison exercise (Level S) was devoted to the subject (see Section 5.5). SA is an important tool in gaining understanding of complex systems. However, the strongest reason for interest in SA in the application of PSA to waste disposal safety analyses is that it can help to guide the acquisition of data. The degree of uncertainty in each of the model parameters used in a PSA calculation is not an inherent property, but can generally be reduced by further research, particularly the gathering of experimental information. Since this process is expensive, it is important to know which parameters most sensitively affect the risks or other performance measures, so that effort to reduce uncertainty can be directed with greatest efficiency.

Presentation of PSA results

The results of a PSA calculation carry a great deal of information, and it is important to present this information in a way that will communicate clearly. Various forms of graphical presentation are commonly used to achieve this aim. The two main objectives are to address the given regulatory targets (these differ from one country to another), and to provide understanding of the system being modelled.

Regulations that are framed in terms of a single output quantity, such as cumulative release to a given boundary during a given period of time, are well served by plotting the full distribution of that output quantity. One way to portray this distribution is in terms of the Complementary Cumulative Distribution Function (CCDF). For a quantity q, the CCDF $\overline{F}(Q)$ gives, for any value Q the probability that q is greater than Q. $\overline{F}(Q)$ is always a monotonically decreasing function, tending to value 1 as Q becomes smaller, and to value 0 as Q becomes larger. If it is desired to keep separate two different kinds of contributions to the total uncertainty in q, it is possible to plot a family of CCDFs. Each single plot shows the distribution of q values arising from one class of uncertainty. The spread in the position and shape of the different CCDFs provides a portrayal of the contribution from the other class of uncertainty.

Regulations that are framed in terms of the mean of some quantity that depends on time, for instance annual individual dose or risk, are addressed by plotting the calculated mean(s) as a function of time. Variants of this kind of plot are useful for showing additional information. The confidence intervals associated with estimating the means can be shown using error bars, or line plots of the upper and lower interval limits. The contributions of different radionuclides to total dose or risk can be plotted along with the total to show clearly the way that the relative importance of the contributions varies with time.

Sensitivity analysis is concerned with the relationship between individual input and output quantities. If a given output is plotted against a single model parameter in the form of a scatter diagram, the degree and nature of any correlation between them can be made evident. In this case, although the system model gives a well-defined functional relationship between the inputs and outputs, a plot of one output against one input quantity shows random scatter, because the results plotted include the influence of other parameters that have been randomly sampled.

All of the graphical presentation forms mentioned can contribute to understanding of the system. However, further communication of the significance to overall performance of selected contributing processes, or of selected subsystems, is often sought by means a variety of plots. For instance, in a multi-barrier disposal system, the different barriers (primary containment, chemical conditions in the repository, engineered and natural barriers to migration, biosphere processes of dilution) contribute with different levels of efficiency, depending on the radionuclide concerned, and on time. Measures of individual barrier performance can be devised (such as the fractions of activity contained and released at each stage), and portrayed graphically using column or pie charts.

3. Development and Application of PSA

This chapter presents a brief history of the development of PSA for waste disposal assessments. The account is closely based on the excellent summary provided in a recent article by Thompson and Sagar [ ].

Other accounts can be found in ???

3.1 World-wide Development of PSA Methodology

During the late 1970s and early 1980s, pioneer work on the application to radioactive waste disposal of PSA using Monte Carlo methods was undertaken in the USA by the Sandia National Laboratory. The studies were applied to hypothetical high-level waste (HLW) disposal facilities in a variety of geological settings. The program SWIFT II was developed for detailed groundwater flow modelling, and results obtained with this program were used to construct a network flow model composed of one-dimensional segments. The network flow-and-transport program NEFTRAN was then used in probabilistic mode to estimate radionuclide releases with account taken of parameter uncertainty.

At about the same time, work began in Canada on SYVAC (System Variability Analysis Code) version 1, again with a view to evaluating the potential deep geologic disposal of HLW. Analytic solutions for the source term and for groundwater-mediated transport in one dimension were evaluated and combined numerically, using parameter values sampled from PDFs based on expert judgement.

SYVAC 1 was taken up in the UK by the DoE team, who adapted the program to UK-specific deep and shallow disposal concepts by providing new submodels. The resulting programs were named SYVAC D and SYVAC A/C. Meanwhile, separate development in Canada led to versions SYVAC 2 and SYVAC 3, incorporating multiple source-term submodels, network geosphere transport submodels, and a compartment model of biosphere transport and radiological exposure. These programs, and the extensive databases of parameter values and distributions, were developed under a rigorous Quality Assurance regime.

During the 1980s, several European groups undertook development of PSA programs, including LISA (developed at Ispra under the CEC radioactive waste management research programme), EMOS (in Germany), PROPER (developed at SKB) and MASCOT (developed by AEA on behalf of UK Nirex

Ltd). Many of these codes were developed in the context of cooperation and information exchange provided by the PSAC User Group, as PSAG was then known.

Interest in representing the processes occurring in the biosphere, as part of probabilistic safety-assessment modelling, varied among the national groups involved. This was largely conditioned by the regulatory frameworks being addressed. Most relevant US regulations, for instance, are based on releases of radioactivity to a notional boundary around a facility, rather than on doses that would arise in a specific set of biosphere conditions with a specific pattern of human behaviour. The length of time into the future for which assessment of releases, doses or risks is required is also relevant to the degree of concern with changes in the environment.

Among those interested in realistic biosphere modelling over a long time-frame, the UK DoE programme stood out in giving attention to modelling the future evolution of climatic conditions and their influence on the repository environment. This work gave rise to the TIME2 and TIME4 codes for generating descriptions of environmental evolution in a way that could be driven by random selection of parameters, and thus incorporated into the PSA framework. To complete the capability of conducting PSA analyses with time-dependent models, the VANDAL code was developed (in several versions), incorporating submodels for repository releases and groundwater flow and transport, and a dynamic biosphere submodel.

The effect of possible future human actions causing releases of radioactivity or otherwise affecting the system performance requires a rather different approach than that for other exposure pathways. Particular attention to this class of phenomena was given in work undertaken on behalf of the USDOE for the Waste Isolation Pilot Plant (WIPP) in New Mexico. Various intrusive activities, including exploratory drilling, well-water abstraction and solution mining for potash were all represented as independent sequences of events whose timings and magnitudes could be obtained by sampling from probability distributions.

3.2 Applications of PSA

The intention in this subsection is to provide a brief survey of the applications of PSA that have been made during the life of PSAG. Can continue to borrow from Thompson & Sagar's account, but since this concentrates on UK and US applications, need supplementary information on other applications, e.g. in Canada , and CEC.

IT WOULD BE HELPFUL IF VARIOUS PEOPLE READING THIS DRAFT COULD PROVIDE ONE TO TWO PARAGRAPHS ABOUT THE PSA APPLICATIONS WITH WHICH THEY ARE FAMILIAR.

4. Topics Discussed by PSAG

Statistical sampling strategies

In Chapter 2, reference was made to the fact that probabilistic system assessment need not necessarily be based on the Monte Carlo method. However, all the experience represented by PSAG is with methods involving random sampling of parameter values from specified PDFs. Several different sampling schemes have been used. In Simple Monte Carlo (SMC), each independent parameter is sampled directly from its 'true' PDF, that is, the sample values are all equally likely according to the PDF that is specified as quantifying expert opinion as to appropriateness of using different values of that parameter with the given models and for the given application. A variant is Latin Hypercube Sampling (LHS), in which an attempt is made to avoid fluctuations in sampling density associated with the limited number of sample cases. This is done by ensuring that for each parameter, each of a number of equally probable discrete sub-ranges is represented exactly once. Some PSAG member groups have used various forms of Importance Sampling (IS). With IS, the sampling process is similar to that in SMC (no restriction on representation of sub-ranges as in LHS), except that the PDFs from which the samples are drawn are different from the 'true' PDFs. The differences between the 'true' and sampling PDFs are determined in a way that aims to improve the accuracy, for a given number of sample cases, of estimation of the mean of a given output quantity. In effect, IS focuses on a particular part of 'parameter space', that which is most important in determining the mean of the output quantity of interest. To compensate for the bias in the sampling, the estimation formulae for means and other statistics have to be modified.

The three sampling schemes just described were all used in the Level E exercise, but no firm conclusions could be drawn about their relative merits on the basis of that exercise, because the number of contributions was insufficient to indicate any systematic tendency for one scheme to produce answers closer to the exact mean than another.

The measure of performance of a sampling scheme is that the means or other statistics that are estimated from the sample cases converge rapidly to the true result as the sample size (number of cases) increases. Of course, in general the true result is unknown, and it is difficult to measure the convergence simply by watching the progress of the estimated mean or other statistic as the sample size is increased. This is because, when the distribution of some consequence value is skewed, values that are much larger than the mean can occur with low probability. By definition, such outlying values arise 'without warning' at infrequent intervals as samples are taken. The running estimate of the mean can therefore appear to stabilise, only to be significantly perturbed by a large sample output value appearing.

In one PSAG Topical Session, Iman [ ] argued that LHS would give convergence significantly better than that of SMC when the number of sample cases is not very large. This claimed advantage appears to be relied upon by a significant number of users of the PSA methodology. A mathematical demonstration that the estimation accuracy of LHS is better than or equal to that of SMC is possible in the special circumstance that the consequence depends monotonically on each of the model parameters. Even in this case, the demonstration provides no formula for the degree of benefit.

Robinson Roberts and Sinclair [ ] presented a paper at one Topical Session that gave results of a comparison between sampling schemes applied to a model system similar to that used in the Level E test case. This work was based on repeating the application of each sampling scheme a large number of times. On this basis, it was shown that, for the model system under consideration,

a) LHS did not perform significantly better than SMC, both being slow to converge as the sample size increased, for the reason that the largest values of consequence tended to arise from a rather localised region of parameter space;

b) IS gave significantly better performance than SMC or LHS (a factor of 10 on the number of samples required to attain a given accuracy). This improved in performance applied to estimating means of the output quantities for which the 'focusing' of the IS was designed;

c) IS performs badly (worse than SMC or LHS) for statistical results for which it is not 'focused';

d) a modified version of Importance Sampling (MIS) could be used that kept the good performance of IS for statistics on which it is focused, while retaining adequate performance (similar to that of SMC or LHS) for results on which it is not focused.

Different treatments of uncertainty

The assessment codes used by members of the PSAG are all probabilistic in nature, and are therefore aimed at taking into account uncertainties regarding the systems to be modelled. However, the manner in which uncertainty is treated can be affected by the view taken of the nature of uncertainty. This question was the subject of one of the topical sessions of the PSAG, and has been discussed on several other occasions.

It is often said that the uncertainties encountered in radioactive waste disposal safety assessments are mainly subjective in nature. The exact meaning of this statement is difficult to determine, but the main idea is that the uncertainties arise because of a lack of precise knowledge, although that knowledge could in principle be gained by sufficient investigation of the system. This is contrasted with a situation in which the behaviour of a system could be governed by stochastic events or processes operating in the future, about which only probabilistic statements could be made in the present, however much investigation were carried out. This second situation is encountered, for instance, in reactor safety studies where component failure is a stochastic process. Classification of uncertainties into Type A (stochastic) and Type B (subjective) is proposed by Hofer [Madrid], although many statisticians question that any such distinction is valid.

The justification for using probability theory to manipulate subjective uncertainties is through the Bayesian interpretation. This was clearly presented to the Group by Smith [Madrid], who explained it in terms of a betting analogy. If one is prepared to bet on the true outcome of any uncertain event in a rational (i.e. self-consistent) way, then such betting must be governed by the rules of probability theory.

Others have argued for the use of approaches to handling uncertainties based on Fuzzy Set theory. Shaw and Nies [Madrid] both presented the mechanics of this approach, using 'min-max' rules to combine elementary fuzzy belief measures, where the rules of probability would be used to combine elementary probabilities. The main objections to the Fuzzy approach were that it is not necessary (because the probabilistic approach is said to be able to handle all cases), and that there is no equivalent to Bayes' Theorem for updating prior estimates.

Robinson [Paris 1992] presented an interpretation of the Fuzzy approach as a generalisation of interval analysis, and spelled out some details of how it could be implemented using Monte Carlo methods not dissimilar in mechanics to those of the traditional PSA.

The key difference between the probabilistic and Fuzzy approaches is that they address different aspects of uncertainty. The probabilistic approach considers how expert judgement leads to a rational bet on what the outcome will be. The Fuzzy approach considers how expert judgement constrains what the possible outcomes could be.

Reduction of research codes to PSA submodels

Brief account required

Spatial variability

Brief account required

Derivation of parameter PDFs

Brief account required

Any other Topical Discussions of which anyone has a decent record?

5. The PSACOIN Exercises

The PSACOIN exercises for intercomparison of PSA codes have constituted the activity consuming most of the time of PSAG, and that which has resulted in the most lasting effect, in that extensive reports have been published. The conduct of such exercises was envisaged from the beginning of the life of the PSAG.

International code verification, validation and intercomparison exercises are a well-established method for helping to increase confidence in the investigation and assessment methods employed in radioactive waste management. Most exercises of this type have focused on models for particular subsystems of a disposal system. Thus, HYDROCOIN [ ] and INTRAVAL [ ] have concentrated on groundwater flow and radionuclide transport through the geosphere; BIOMOVS [ ] has specialised in biosphere transport models; geochemical and geomechanical models have been addressed by the CHEMVAL [ ] and COSA [ ] exercises, respectively.

The PSACOIN series differs somewhat in attending to models of the entire disposal system. The participants have been distinguished, not by expertise in any particular class of physical or chemical phenomena, but by interest in and involvement with the development of probabilistic assessment tools for waste disposal systems. The aim, therefore, has not been specifically to validate models aqainst 'real-world' behaviour, but rather to build confidence in the correct operation of what can be quite large and complex computer programs, incorporating (or at least controlling) large numbers of submodels for different subsystems, and operating in a mode (the Monte Carlo PSA method) which may appear at first sight to lend a degree of obscurity to the results.

Indeed, there are some problems peculiar to comparing probabilistic results. The twin problem is to verify the correct operation of the underlying models, and to verify the correct application of the procedures to implement treatment of parameter uncertainty and the way it produces uncertainty in the output quantities. If two PSA codes produce different results, it is necessary to determine whether the models themselves are operating differently, or whether the differences belong more to the set of sampled cases with randomly selected parameter values.

Work on the PSACOIN exercises began at the first meeting of PSAG. A series was envisaged in which the basic methodology, independent of the particular models, would first be tested, followed by a progression towards modelling of the kind that will be involved in real national safety assessments. Because of this envisaged pattern of increasing complexity and realism, the exercises were named as

'Level *X*', where *X* was to be numeric. As the series progressed, however, certain exercises to address particular issues were devised, and non-numeric Level names were added.

After initial conception of each exercise, the usual pattern for carrying it forward has been to appoint a Task Group to draw up a written specification and a questionnaire for eliciting contributions in a standard form. This work may involve pilot running of the test case to discover any difficulties inadvertently introduced by the draft specification. With the support of the NEA, the Task Group analyses the responses, and draws up a written report for publication by the OECD. At every stage of this process, approval from the whole PSAG has been obtained, and in many cases, there has been iteration of participants' contributions, in the light of questions raised by the Task Group's analysis.

In the following subsections, each of the PSACOIN exercises is described briefly.

5.1 Level 0

This first exercise aimed to test the 'executive' functions of PSA codes, including in particular the process of random selection of parameter values from different probability distributions, and the application of mathematical system model composed of linked submodels. The exercise also aimed to test post-processing, that is statistical analysis of the results, including sensitivity analysis. For these purposes, the exercise specified an extremely simplified model of release of radionuclides from a repository vault, one-dimensional transport through the geosphere, and doses arising from consumption of drinking water. Twelve organisations contributed results.

The 'simple' model turned out to pose certain problems for statistical convergence unlikely to be typical of real applications, but after four iterations a generally good agreement was obtained for estimates of mean dose versus time, and it was concluded that the executive parts of the PSA codes and their post-processors were operating as expected. Lessons were learned with regard to specification of exercises, analysis of results for comparison, and presentation of PSA outputs. Unresolved issues over the rating by 'importance' of the model parameters led to the conception of further exercises.

The Level 0 report was published in 1987.

5.2 Level E

Level E shared the same purpose as Level 0, with the important addition of comparison with an exact probabilistic solution that enabled mean dose against time to be calculated precisely for a model disposal system. The model, while still relatively simple, was more realistic than that used for Level 0. It involved release from a vault by groundwater leaching, a two-layer geosphere transport path, and a drinking-water dose model. The exercise supplemented the probabilistic case with several deterministic cases (fixed combinations of model parameters), in order to assist with establishing correct code operation and data input. This practice was followed in later exercises.

The 'exact solution' for the mean dose against time was made possible because certain aspects of the mathematical modelling were amenable to an analytic approach. In the source term (the release rates of radionuclides from the repository), no nonlinear effects (such as solubility limitation) were specified to occur, but only simple leaching, as by a constant flux of groundwater, giving radionuclide fluxes proportional to the remaining inventory. Similarly, each of the two geosphere layers were treated in one dimension, with advection and dispersion occuring in a uniform groundwater flow field, and with a simplifying choice of boundary conditions. The drinking-water dose model simply required a multiplying factor. All of this meant that, after applying a Laplace transform to the time dimension, a solution for a given choice of submodel parameters could be written down analytically. The PDFs for the submodel parameters were then specified to have sufficiently simple forms that averaging of the Laplace-transformed solution over the parameter distributions could also be done analytically. The only step requiring numerical evaluations was the final inversion of the Laplace transform to give mean dose as a function of time. This step could be carried out to give results that were accurate to some 4 decimal places, and certainly suffered no uncertainty associated with limited sample-set size.

Of course, the extraction of the exact solution was not something that the exercise participants were called upon to repeat. The purpose was for ordinary Monte Carlo PSA codes to be applied to the problem, ignoring the analytic route to the solution. Different participants tackled the problem using different parameter sampling schemes (simle random sampling, Latin Hypercube sampling, Importance sampling), so there was an opportunity to compare whether these choices had any influence on the proximity of the results to the exact solution.

At the time the exercise was being framed, the issue of convergence of PSA results was a matter for debate within the PSAG. A proposed approach (using the 'Shapiro-Wilkes' test) to testing the validity of specifying confidence limits on estimates of mean values was recommended to participants for evaluation.

The different predictions obtained with the 10 participating PSA codes were found to agree generally very well with the exact solution and with each other, despite the use of different sampling methods, different time-step algorithms, and sample sizes ranging from 100 to 10 000.

The study was unable to be conclusive about the relative merits of different parameter sampling schemes, which at least indicated the absence of any overwhelming advantage to any one scheme for this particular problem. The Shapiro-Wilkes test was not found to be a reliable indicator of convergence. However, the exercise was useful in providing an occasion to explore the assignment of confidence limits on the basis of Normal statistics, Chebyshev's theorem, or the Gutman analysis (see Chapter 2 for a brief explanation).

The Level E report was published in 1989.

5.3 Level 1a

The Level 0 and Level E exercises had required most participants to write special-purpose code to implement the models specified. Level 1 began the trend to greater realism by encouraging the use of production submodels already developed by participants for use in their national programmes.

The Level 1a test case addressed a deep repository concept; it specified modelling of mobilisation of radionuclides within the vault, release into a two-stage geosphere transport path, and calculation of doses from drinking well-water. One aim of the exercise was to begin to evaluate the relative importance of the uncertainties associated with data values and those associated with the possibility of taking different mathematical approaches to modelling a given set of processes. This aim could only partially be fulfilled, because the test-case specification was fairly prescriptive with regard to geosphere transport modelling methods, although it gave more freedom with respect to the vault modelling. Some of the modelling differences that arose among the contributions concerned the treatment of advective and diffusive mechanisms for the transport of radionuclides out of the repository.

It was concluded that to make a more instructive comparison of the effects of parameter-value uncertainty and modelling uncertainty would require an exercise in which the modelling approaches were constrained only by the existence of a body of data (whether real experimental data, or synthetic field data created for an exercise), which would serve to justify or rule out any particular modelling approach. In the Level 1a exercise, modelling choices were constrained by arbitrary prescriptions in the test-case specification, and in some cases by the fact that participants had chosen to use existing models not necessarily ideally designed for this particular application.

The Level 1a report was published in 1990.

5.4 Level 1b

The Level 1b test case contrasted with previous exercises in focusing on biosphere modelling, to the extent of excluding modelling of the geosphere altogether. In the earlier exercises, the measure of radiological consequence used was individual dose, assumed to arise through consumption of drinking water. Transport processes within the biosphere were not modelled; the calculated rate of release of each radionuclide from the geosphere was simply multiplied by a constant to obtain the annual amount ingested, and a further radionuclide-dependent factor then gave the resulting dose.

In Level 1b, by contrast, a simple compartment model was used to represent transport processes within the biosphere, and seven different exposure pathways were considered in arriving at the total dose that an individual could receive. A simple source term (the rates of release of radionuclides into the biosphere) was specified, which did not seek to represent in any significant detail processes involved in release from a repository and transport through the geosphere. This specification could be considered as representing approximately a near-surface disposal facility, but no use was made in the exercise of such a conceptual framework.

The exercise specification, being prescriptive with regard to the mathematical representation, did not give an opportunity to compare different interpretations or treatments of a common body of information. Rather, the objectives of the exercise were to give participants experience in the application of PSA methodology to biosphere transport and exposure submodels, and to explore the effects of parameter uncertainty in these submodels, in the context of a system where individuals can be exposed by several pathways. To the extent that the codes used by participants were applicable to other systems, the exercise would also usefully contribute to the verification of these codes.

Comparison of the contributions obtained from participants, using a number of different codes, showed good agreement in general, not only for the probabilistic results, but for a deterministic 'central case' (in which every parameter was fixed at a value central to the PDF defining its uncertainty). It was felt that the exercise provided a demonstration of the practicability of incorporating into PSA models representations of biosphere processes with a modest level of complexity. The advantages of so doing include being able to allow explicitly for the influence of, and the uncertainty in, a number of processes that may interact in a complex way.

The results were interesting in a number of ways, illustrating several effects that may arise in PSA applications incorporating biosphere models. Of course, the importance of some of the effects may depend on the system being modelled.

- The peak mean dose calculated in the probabilistic case was about 5 times greater than the peak dose in the deterministic central case. Thus, consideration of uncertainty in biosphere parameters is important, although uncertainty in other subsystems, particularly transport in the geosphere, can sometimes give rise to somewhat greater effects.
- No single model parameter was found to be a sensitive determiner of doses. The uncertainty in dose arose from the combined effect of many parameter uncertainties.
- The drinking water pathway for radiological exposure, which is in many applications the only one considered, was far from being the most important pathway in the system studied in this exercise. The relative importance of the various exposure pathways was found to vary as a function of time.
- Some of the timescales for redistribution of activity in the biosphere were found to be long (in excess of 10^4 years). Depending on the other timescales in the overall system (such as those for release from the repository and for transport in the geosphere), this fact could make it important to represent transport within the biosphere, rather than using an equilibrium model that represents the relationship between flux released from the geosphere and dose received via a given pathway as a time-independent proportionality, ignoring transient effects due to transport delays.
- There was evidence that the nature of the interface between the geosphere and the biosphere requires careful consideration. Unrealistic specification of the boundary conditions could lead to inaccurate calculation of concentrations in the biosphere.

5.5 Level S

This exercise was in several ways different from previous exercises. First, its focus lay on one particular stage of PSA methodologies, sensitivity analysis. Thus, it was not concerned with system modelling itself, but with data post-processing. A common data set could thus be made available to participants, to use if they wished. The model system for study was the same as that used in the Level E exercise.

Secondly, although a quantitative set of questions formed part of the exercise, there were also more general questions about sensitivity analysis, and participants were encouraged to submit analyses of the common data set other than those directed at answering the basic questions, as well as original ideas on the subject of sensitivity analysis in general.

In part, the Level S exercise grew out of a question that arose in many of the preceding exercises, and which was not simple to resolve. In the earlier exercises, participants were asked to rank the parameters of the systems model in question in order of importance. The widely varying responses indicated that the seemingly simple question "which are the most important parameters of a given system model?" has no unique meaning, and can be answered in many ways.

One traditional approach to sensitivity analysis addresses the question with respect to a reference point in 'parameter space'. The local variation of a model output of interest in the vicinity of this point is examined as the parameters are changed in some systematic way. Such an approach may be called 'deterministic', to contrast it with the situation in which the parameter-value combinations are sampled randomly from the distributions that characterise their uncertainties. Since these samples, together with the corresponding model output values, are available as the result of a PSA calculation, it is plainly of practical interest to conduct a sensitivity analysis in this way, without the need for a separate set of calculations.

In the Level S exercise, and indeed in other PSACOIN exercises, many participants performed sensitivity analyses based on randomly sampled parameter values either by extracting correlation coefficients between the parameters and one or more model outputs (such as dose at a particular time), or by performing regression analyses. These approaches are closely related and can be thought of as fitting simple functional forms (e.g. linear) to the dependence of the output quantity on the parameters. The fits obtained depend on the parameter distributions. In the special case of very narrow distributions, a linear form would in general provide a good fit, and the correlation or regression coefficients would obviously bear a simple relationship to the gradients that could be obtained from deterministic sensitivity analysis. However, this limit does not usually apply in the case of PSA for waste disposal systems. In any case, the regression and correlation analyses focus, like deterministic sensitivity analysis, on the variation of a given model output value as the input parameters span their ranges of interest.

The Level S problem specification and questionnaire, however, encouraged yet another approach, termed 'distribution sensitivity analysis' in the final report for the exercise. Here, the focus is on changes to the characteristics of the whole distribution of output values produced by a PSA, as changes are made to the input parameter distributions. The motivation for taking such an approach is as follows. First, safety assessments are concerned with the distribution of consequence values, or with broad characteristics of this distribution, such as mean and spread. Secondly, the parameter distributions are not fixed absolutely, but are potentially subject to change as site characterisation or other data gathering reduces the uncertainty in some parameter values. In order to optimise the direction of such experimental work, it is of interest to estimate how the consequence distribution characteristics, which are the end-points of a safety assessment, would be affected by a hypothetical change in the distribution characterising uncertainty in one of the parameters.

In the model system for Level S, the 12 uncertain parameters all had either uniform or log-uniform distributions. The quantitative questions did not suggest consideration of any departures from this class of shapes, simply shifts and symmetric changes of width. The model output quantities considered were mean dose at particular times, and maxima of dose up to particular times. The distributions of these output quantities did not, of course, conform to any standard shape. The questions about the sensitivity of the output distributions concerned only the mean of a given dose value, and the 'spread' of the distribution, which was typically quantified by the variance or standard deviation. In order to allow consideration of shifted input-parameter distributions, the supplied common data set included randomly sampled parameter values covering wider distributions than the specified originals. By appropriate censoring of the data set, samples could be obtained covering the original ranges and shifted ranges as required.

The exercise achieved its objective of seeing new techniques of sensitivity analysis developed. Where numerical agreement was to be expected between different participants' responses to specific questions, it was generally found. Ranking of variables by importance in terms of 'distribution sensitivity' was satisfactorily consistent, at least for the most influential parameters.

A major conclusion of the exercise concerns statistical error in the probabilistic sensitivity analysis. Many of the estimated quantities cited in the contributions were subject to standard errors that exceeded the corresponding estimated quantities in size. In other words, the estimated quantities could not be significantly distinguished from zero. As with all statistics calculated from PSA results, a reduction of error is to be expected as the number of sample cases is increased. However, it appeared that a sample size adequate to estimate mean doses with sufficient accuracy for practical purposes may often be inadequate for estimation of sensitivities of the kind considered in this exercise. This outcome emphasises the need for error analysis to accompany all statistical calculations. Some doubt is cast on the suitability of standard PSA calculations for providing accurate sensitivity information, but the question could not be resolved within the scope of this exercise.

A further question left outstanding is the appropriate way of handling what is a common characteristic of PSA calculations for waste disposal systems, that many of the sample consequence values are zero, or effectively so. Correlation or regression analyses based on ranked output values can be significantly affected by this large number of 'tied' values, which is not typical of many other applications of statistical analysis.

The Level S report was published in 1993.

5.6 Level 2

The long period of discussion about the direction for Level 2

Purposes of exercise as finally adopted

Planned phasing

Summary of results for Phase 1

Lessons

6. Summary of Achievements of PSAG

By sponsoring the existence and activities of PSAG, the NEA has provided important and timely support for a number of OECD member countries in a period in which they were seeking to develop a capability to conduct safety assessments for potential and actual radioactive waste disposal facilities. The whole process of developing concepts for the disposal of radioactive waste, identifying and evaluating potential disposal sites, and assessing the long-term radiological safety of specific facilities, is one that has aroused strong interest by a variety of interested parties, and will continue to be subject to intense and searching scrutiny. In this context, most countries have seen as very valuable the search for international consensus with regard to philosophy and methodology, and the possibility to enhance confidence in their assessment tools by international cooperation and mutual review.

The major goals of the PSAG have been to promote the exchange of information, ideas and computational tools between participating organisations, and to help to build confidence in PSA methodology in general, and in the PSA codes used by participants in particular.

Information exchange

The effective exchange of information was encouraged by drawing together from many countries experts in a highly specialised field. The PSAG membership has always included many who were actively engaged in the development of PSA codes, and who were therefore intimately familiar with the practical details of implementing the methodology. The discussions of PSAG have always therefore been strongly 'earthed' in the practicalities of applying PSA. PSAG was a group where a complex methodology was applied in practice, not simply talked about.

PSAG has also benefited from a significant presence of those who were involved in shaping the direction of national programmes for assessing waste disposal facilities. The Group therefore not only constituted a valuable source of first-hand information, but also provided a contributory influence in the formulation of national approaches to this technical problem. PSAG member countries were at different stages in their development of safety assessment methodology, and for some groups, PSAG was the starting point for their developments. Participation in PSAG has come from more than 10 different OECD member countries, as well as from the CEC and the IAEA. The work of the PSAG has continually been reviewed by the NEA Performance Assessment Advisory Group.

The activities involving information exchange and technical discussion have covered wide ground, including the design of PSA codes, development of specific submodels for PSA codes, the theory of statistical convergence, the fundamental basis for treating uncertainty in probabilistic terms, reduction of research codes to practical assessment tools, and elicitation of expert opinion.

Code Intercomparisons

The goal of PSAG to help build confidence in the computational tools used by member groups in safety assessments was forwarded principally through the PSACOIN intercomparison exercises. These were conducted in a series of 'Levels', designed first to test the basic framework of PSA codes, and then to move progressively towards greater realism, in the sense of studying test cases similar to real assessment problems, and in using models drawn from those actually being used in national programmes.

Levels 0 and E tested the executive functions of the participating PSA codes, including random generation of parameter values from specified probability distributions, and the statistical analysis of results. Practical attention to these questions brought to light the need to select effective ways of presenting PSA results, and the difficulties of giving precise meaning to the question of which model parameters are the most important.

The Level 1a and 1b test cases represented respectively deep and shallow disposal systems in a more realistic way, giving some freedom to participants to use submodels developed for use in their national programmes. These exercises, by raising the question of how wide a range of different models could validly be applied to a particular problem, enlightened preparation for a Level 2 or later exercise in which uncertainty associated with model choice could be more fully examined. Level 1b gave a valuable opportunity for members to gain experience in the incorporation of some complexity into the biosphere models used within a PSA calculation.

A general result of Levels 0, E, 1a and 1b was that the participating codes gave substantially agreeing results, where such agreement could be expected in the light of the models being used.

The Level S exercise specifically addressed methods for sensitivity analysis. It differed from the other PSACOIN exercises in concentrating more on methodology than on comparison of results obtained using different codes applied to the same problem. The published report contained articles contributed by participants, and some of these articles presented new approaches and techniques for sensitivity analysis.

Finally, Level 2 represented a major increase in complexity and realism, attempting to address the question of the relative importance of parameter-value uncertainty and uncertainty with regard to choice of conceptual and mathematical models where alternatives exist that cannot be ruled out on the basis of available information.

Final status

At the time of the final PSAG meeting, the planned course of the Level 2 exercise was only partly covered. Furthermore, not all technical, philosophical and practical questions related to PSA have been fully resolved (see Chapter 7). Further international attention to these matters may be achieved in the future through other means. At this stage, some of the national groups that have participated in PSAG are very actively engaged in preparing safety cases for licence applications for specific disposal facilities. Others are continuing in the process of justification of disposal concepts and examining the relative merits of alternative disposal sites. The work of PSAG has created a body of reports on technical issues that will continue to be relevant to these activities, and the published PSACOIN exercise results will provide a set of benchmark tests that can be used as part of verification of new PSA codes.

7. Unresolved Issues

Despite nine years of fruitful work, the PSAG has not, of course, completely answered all questions regarding PSA and its application to radioactive waste management. The way forward for international cooperation in this technical area is not yet clear. This Chapter provides a brief account of some of the issues that appear to remain unresolved.

7.1 Unresolved Issues in the Fundamentals of the Treatment of Uncertainty

The validity of the probabilistic approach

Most countries involved in the use of nuclear power have established some guidelines for the evaluation of schemes for disposal of radioactive wastes. However, not all regulatory frameworks are very explicit about the way in which uncertainty should be taken into account. Among those that either require or allow a quantitative treatment of uncertainty, there is even less guidance as to the fundamental nature of uncertainty, and what language is appropriate to discuss it.

The work of the PSAG has all been based on probabilistic approaches to treating uncertainty. This means that it has been assumed that the theory of probability provides an appropriate basis for quantifying uncertainties and for combining uncertainties. Nevertheless, PSAG has been open to the debate on the question whether other philosophical bases are appropriate, and Fuzzy Set theory in particular has been discussed (see Chapter 4). It is also hard to avoid the impression that, while many are applying the PSA methodology, there is a level of unease about the appropriateness of using probabilistic language when eliciting expert judgement about parameter values when there is a marked lack of direct experimental evidence. Thus, the PSAG has also debated the idea that different types of uncertainty should be distinguished, and handled differently. The PSAG did not reach definitive views in any of these debates.

It is possibly to avoid open questions such as these that some countries are not currently using PSA in preparing assessments for proposed disposal facilities. Rather, deterministic analyses are made, using conservatism as the main means of dealing with uncertainty. This means choosing models and parameter values in a way that should tend to overestimate the risks or other similar performance measures. The conservative approach is respectable, but open to some dangers. First, in a system with many components and many sources of uncertainties, many conservative assumptions may have to be made. In combination, they may lead to prediction of unacceptable system performance, whereas an assessment that was realistic at every turn may well have predicted acceptable performance. Secondly, it can sometimes be difficult to make an assumption that is conservative in every way. For instance, ignoring some system feature or simplifying a process may tend to make the consequences larger at most times, but smaller at others. In this event, the conservatism cannot be simply defended. Finally, it is rarely possible to agree an absolute bound on the value of a parameter for which uncertainty exists. Hence, the consequences calculated by an analysis with so-called conservative parameter settings will rarely be absolutely bounding. To resolve this difficulty, one is tempted to describe the adopted parameter values as those for which there is only a small probability of their being non-conservative. Such an approach followed logically would lead to adoption of the PSA methodology, at least for characterising the upper tail of the consequence distribution.

The PSAG in its activities has concentrated on providing help to its members in developing the PSA approach, not on convincing non-members of the superiority of PSA over alternative approaches.

Can PSA only handle some types of uncertainty?

It is sometimes asserted that PSA is a tool for dealing with one type of uncertainty, that associated with the values of model parameters, while other types of uncertainty cannot be dealt with using PSA. Such a clear-cut statement is open to debate, and it is a matter still to be determined exactly which types of uncertainty can, or should, be treated in this way.

One thing is certain: if an aspect of uncertainty about the system being studied cannot be quantified in probabilistic terms, it cannot be treated by a probabilistic assessment method. For instance, in any assessment, it is possible that human errors have been made in assembling the data, in reporting the results, or in the writing of the computer code. But it appears to be impossible to assign a probability distribution to the magnitude of any such undetected errors, so it is impossible to deduce a quantitative estimate of the impact of outright errors on an assessment result.

It is more difficult to decide about the applicability of PSA to the class of uncertainty associated with the possibility of using several alternative models to represent the same phenomena. Conceptual model uncertainty (CMU) was at the centre of the concerns of the PSACOIN Level 2 exercise, but that exercise was not, up to the time of writing, able to draw firm conclusions about the relative importance fo CMU and other classes of uncertainty, nor about the amenability of CMU to treatment within the framework of a PSA calculation. CMU was also the subject of a recent NEA Workshop [ ], but again, the question of the applicability of PSA was not settled.

Sometimes, the existence of a range of applicable models, none of which can be ruled out by the available evidence, can be represented in parametric terms. For instance, for transport of radionuclides by groundwater flowing in a fractured rock, the question arises whether the rock matrix between the fractures is accessible to the radionuclides, giving the possibility of greatly retarding their transport. A model that includes diffusion into the rock matrix and a model based on transport only in the fractures may be considered as alternative models, with some uncertainty as to which is applicable in a particular medium. However, it is perfectly possible to construct a model that represents both fracture transport and rock-matrix diffusion, including one or more parameters that affect the ease of access of solutes from the fractures into the matrix. Assigning PDFs to these parameters could have a very similar effect to attributing probabilities to the two alternative models. Thus, one outlook, in which the analysis was treated as a single model with parameter uncertainty, could be considered practically equivalent to another outlook, in which conceptual model uncertainty was said to exist.

It would thus appear that PSA is capable, from a computational point of view, of handling any type of uncertainty that can be characterised by probabilities. Casting uncertainties in parametric form is not so difficult: even choices between apparently unrelated alternative representations can be cast mathematically in terms of a discrete probability distribution for an integer that selects between the alternatives. The difficulty lies in attributing the probability distributions in a way that is defensible.

Risk dilution

Risk dilution is a term that is sometimes used to describe a situation in which, counter to normal expectations, a reduction in the uncertainty of the input parameters of a model leads to an increase in the mean of an output quantity. When this trend comes about, there is a danger, which may be explained as follows. In assigning PDFs to describe the uncertainty in the parameters, there may be a tendency to overestimate the uncertainty, that is to overestimate the width of the distributions. If this overestimation results in the mean consequences being lowered, the unfortunate effect is that what appears to be a conservative step (overstating the degree of uncertainty) leads to an overoptimistic assessment of the system performance.

One circumstance in which risk dilution is a danger is that in which the consequence (for instance individual dose rate) has a peak in time, and the time of the peak is affected by one or more of the uncertain parameters. Averaging over the possible parameter values amounts to averaging over alternative situations in which the peak in the consequences occurs at different times. At any given time, the mean consequence is less than the consequence value that would apply if the peak occurred at that time. And the wider the distribution of the parameters, the wider the range of peak times, and the lower the mean consequence at any one time.

How to detect the existence of risk dilution, and to guard against unwarranted dilution, is a question that PSAG has not solved.

Time dependence and other complexities of model

Not only are time-dependent models harder to implement than constant ones, there can be a problem of specifying parameter distributions when the parameters can be general functions of time (infinite-dimensional objects). The question of when it is necessary to model time-dependences explicitly, when they can be ignored, and when time-invariant scenarios can be used as a substitute, is not resolved.

Computer power continues to increase, giving the possibility of using more and more complex models. Is this trend a real gain? Will it increase confidence in the results, or tend to obscure the *dominant principles? Is there some principle determining the appropriate level of model complexity for a given level of site characterisation?*

Scenarios versus all-embracing simulation

This paragraph should briefly expose the debate about the relative merits of these two approaches.

7.2 Unresolved Issues in PSA Methodology

Convergence

The question of convergence was referred to briefly in the discussion of statistical sampling schemes in Chapter 4. It is an important issue because the results produced by an application of PSA must be reliable, yet the method by its nature involves repeated calculations using numerical models, and these can consume much computer time, the more so as the models are made more realistic. It appears that in many applications of PSA to date, the sample sizes used have been determined mostly by intuition, operating within constraints of expense, or elapsed time, associated with computer execution time.

Achieving convergence, and having confidence in the fact, may be more difficult in PSA applied to waste disposal systems that in many other applications of Monte Carlo simulation, because of the skewed distributions of output values that seem to arise. If, for instance, a value that is at least 100 times bigger than the mean occurs with a probability of 1 in 1000, a sample size of 100 cases would have only one chance in ten of encountering such an extreme value. Yet if such an extreme value did arise, it would contribute to the estimated mean at least ten times more strongly than it should. The accuracy of the mean estimation would thus be very poor with a sample size of 100; a sample size of several thousand would be more appropriate.

In the PSACOIN Level E exercise, a scheme for testing convergence was examined, based on dividing the sample results up into batches, and looking at the distribution of the means of the batches. The results obtained in this exercise gave no grounds for believing in the efficacy of this test scheme. The difficulty of locating rare events by random searching is often compared with 'looking for a needle in a haystack'. Conducting PSA when the degree of skewness of the distribution of the output quantity of interest is unknown can be compared with exploring a haystack that may or may not contain a needle. The fact that one has not yet been found is no good evidence that none is there.

The PSA method makes it possible to accompany estimates of means or other statistics with a measure of the estimation accuracy. However, such error analysis is based on the variability observed in the sample cases that were examined, and there is always some doubt that the error estimate may be smaller than is really justified. It would appear that to be assured of convergence, some information about the shape of the consequence distribution is required, independent of the evidence provided empirically by the samples taken so far. Automatic 'stopping rules' should therefore be viewed with caution.

It is likely that the number of sample cases required for convergence and for attaining a given estimation accuracy depend on the nature of the output quantity. For instance, the cumulative release to the surface of a parent radionuclide cannot possibly exceed the disposed inventory. There is therefore no question of the upper tail of the distribution extending beyond this limit. For this reason, the mean of a cumulative release may converge much more rapidly than, say, the mean release rate at a certain time. It appears, however, than no systematic studies have yet been made on the relationship between convergence and the choice of performance measure.

The Level S PSACOIN exercise drew particular attention to the fact that measures of parameter sensitivity obtained on the basis of a PSA calculation can be much further from convergence than the means or other statistics calculated as part of uncertainty analysis.

Parameter correlation

The question of correlations among the uncertain parameters of a model to which PSA is applied has only briefly been considered by the PSAG. The matters requiring attention lie in two areas. First, it is difficult to communicate the meaning of correlation where subjective uncertainty is concerned. In the case of two or more stochastic quantities that may be sampled by repeated observations, the joint probability distribution may be interpreted in frequentist terms. If the observation of a large value for one variable is frequently accompanied by the observation of a large value for another variable, the two can be said to have a positive correlation. However, for most of the uncertain parameters of models of the long-term performance of waste disposal systems, there is no question of repeated observations, although some of the parameters may relate to quantities, such as rock properties, that in principle could be measured at different locations or times. Rather, the PDF attributed to each parameter represents a degree of belief on the part of experts (conditioned perhaps by experimental observations of measurable quantities) in the appropriateness of using different values of that parameter in the assessment. Correlation between two parameters then has to be expressed in terms such as the following. If, by some means, the question of the value of one of the parameters could be settled, and the resulting value was at the high end of the distribution expressing current belief about the alternatives, then if this would result in the expert opinion about the value of a second parameter being revised upward, a positive correlation could be said to exist.

The second type of outstanding problem concerning parameter correlations is that of specifying the correlations quantitatively, and implementing the correlations in the Monte Carlo sampling. It may be easy to elicit an expert judgement of the correlation coefficient applying between two parameters, but simply specifying the correlation coefficient is insufficient to specify the full joint PDF, which gives the probability density in the neighbourhood of any given pair of values. Discussions on this subject within the PSAG have revealed that some members have tried a scheme in which tables of sampled parameter values have their entries permuted at random, until the correlation coefficients reach the required values. However, this is an arbitrary and non-unique choice of basis for achieving that degree of correlation.

It would appear preferable, when possible, to discover the physical basis for belief in a correlation between two parameters. Usually, it means that they both depend on another set of variables, involved in the fundamental processes underlying the meaning of the parameters. If those underlying variables can reasonably be attributed independent (uncorrelated) distributions, and if analytic expressions can be stated relating these independent variables to the model parameters, the those parameters will become dependent random quantities that in general will be correlated.

Types of system model

The model systems studies in the PSACOIN exercises, and much of the experience of applying PSA that has been discussed by PSAG members, have been concerned with the groundwater pathway for transport of radionuclides from a repository to the accessible environment. It may be that there are particular problems associated with applying PSA to models of other major pathways, and if so, the PSAG has not explored such problems. The Human Intrusion pathway, for instance, may be dominated by the effects of intrusion events that are relatively short-lived, and occur at unpredictable intervals, whereas groundwater-mediated transport gives rise to long-lasting releases, for which uncertainty in the parameters results in relatively minor variations in timing and magnitude of the consequences. Convergence for PSA applied to human intrusion events may well be much harder to obtain.

7.3 Unresolved Issues in the Application of PSA

Visualisation of model results in the presence of uncertainty

The concept of uncertainty is difficult to communicate, especially when the quantities calculated are functions of space and time. PSAG members have repeatedly expressed the feeling that further advances in the art of communicating such results must be possible, but no substantial advances have been made.

Termination of the data-gathering/assessment cycle

How is PSA affected by, and how can it contribute to termination of, the cycle of assessment, data gathering, re-assessment?

References

ATTACHMENT 7

Uncertainty in Performance Assessment

Peter Robinson
Intera Information Technologies, Environmental Division
Henley-on-Thames, Oxfordshire, UK

Presentation to the 15th NEA PSAG Meeting

Introduction

Safety assessments for radioactive waste disposal proposals must make predictions of the performance of complex systems into the distance future. Inevitably, these assessments must be performed in the presence of various uncertainties. Often these are handled using a probabilistic approach.

It is the purpose of this talk to set out the case for an alternative way of treating uncertainties, which may offer a more defensible and practicable approach. The ideas discussed here are to be presented in a fuller form at the Spectrum '94 meeting in August.

Origins of Uncertainty

It is useful to classify the various sources of uncertainty according to their fundamental origin.

In the first place, we distinguish between variability and uncertainty. Variability in natural systems, e.g. in permeabilities, has to be characterised, and this characterisation will be subject to some uncertainty, but variability itself needs to be included within a model.

Where experimental measurements of a fundamental quantity have been made there will be variability between different results. This variability can have many causes – and expert interpretation is required to understand it in each particular case. It is a mistake to directly equate the variability among experimental results with uncertainty in the parameter required for a particular assessment.

Some processes and events exhibit apparently random behaviour. For example, human actions, or seismic events are not susceptible to precise predictions.

Most commonly the root cause of uncertainty is simply ignorance. Of course, ignorance is rarely total. We may know constraints on possible behaviours, or have indirect experimental knowledge. By bringing together all the information, we can hope to make some statements about each issue that arises.

This process, the application of expert judgement, is all pervading in performance assessments. An assessment calculation of the performance of a waste disposal system will be predicted by a large number of decisions taken by experts. At each decision point many alternative choices could be made. This is the uncertainty that we must address – how do these choices cumulate to affect the final prediction of performance?

Possible approaches

In order to deal with these uncertainties we would like an approach which can treat them all in a consistent and defensible way. Apart from the probabilistic approach, which has been much discussed, a number of potential approaches could be used. In order to keep the discussion from being overly complicated, we focus mainly on the particular issue of parameter uncertainty.

Best Estimates

The simplest, and probably earliest approach used in radioactive waste assessments, is to make a "best estimate" at each decision point. In this approach, the experts are asked to take all available information and make a judgement as to a single value (or model choice etc.) which best characterises the current knowledge. Although the best-estimate approach has simplicity on its side, it fails to indicate the impact of uncertainties and has not been used in large assessments.

Conservative Estimates

A refinement of the best-estimate approach is the conservative approach. In this approach the experts are asked to choose a value (or model) which will lead to an overestimate of the real consequences. The idea is simply that if the conservative estimates can be used and regulatory targets met then all is well. This approach has a number of draw-backs. One is that it is by no means always clear what a conservative value for a parameter is. Another problem is that it is all too easy for conservatism to turn to pessimism and for everything to be ignored or discounted. Then the safety case is not made and the approach has defeated itself.

Nested Sets

The approach which we put forward here builds on the ideas of a conservative approach, but has features in common with a probabilistic approach.

We start by asking what it is reasonable to ask of experts. It is our view that experts can be expected to classify the possible answers that they could give.

As a start, we require the experts to judge which answers can be said to be uncontroversial, i.e. a central range of possibilities which could not be ruled out. This could apply to choices of processes to include as much as to parameter values. We then calculate the consequences for any combination of such possibilities.

Next, we would consider a wider range of possibilities for each decision, characterised as "possible". Again all combinations would be calculated.

This can be extended to more and more speculative possibilities as desired, with the levels being chosen to suit the assessment objectives.

Let us consider what we have done in the above procedure. For each choice we have nested sets of possibilities – becoming more speculative as they grow. By undertaking calculations we derive similar nested sets of consequences. Given a large set of calculations, the category of a consequence is the maximum over all sets of choices leading to it, whilst the category for each such set is the minimum of all categories of its elements.

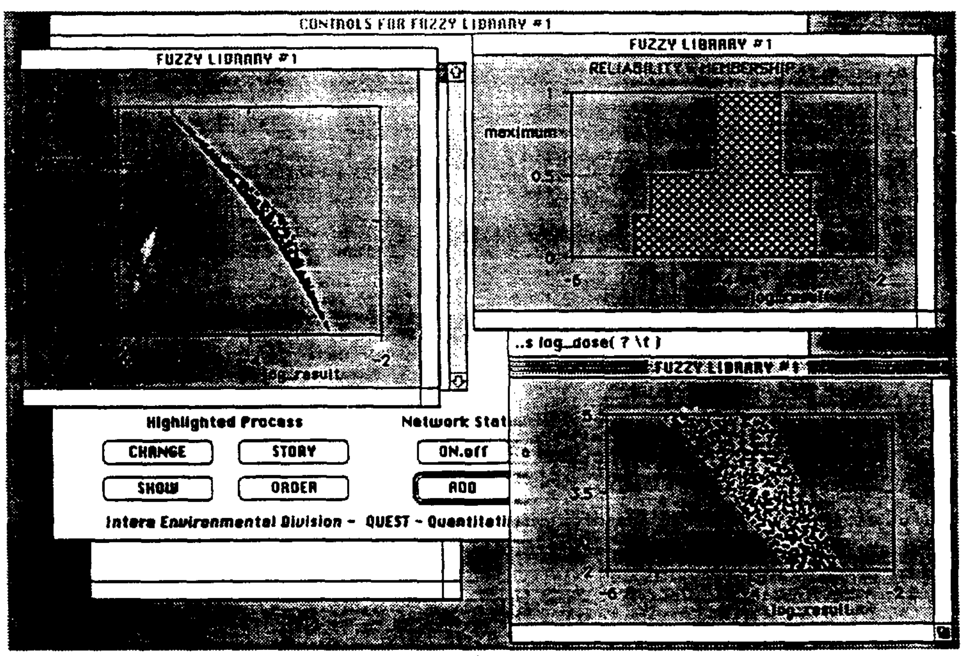
In fact, the categories correspond to membership function values for fuzzy sets and the min-max combination rules are precisely those commonly used in fuzzy set analysis.

Practicalities

In implementing the nested set or fuzzy approach a number of practical issues must be addressed. We make a number of observations below on some of these.

Calculational Procedures

As part of a recently completed project jointly funded by the European Union and UK Nirex Ltd, we have developed a tool to undertake a fuzzy or probabilistic analysis of parameter uncertainties. The QUEST application uses Monte Carlo sampling and the min-max combination rules to produce output membership functions. The fact that large numbers of samples are used allows sensitivity analysis to be undertaken, e.g. by taking scatter plots of various inputs against key outputs. A typical QUEST output is shown below.

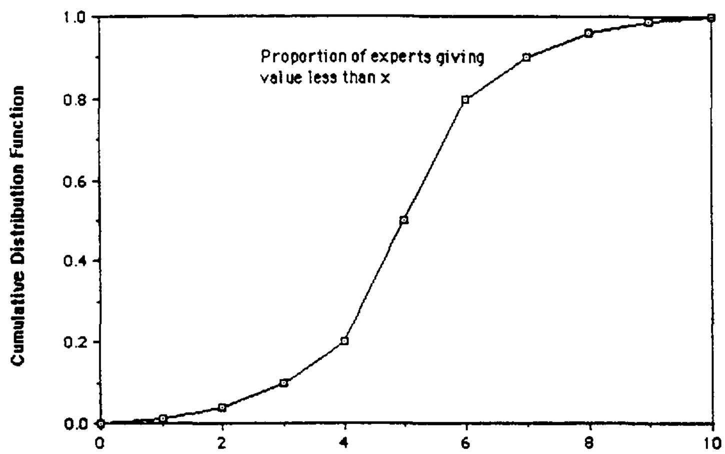


Elicitation

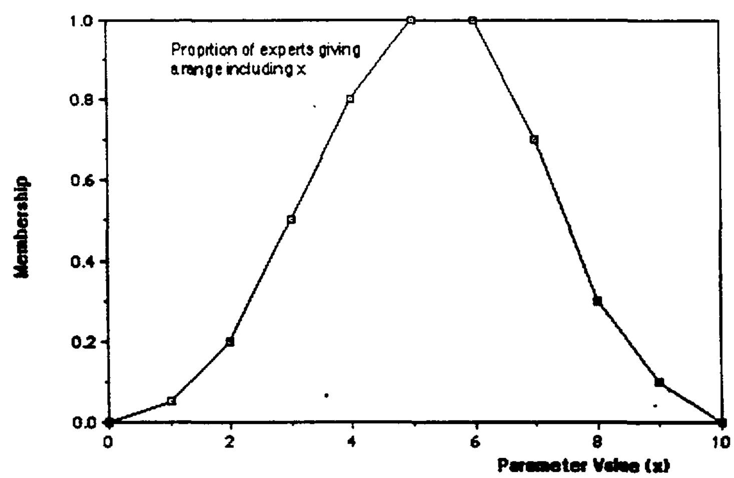
The nested-set approach does not solve the difficulties of eliciting information from experts. In order to be consistent across all uncertainties the definitions of each membership category needs to be carefully described.

The elicitation process can be viewed in terms of voting by a large group of experts, as shown in the following figures.

Expert Elicitation of Probabilities



Expert Elicitation of Membership



Conclusions

- The nested set approach provides a practicable method of treating uncertainties.
- Monte-Carlo sampling can be used to implement the approach.
- The inputs required as less detailed than the probabilistic approach.
- Outputs are different to the probabilistic approach, for example showing what could reasonably happen, or how unreasonable one must be to give an unacceptable consequence.
- A different regulatory view would be needed to interpret nested set results.

ATTACHMENT 8