SKI Report 95:17

Uncertainties in Safety Analysis

A Literature Review

Christian Ekberg

May 1995

ISSN 1104-1374
ISRN SKI-R--95/17--SE

SKi
STATENS KÄRNKRAFTINSPEKTION
Swedish Nuclear Power Inspectorate

SKI Report 95:17

Uncertainties in Safety Analysis

A Literature Review

Christian Ekberg

Department of Nuclear Chemistry, Chalmers University
of Technology, S-412 96 GÖTEBORG

May 1995

Summary

In the construction of an underground repository for spent nuclear fuel, there are many factors to take into account relating to many various disciplines, e.g. geology, rock mechanics, hydrology and geochemistry. For performance assessment of the repository, mathematical models are used. In the use of these models, uncertainties of various types may occur. In an attempt to categorise the uncertainties the following list may be useful.

1. Future states of the disposal system, (scenario uncertainties)
2. Models used to simulate these future states, (conceptual uncertainties)
3. Data and parameters used in the modelling effort

The first group, relating to the future states of the disposal system, comprises such events as ice ages, continental drift and human intrusion. The effects associated with such events have to be estimated. One of the sources of knowledge of the geological future lies in the investigation of the past. This may in fact be the strongest key to the future when combined with proper models.

The models used in the modelling effort may be divided into several parts:

1. Waste package and fuel dissolution models
2. Repository models
3. Ground water flow and radionuclide transport models
4. Biosphere models

There is also the possibility to incorporate all these parts into one large model or model package, e.g. in probabilistic modelling of the whole system, but usually a lot of simplifications have then to be made.

Computer codes need data, as input, to work. These data are usually subjected to errors and uncertainties of varying magnitude and origin. It is, however, possible to investigate the propagation of uncertainties from input data to results. One common approach is to investigate these effects with some kind of Monte Carlo sampling technique followed by a statistical evaluation, giving for example a distribution function of the output when the input is varied. This function may be used to evaluate a confidence interval for the model result. The uncertainties may be propagated through the program chain, from the fuel to the biosphere, by using the confidence interval from the preceding simulation as an input to the next, thus giving a more complete uncertainty analysis for the whole problem.

The results from a model may also be investigated by a sensitivity analysis. Such an analysis is usually performed in order to reduce the number of variables being investigated in an uncertainty analysis. However, a sensitivity analysis may also be used to calculate coefficients for some expression describing the simulated process.

The purpose of the presented work has been to give a short summary of the origins of many uncertainties arising in the designing and performance assessment of a repository for spent nuclear fuel. Some different methods to treat these uncertainties is also included. The methods and conclusions are in many cases general in the sense that they are applicable to many other disciplines were simulations are used.

As a conclusion it may be noted that uncertainties of different origin have been discussed and debated, but one large group, e.g. computer simulations, were the methods to make a more explicit investigation exists, have not been investigated in a satisfying way.

Contents

Contents ctd.

1. INTRODUCTION

In Sweden, about half of the electrical power is generated form nuclear facilities. This energy is produced by twelve nuclear reactors with a consumption of nuclear fuel at a rate of ~250 tons per year. The current political aim is to have no nuclear reactors operating after the year 2010. By this time and with current energy production the amount of spent fuel will be 7000-8000 tons.

World-wide there are several ideas about dispose of this spent nuclear fuel. The one accepted in Sweden is to emplace it in the bedrock several hundred meters below the ground surface. According to present plans the fuel will be stored in copper or copper/steel canisters which are placed in cylindrical holes drilled in tunnel floors in the rock. The remaining space in the holes are then filled with compacted sodium bentonite clay. The tunnels and shafts in the rock are to be filled with a mixture of sand and sodium bentonite clay [KBS-3] [SKB 1].

In connection with a safety analysis of such a repository there are many factors to take into account, e.g. hydrology, geology, rock mechanics and geochemistry. Since the repository is supposed to function about hundred thousand years without maintenance many uncertainties arise in the investigation of each concerned discipline. From the geological point of view, such an uncertainty is the behavoiur of the rock and thus the repository in the case of a glaciation. A characterisation of the chemical environment is of importance for the prediction of canister corrosion as well as for prediction of the transport of released radionuclides from a failed canister.

In order to predict some of the changes in the vicinity of the repository, computer simulations have to be made. These simulations may range from climate modelling to simulations of the dissolution of the spent fuel.

The computer simulations are, however, often subjected to uncertainties of different origin. It is interesting to know which variables change the result signifficantily when they are varied. Therefore a sensitivity analysis is often made together with an uncertainty analysis. The purpose of the latter usually is to investigate the effect of different constellations of indata and parameters. In this paper some different origins for uncertainties and a short review of different uncertainty and sensitivity techniques, are described.

2. Overview of uncertainties

Uncertainty analyses are one of the contributors to information needed in order to ensure that the wastes from nuclear power plants are stored safely for the time needed to render it harmless in the sense of radiation. Decisions have to be taken at various stages during the decision / construction process. Some of the crucial questions may be [Johnston 1987]:

How to select a site?
How to investigate a site?
How to design a repository?
Whether to authorise its construction and operation?

In many cases computer simulations are made to visualise and investigate properties connected with the repository. It should, however, be recognised that the calculated results are by no means a prediction of what really will happen, i.e. all such calculations are encumbered with uncertainties. There are several ways to group these uncertainties. Once grouped, it may prove difficult to determine where the borders between different groups should be drawn, but it is necessary to do this in order to get a good overview and to be sure that as may as possible uncertainties are taken into account. The more important sources for uncertainties may be grouped in the following way [Cranwell 1987, 1]:

1. Future states of the disposal system [including its environment]
2. Models used to simulate these future states
3. Data and parameters used in the modelling effort

Future states of the disposal systems are generally referred to as "scenarios". Within this group are events like earthquakes, climate changes and human intrusions.

Uncertainties associated with the models used to simulate the future states of the repository arise mainly from the fact that not all mathematical models are good representations of reality. In addition to that, they may be used incorrectly in a computer program. One way of using a model incorrectly in a computer program is to extend the calculations to a region where the model no longer is valid.

The uncertainties associated with data and parameters are probably the most easily quantifiable ones since they may be derived from error propagation analysis or some statistical method used on the results from a computer simulation. There exist some other reviews of uncertainties which might be interesting, e.g. [SKI 2].

3. Future states of the disposal system.

One of the most complex contributors to the total uncertainty concerning the behaviour of a disposal system is the question of future events in the site or it's surroundings. The identification of these events is not only vital in the safety evaluation of a repository but may also work as a guide to data collection. The events may be classified in the following way [INTERA 1].

1. Naturally occuring geologic events.
2. Events caused by the actions of humans.
3. Events caused by the repository system.

It is easy to find unreasonably many events that may occur in the future. In addition to that, it is impossible to take any action against the main part of these. This statement stresses the importance of discussions on likelihood and severity of each event.

Many approaches to solve this problem exist, e.g. [Cranwell 1982], [Cranwell 1987, 2]. One of them is to try quantifying the effects of a certain event. These effects may then give a new event and so on. After each event the effect may be the subject for computer simulations and speculations. If the event is placed in a box, the calculated of predicted effects may then join different boxes. This visualisation may be performed in a number of ways, for example some kind of tree structure [SKI 1], but the branches of the tree will mingle to the point of undistinguishability. However, the tree approach has been argued not to be useful for analysing geologic processes or their int' actions [Burkholder 1981].

It is vital to remember that in most cases it is impossible to foresee the future, but in the case of nuclear fuel repositories, it is essential that an attempt is made. The length of time discussed in these scenarios are more far off in the future than the bronze-age man are in the past and therefore history may not be enough to try to mirror the future.

3.1 Geological events.

The possibility for many natural geological events to occur may depend on the choice of place for the repository. For example this holds for earthquakes and volcanic eruptions. Most other processes of changing the properties of the rock barrier are very slow, therefore these uncertainties may be divided into different intervals: present, at about 10000 years and beyond 100000 years in the future [INTERA 1].

3.1.1 The present situation.

At present, most uncertainties lie in parameters and modelling. These may include spatial variation of parameters and choice of computer model. The uncertainties thus introduced may be found in separate Sections.

3.1.2 About 10 000 years into the future.

For Swedish conditions, one very likely event is the occurrence of a new ice-age, i.e. a time when the surface of the repository is covered with thick land-ice. Some effects of this event are more or less possible to foresee, for example formation of pluvial lakes, stresses due to ice load and changes in sea-level. On the other hand it is almost impossible to predict the hydraulic conductivity and the flow paths in the rock during or after a glaciation. These effects may be reason enough to place the repository at a place where no glaciation may occur [INTERA 1]. However, this only applies to countries were such a choice exists. Since every country is responsible for its own radioactive waste, some countries, for example Sweden, do not have the possibility to place the repository where a glaciation can be avoided.

3.1.3 About 100 000 years into the future and beyond.

At this length of time many events are possible. There may be meteor downfall, volcanic eruptions, earth-quakes, landslides and ice ages [Campbell 1978]. The only method available for prediction of these events, is the past. It is however argued that ice ages may be determined by astronomical calculations. Stretches of the rock and continental drift are in many cases possible to foresee, but the effect of these events on the repository and the surrounding rock may be vast or none existing [INTERA 1]. In anyway, the effort should lie on determine to which extent an event calls for extra precautions or not. An investigation may indicate that the effects of these events are small in comparison to other during this time interval.

3.2 Events induced by humans.

The repository may be breached for many reasons. The most likely are perhaps the desire for the metal deposits in the fuel or ores close to the repository. Both of these events are probably due to some kind of collapse of the society in that country or on earth as a whole [SKI 3]. Otherwise these actions may be prevented or at least the explorers may be warned to avoid some regions. However, the collapse may not be

necessary if the want for the spent fuel is greater than the knowledge of how to retrieve it safely.

To foresee what happens if the repository is intentionally violated is almost impossible since the effects depend on the technological level of that time. It is however possible to construct the repository in a way that will make it difficult to breach it at present time or in the near future. There is, however, a severe drawback with this possibility. A safer technique for either transforming or disposing the waste may be discovered a short time after the closure of the repository, but a good technique to open it may not yet exist.

The event of unintentional breaching of the repository is more easily handled. The recommended approach is to use a conservative direct-release analysis combined with expert opinions [INTERA 1]. However, most cases of unintentional violation probably originate from mining explorations. These may to some extent be avoided by placing the repository in geological uninteresting place.

3.3 Features, processes caused by the repository (system uncertainties).

As an alien part of the rock, the repository may interact in several ways with the host rock. As a whole, these interactions and their effects are more of a continuing process than an event although they still need a lot of investigations. However, many of these processes are already quantified and analysed. Next to the question of whether it is mechanicaliy possible to build the repository, is how to build it to achieve as low as possible negative interaction with the rock itself. Unfortunately there still exits some uncertainties. For example how the pressure will be compensated with the swelling pressure of the bentonite and whether it is necessary to use concrete in the construction of the HLW repository. Other uncertainties may arise if another concept for the storage is selected, e.g. if the repository is built in bedded salt [Bingham 1979].

4. Modelling the future events and states.

If an event is deemed likely to occur, it is important to try to estimate its effect. This is often made by computer simulations. The simulations are, however, often subjected to the drawback of no or very few data to use as input into the model. It is also vital to remember that no simulation is true i.e. it is not completely simulating what is actually happening. Further it may be noted that in some cases the effect of a change in only one input parameter may change the result considerably [SKI 4]. Therefore it may be necessary to make a sensitivity analysis of a computer model in order to determine a ranking list of the variables. In order to discuss models, a few definitions are necessary [Nordstrom 1992].

model: a hypothesis, theory or a combination of theories that provides new insight or a new interpretation of an old problem.

chemical model: a theoretical construct that permits the calculation of physiochemical properties of substances (thermodynamic, kinetic and quantum mechanic properties).

geochemical (or hydrogeochemical) model: a chemical model developed for geologic systems (water rock interactions), includes natural solutes and mineral species.

verification: a test of a model to see that it calculates what it is supposed to calculate, i.e. to determine that the algorithm and code are correctly written.

validation: a test of a model to see that it reproduces experimental laboratory measurements (field measurements are not laboratory experiments).

One may argue that the verification is rather a test of how a model is incorporated in a computer program. The modelling of the performance of a repository for spent nuclear fuel may be divided according to Figure 2 [Campbell 1988].

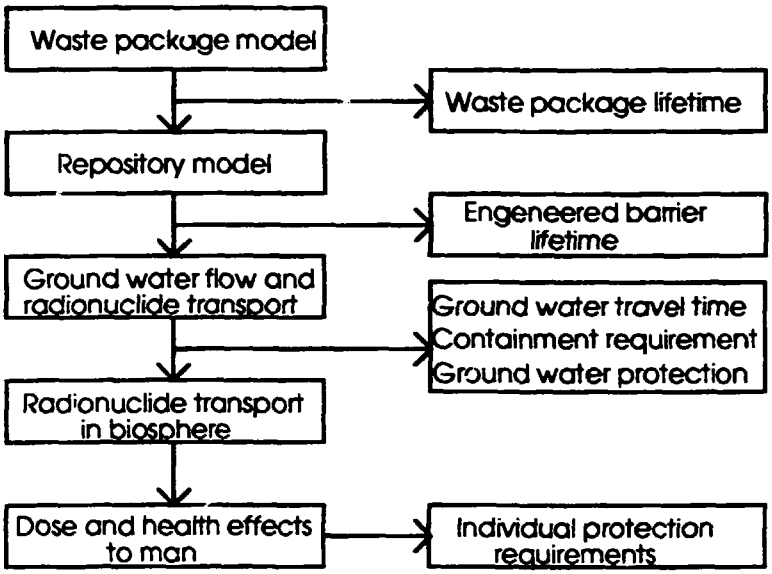

Figure 2. Illustration of consequence modelling sequence.

Each of the boxes in the left column of Figure 2 represents complex systems that may be modelled with more or less simplifications. The right column shows what the expected results of the modelling are. It should be noted that the results from the "Groundwater flow and radionuclide transport" box, is a requirement in USA and not in Sweden. Some programs are taking the whole Figure 2 into account, but in those cases with a lot of simplifications. The different models are further discussed below.

There is also the problem of validating the results from the models. In some cases it is impossible to validate a model [Nordstrom 1992] and therefore it is sometimes more important to increase the understanding of the system being modelled, by for example sensitivity analysis, rather than trying to validate against field data.

4.1 Waste package model

A very common and thoroughly examined event in the investigation for a nuclear waste repository is the "canister failure" event. One reason for this is that the probability for a canister failure approaches one as the time increases. The computer simulations of this event may be done in a number of ways. The simplest way is to calculate how long a given thickness of the canister may withstand corrosion [Kurth 1986][Sutcliffe 1984]. This approach may also include some uncertainties for example the amounts of Cl and O_2 in the water [Song 1989]. Another example is to use some kind of probability for the canister to fail, not considering why.

If the canister fails and water reaches the fuel, this will start to dissolve. Since among other things the composition and spatial distribution of the elements in the fuel

are not well known, modelling is not easy. Attempts have been made but it has been difficult to validate the models with experiments [Börjesson 1994], [Shoesmith 1992].

4.2 Repository models

The repository models are mainly concerned with the underground facility, which consist of for example backfill materials and the underground structure. The main inputs to these models are the concentrations from the waste package models. The main output from the repository models is the rate of radio nuclide escape to the circulating ground water. Naturally this only holds for codes associated with chemistry calculations. There are many other disciplines that repository models consist of, for example rock mechanical and hydrological models.

The simplest chemical repository models only use some equation to relate the incoming concentrations to the released ones. These are sometimes incorporated into a complete model as described in Section 5.5. However, a more complicated code may simulate fluid movement and radio nuclide transport through the repository by using for example hydraulic and sorption properties of the backfill material, the geometry of the mined area and the solubilities of the released radionuclides. These models are very complicated and are almost impossible to validate unless they are constructed of simpler parts which may be validated separately.

4.3 Ground water flow and radionuclide transport

The ground water flow and radionuclide migration codes, e.g. CALIBRE [SKI 5] and CRYSTAL [SKI 6], are using the result from the repository models to calculate the transport of radionuclides, by the ground water, from the repository to the environment. Such codes exist in several categories depending on which phenomena are deemed to be more important. Some programs concentrate on the transport problem and use analytical solutions for the transport [INTERA 2]. A more realistic approach couples the transport with chemical reactions and also calculates the sorption on the host rock. Such a code usually performs the calculations in two or three dimensions. Another approach is to concentrate on other parameters, for example how fluid flow, heat, brine transport influences on radionuclide transport [Reeves 1986]. These codes are however usually very slow in producing their results due to the complexity of the simulated system.

The output of the transport models is usually either expressed as concentrations in the ground water or as time-dependent discharge rates. The

discharge rates may be integrated to give the cumulative releases of radionuclides to be compared with the governmental approved release, and the ground water concentrations may be directly compared with the regulated amount. The latter statement applies only to USA since there is no such regulations in Sweden.

4.4 Radionuclide transport in biosphere and dose and health effects to man

In order to meet the regulated doses to man from a repository it is necessary to simulate the radionuclide transport at the surface and how the nuclides may effect biological life. The result from the ground water transport model is used in a biosphere transport model to simulate migration through the surface layers and estimate the concentrations in the atmosphere, surface waters and soil [Helton 1981]. The total transport rate of elements must be linked to both biological phenomena such as photosynthesis, ingestion and excision, and physical phenomena such as diffusion and sorption. Thus both approaches are needed in a biosphere model.

One approach to build a biosphere model is to base it on several subsystems in which the ecological processes are supposed to balance each other and thus the transport through the system may be reduced to a simple net flow, as used in BIOPATH [Bergström 1983]. Such compartments may be for example: soil, superficial groundwater, atmosphere and biota. The concentrations achieved in the different compartments may be used to predict the intake through inhalation, ingestion and external exposure. These intakes may then be converted into dose levels received by organs in the body. However, not enough data and understanding of the participating processes exist to make correct estimations of the intake as a function of the surface concentrations [Campbell 1978].

4.5 Combined models.

It is possible to make one total model that will predict the dose to humans on the earth after, for example, a canister failure [Ahn 1990]. The main problems with these models are that a lot of simplifications of the system are made. The results from these simulations may or may not be approved by the expert opinion, but the existence of codes that simulate the whole transport of radionuclides from the repository to man, is vital for the understanding of the sensitivity for the different parts of nuclide release to the surface since such an analysis may point out the main contributors to a high dose on the surface. There exist computer codes that investigate the sensitivity and uncertainty in the complete release codes, e.g. SYVAC [Wuschke 1981]. The SYVAC code consists of three submodels representing the three major constituents of

the disposal system: the vault, the geosphere and the biosphere. The submodels are treated in sequence and an estimated dose to an individual may be calculated for each scenario
[Shemilt 1989]. Such an approach gives further important information of the range of effects from different events and are therefore of great importance.

5. Evaluation of models.

If there exists a possibility to verify and validate a computer model it is important that these tests are made in order to ascertain the correctness of the chosen approach. Unfortunately this is not always done properly, which may result in severe misjudgements when designing a repository.

5.1 Validation

Validation of a model is essentially performed to examine if experimental data can be reasonably reproduced. It is desirably that a validation of a model is made before it is put to any practical use. However, it is possible to make this validation in a number of ways, each giving different results to the question of whether experimental data are matched or not. There exist some attempts to make validation of models, for example the CHEMVAL project [Read 1991] and [INTRAVAL]. The latter is mainly concerned with transport of radioactive substances in the geosphere, both experimental and modelled, and the former is more of a calculation exercise where the results from different computer codes are compared. It is noteworthy that one conclusion of the INTRAVAL project is that most of the present performance assessment migration models may be over conservative in their result.

There are several parameters that should be chosen for the validation of a model. Some are perhaps not possible to use, but as many as possible ought to be taken into account. Some examples, from the geochemical point of view, are mineral solubility as function of solute concentration and mean activity coefficients. It is the opinion of [Nordstrom 1992] that only the major ions are to be taken into account in a validation since a small error in the distribution of the major elements may cause a large error in the distribution of some trace elements. The values chosen for validation should not be field measurements but rather laboratory experiments [Nordstrom 1992], since it is possible, in a laboratory, to keep the major variables controlled. Some models e.g. geochemical models, are almost impossible to validate since we do not have a complete knowledge of any hydrogeochemical system, only approximations. Therefore, the only method of validation is in this case invalidation which may be a very powerful tool.

The lack of organised validation of models seems to apply to most disciplines and great effort should be used to construct validation strategies that most codes should go through before use in any real simulation.

5.2 Verification

The verification of a model is mainly made to confirm that the computer code functions correctly. This includes the choise of the mathematical expressions and data bases used. It is also possible to examine the model with respect to the sensitivity for changes in the input data. Therefore simulations with various data bases, were the results are compared, are needed. However, it is essential to remember that different databases may have the same origin and therefore are equipped with the same faults. There exist no general method to verify computer codes and therefore it must be the responsibility of the programmer to verify the model in the best way possible. There exists a suggestion on how this should be done by [Nordstrom 1992]. However, since this procedure is not wholly general there will still exist custom made verification processes the reliability of which experts may argue.

6. Data and parameter uncertainties

Data and parameter uncertainties are perhaps the type that is most easily quantifiable. These uncertainties may arise from several sources, such as [Cranwell 1987, 2]:

1. measurement errors
2. paucity of data
3. misinterpretation of data
4. spatial variation of parameters
5. assumptions regarding behaviour of the system, [conceptual uncertainties]

6.1 Measurement errors and misinterpretation of data.

A lack of precision in the use of an equation may be due to either measurement errors or misinterpretation of data. It may be very difficult to determine which. Consider the following example [Moffat 1988]: In some model, the difference between two bulk temperatures is needed. If a very sensitive probe is inserted in the bulk, and the temperature is measured, the measured temperature is probably not the bulk average temperature that was needed in the equation. The main cause of this problem is that there is no clear definition of which values to be use. The result may therefore be more or less inaccurate. It is most vital to understand what is needed in the equations and how to measure it.

Pure measurement errors may be found by re calibration of the instrument. This calibration may, however, itself be incorrect. It is sometimes easy to detect the possible uncertainties in the calibration since usually new measurements are introduced. Another case of "measurement errors" is when the result from the measurement is not correctly read or transmitted, for example if a pH-meter shows a pH of 4.5 and the experimentalist writes down 4.7.

Errors which origin lies in misinterpretation of which data to be introduced into an equation, are often difficult to locate. In most cases, it may not be possible to detect these errors, unless a clearly impossible answer has been obtained. The responsibility is therefore taken by the experimentalist who is measuring the values. It is vital that experimentalists consider, what they have measured, very thoroughly. There may be many small factors that will make the measurement incorrect. To use the obtained data without knowing how they were measured may be fatal. As an example, a correct measurement of the temperature of a gas flow with probes, may have the following form [Moffat 1988]:

$$T_{gas} = T_{probe} + C_{rad} + C_{vel} + C_{cond} \qquad (1)$$

where the C's are correction therms for radiation, velocity and conduction errors. It may be seen from equation 1 that if the temperature was measured with a thermometer the values obtained may not be used in an equation that needs the gas temperature, (T_{gas}), and not the probe temperature (T_{probe}), thus giving the wrong result.

The processes from the true value to the correct value through measurement errors and misinterpretation of data may be seen as a chain with clearly distinguishable nodes. A more detailed figure may be found in [Moffat 1973].

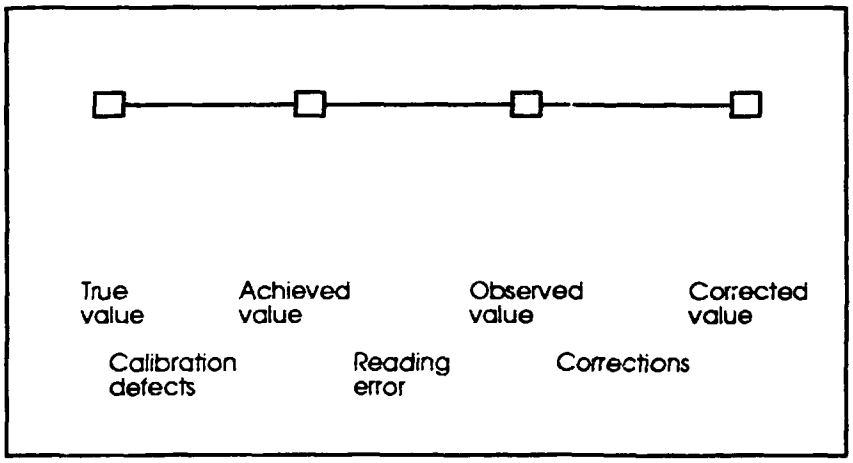

Figure 1. The Measurement Chain

It may be seen from Figure 1 that there are many things to take into account when drawing conclusions from experimental data. However, if all data are regarded with a certain amount of suspicion, the effect of these errors could be reduced significantly.

6.2 Paucity of data

Paucity of data is perhaps the most common source for lack of precision in the results of many models since many vital parameters may be missing. In some cases there are not enough measurements made to make a good estimate of the true values. In some cases a closer examination of the available data might reveal that the introduced uncertainty is of minor importance.

In other cases however, there may be no data available at all. Then there are very few possibilities open for a modellist to calculate a result. The most common way to solve this problem may be to perform a sensitivity analysis of the model, thus determine whether the concerned parameter is important or not. If this approach is not possible, a common method may be to get an expert's opinion on the subject. This may

result in great uncertainties in the result of a model. If the result is clearly wrong it may be wise to reconsider the parameter estimated by the expert first. In many cases the result is not obviously wrong though the parameter estimation may be wrong. In those cases may be almost impossible to find and correct a value that may be of great importance in later simulations.

6.3 Spatial variation of data

Spatial variation of data is perhaps one of the most important contributants to uncertainties in the determination of the geology for a disposal system. In a rock, the minerals may appear almost at random and thus making accurate predictions of their distribution almost impossible since taking samples from the rock changes its properties dramatically, e.g. creating new fractures or dilute the groundwater with drilling water. It is possible, however, to argue that the water only comes in contact with the fracture filling minerals, but it is impossible to state this in any given circumstance, and thus the entire composition of the rock may be important. This may be modelled in a number of ways, e.g. with the minerals distributed at random over the fracture surface [Emrén 1991].

Another example of spatial variations in the rock is the hydraulic parameters such as conductivity and porosity [Gelhar 1976]. These variations are propagated to the ground water flow velocity through the use of Darcy's law and are therefore of great interest. It may be noted that such parameters as the conductivity may, due to its spatial variability, be treated as stochastic parameters [Cranwell 1987, 2]. This is usually not the case since it is easier to use an average. This may, however, induce a large error in the calculations.

Spent nuclear fuel consists of many different species situated in different parts of the fuel pellets. Therefore composition of the fuel may also depend largely on where in the fuel pellets the measurement is made. It is in this case possible to use some kind of an average value, since the variation of the composition is in a very small volume compared to the area exposed to leaching. This, however, does not apply to the distribution of the minerals in the rock.

Some changes in time also belong to this section since they are fast enough to appear to be changes in space rather than time. An example may be the chemical fluctuations in an aqueous system [Schecher 1988]. The measurement of these fluctuations may be an example of misinterpretation errors, as described in section 2.1, if the phenomenon is not known to the experimentalist. The result may be fatally wrong if changes in time are treated as spatial changes.

6.4 Assumption regarding the behaviour of the system

If no or very few experiments are made it may be necessary to guess how a system will behave. This guess is often based on some assumptions that may, or may not, be correct, for example the temperature gradients from the repository to the surrounding rock may give raise to convective flows in the cracks of the rock.

If the geology, hydrology and chemistry for a repository are well determined and as many iteractions as possible are investegated, the assumptions regarding the system should not be major uncertainties, but some parameters may still be guessed rather than measured. However, it is important to bear the uncertainties associated with assumptions regarding the system in mind since some of them may still be used. Unfortunately it may be that their origin has been forgotten and they are now treated as facts.

7. Treatment of data and parameter uncertainties

There are several methods for treating data and parameter uncertainties. The most commonly used are probably among the following.

1. statistical methods
2. stochastic models
3. interpolation techniques
4. differential analysis techniques

7.1 Statistical methods

Statistical methods are perhaps the best developed and most widely used of the techniques for treatment of uncertainties in data. Most of the experimental applications involves a random error that influences the experimentally determined data. These errors may be seen as different results from the same set of experimental conditions. The errors may be estimated by repetition of the experiment under the same conditions. However, this is not the case with most computer codes when an experiment is simulated. Since a computer code is based on a theoretical model of what happens in the experiment and the models usually are deterministic, i.e. a given set of input variables will always give the same output value. Therefore in performing an uncertainty analysis with statistical methods, the design chosen must not include replication [Harper 1983].

Statistical methods may be divided into two different subgroups.

1. experimental design methods
2. sampling methods

7.1.1 Experimental design methods

In experimental design methods, the main concept is to use a specific design to select a specific set of input variables and their mutual order. A typical such method is the factorial experimental design (see Appendix I). These designs vary all the input variables at the same time, which makes it different from the old "one-factor-at-a-time" designs. The efficiency of factorial designs, in the estimation of interactions between the input variables, is well documented and such designs are for example implemented in many optimisation calculations. This approach has also been used in connection with repository calculations [SKI 7], however only for a rather small system. The factorial experimental design makes use of all data in the estimation of the effect of each input variable. It is however clear, as the number of input variables increases, that the number of computer runs needed will increase rapidly. One way to eliminate this

problem is to assume that all high order parameter dependencies are zero. This will only give the main effects of the input variables, but those are usually the only ones of interest. There is however the possibility to use iterated fractional factorial design [Andres 1993]. This method is based on the making of a small number of groups and then assigning each variable randomly to one of the groups. This procedure is repeated a number of times, e.g. 7, and then the result may be evaluated. The main purpose of using iterated fractional factorial design is to reduce the number of iterations without loosing too much information. Experimental design methods are better suited for sensitivity analysis than uncertainty analysis. A sensitivity analysis may be of great interest since the number of variables for an uncertainty analysis may thereby be reduced to the important ones, see Section 7.

7.1.2. Sampling methods

Sampling methods are based on treating the model input parameters as random variables with a given probability distribution and, if necessary, correlations. The parameters are then chosen by the use of some sampling procedure. The model is executed for each set of inputs and the result from each set may be used to calculate the distribution function for the output.

There are three more or less different methods to make the sampling procedure: simple Monte Carlo- (MC), stratified- and Latin Hyper Cube (LHS) [McKay 1979] sampling. There are several reasons for preferring some kind of random sampling [Iman 1987].

1. If properly done MC methods can be designed to avoid some inputs that are impossible to model.
2. The MC approach varies all input parameters simultaneously, thoroughly exploring the input space and can be made very efficient.
3. If the probability distributions assigned to the inputs are meaningful, then statistical estimates of output quantiles, means and variances can be made.

There exist many computer programs that uses this approach, for example the PRISM program package [Gardner 1983] and [SKI 8]. The former is a program package that consists of three main parts: first the input parameters are given together with some significant statistics and from these several sets of input vectors are made; second, the calculation code is run for each set of variables and the results are collected to a file; third, the results are evaluated. Clearly the PRISM package follows the outlining of a program based on sampling methods in such a clear way that it may be seen as a very good example of sampling method implementation.

Since it may be difficult to cover the complete variation space with simple MC, it may be argued that in order to draw any conclusions the number of samples must be greater than what is usually used.

The three different cases of MC models will be discussed in more detail in Appendix II However, the conclusion is that the most effective is (LHS).

The sampling methods are a simple way to investigate the effect of uncertainties in the input of a model, but there is one major drawback. It takes many computer runs to make the result statistically significant. If stratified- or LH sampling is used, the probability distribution functions (p.d.f.) for each input variable are needed. In most cases those functions are not known, but it is possible to approximate them with acceptable accuracy. The main advantage of the MC based simulations is that the result may be analysed with common statistical tools.

7.2 Stochastic Models

Usually "stochastic model" means that the input values can be treated as stochastic variables with a mean and a variance round this mean. The approach then is to solve the equations two times, one for the mean value and one for the variance. These methods are often used for prediction of ground water flow. The main variation often is in parameters like hydraulic conductivity and porosity. Some codes for stochastic modelling of contaminant transport in a one-dimensional flow system, have been developed [Bonano 1987].

7.3 Interpolation Techniques

Interpolation techniques are often used to estimate a complete surface structure from spatially distributed data. One such technique is kriging as described by [Matheron 1969]. The main principle of this method is that observation records come from realisations of some random function. The aim then is to construct linear estemators that are unbiased and have a minimum variance. This technique is mostly used by mining engineers and geologists. However, interpolation techniques may perhaps also have some applications to the calculations for disposal of nuclear waste.

7.4 Differential analysis techniques.

The differential analysis techniques are mainly based on a Taylor expansion and the associated partial derivatives [Iman 1985]. The general idea is to treat the dependent variable of interest as a function f of the independent variables x_1.....x_k. This function

is then expanded as a Taylor series about some vector $X_o = (x_{1o}.....x_{ko})$ of base case values for the variables $X = (x_1.....x_k)$.

$$f(X) = f(X_o) + \sum_j \frac{df(X_o)}{dx_j}(x_j - x_{jo}) \qquad (2)$$

This Taylor series may be expanded to include terms with higher order derivatives but is usually truncated after the first or second order of derivatives. The expansion made in Equation 2 generates a linear model as described in Equation 3.

$$Y = \beta_0 + \sum_j \beta_j x_j \qquad (3)$$

This model may then be used for uncertainty or sensitivity analysis, i.e. calculating only the effects on the linear equation.

8. Sensitivity analysis

If a computer model is used to simulate some complex process it is of great interest for both the programmer and the user to make a sensitivity analysis of the system and the program. The sensitivity analysis is also important for verification of the model. Usually a sensitivity analysis means that the change in the result due to changes in the input is investigated. Such an investigation has several important benefits [Iman 1978]. First, the intuition of the programmer may be tested. Second, unimportant variables or unnecessary model complexity may be revealed. Third, some of the input data may be ranked with respect to their influence on the result. As a result of the second and third advantage the number of variables used for an uncertainty analysis may decrease significantly from the original number. Therefore the uncertainty analysis may be performed more rapidly to save computer time for complex codes.

There are many ways to perform a sensitivity analysis but some of the most important are described below.

8.1 Response surface methods

Response surface replacement for computer simulations is usually based on some experimental design to select input values for the code which generates a result. The method of least squares is then used to estimate the parameters in a linear equation, see Equation 3. This equation is generally known as a fitted response surface and this surface is used as a replacement for the computer model. Linear models are usually expressed with an error term added to represent stochastic variation. Computer models, however, usually produce deterministic output and therefore differences between the linear equation and the computer model may rather be due to lack of fit than stochastic variations. The ranks in sensitivity for each variable may then be seen as the factor for this variable in the equation.

When a linear model is created it may be used to predict not only the sensitivity for each parameter but also for an uncertainty analysis. In the latter case a Monte Carlo sampling, see Appendix II, of the input values is needed to estimate the effect for the dependent variable. It is also possible to obtain the expectation value and the variance for the result directly from Equation 3 in the following way.

$$E(Y) = \beta_0 + \sum_j \beta_j E(x_j) \qquad (4)$$

$$Var(Y) = \sum_j \beta_j^2 Var(X_j) + 2\sum_i \sum_j \beta_i \beta_j Cov(x_i, x_j) \qquad (5)$$

The expectation value and the variance may then be used to give a description of the system response.

8.2 Monte Carlo methods

The Monte Carlo approach may be used directly with the simulation by choosing the input values from the interval of the variables used. In some cases the distribution in this interval may also be taken into account, see Appendix II. The approach may be to hold one variable at a fixed level for example ten simulations and then fix variable number two and continue until all variables are used. Then the variance in the result for each stationary variable is calculated according to Equation 6.

$$\mathrm{Var}(Y) = \sum_{j} [Y_j - E(Y_j)]^2 / n \qquad (6)$$

The variances are then compared and the variable which gives the smallest variance in the result, when held fixed, is deemed to be the most important and the next smallest to be the second most sensitive and so on. This method, however, is somewhat unstable in the result since there must exist a significant difference in the sensitivity of the variables to get a reproducible answer, i.e. the result shuld not be dependent on the seed to the randomiser function. The number of computer runs is also great for simple MC selection so this method must be modified by some other sampling procedure to be time efficient.

8.3 Differential analysis

It is possible to make a sensitivity analysis from a differential analysis, see Section 3.4. The coefficients in such a Taylor expansion may be normalised and thus used to develop ranking of the variables of importance. Further, estimations of the expectation value and the variance may be derived from Equations 4 and 5 respectively, since these are properties of linear models and thus may be used in both linear regression models and Taylor series [Iman 1985]. In many cases the differential analysis approach is too difficult to make and therefore it is important to know something of the system to be analysed in order to choose the right sensitivity analysis approach.

9. Conclusions

In the safety analysis of a repository for spent nuclear fuel according to a multi barrier system, there are many factors to take into account. Unfortunately the information is usually not very certain in terms of specific values and events. This fact raises many uncertainties both in the field of for example, how will the vicinity of the repository look in about ten thousand years and also in trying to predict for example migration of radionuclides from the repository to the surface.

The geological and human induced events are mostly open for educated speculations since the only key to the future is the past and it is impossible to be certain that everything will happen as it did then. As for the pure human intrusion events, there exists nothing at all but speculations. Therefore, for these kinds of events, as for the geological events, it is important that they are discussed from every possible angle and the effects of most of the events are discussed and evaluated according to the knowledge we have today.

One often used method for trying to foresee the future is computer simulations. These simulations may be used in several disciplines, e.g. hydrology, rock mechanics and geochemistry. Sometimes this criterion for dividing the groups of computer codes is not satisfactory. In the case of a repository analysis the codes may be divided into the following groups: waste package models, repository models,, groudwater flow and radionuclide transport models and biosphere models. Since computers usually perform exact calculations, these simulations are greatly affected by uncertainties, both in the simulated model and in the data and parameters used as input. The effects of these uncertainties have been investigated for some time and there exits several methods for the evaluation of the effect in the result of a computer calculation due to uncertainties in the input. One of the problems in these calculations is that usually some uncertainty interval for the investigated parameter is needed, but there exists none. As a result of this, also this method is back to educated guesses. Unfortunately, the models used in the simulations are also associated with uncertainties. There is always the problem of verification and validation of a computer model. This must be done as early and accurately as possible in order to avoid difficult misjudgements in many cases, for example radio nuclide transport.

There is much work done in the field of uncertainties, but there is still much more to be made in order to achieve any confidence in the calculations and predictions made at this point.

10. References

Andres 1993 — Andres, Terry, H. and Wayne C. Hajas, "Using Iterated Fractional Factorial Design to Screen Parameters in Sensitivity Analysis of a Probabilistic Risk Assessment Model", Proceedings of the Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications, 19-23 April, Karlsruhe, Germany, 1993.

Ahn 1990 — Ahn, j., Suzuki, A., "Sensitivity Analysis for a High-Level Radioactive Waste Repository with the Bounding Fracture-Transport Model.", Radioactive Waste Management and the Nuclear Fuel Cycle, Vol. 14(4), p. 257-273, 1990.

Bergström 1983 — Bergström, U., Edlund, O., Evans, S., Röjder, B., "BIOPATH - A Computer Code for Calculations of the Turnover of Nuclides in the Biosphere and the Resulting Doses to Man", Studsvik Energiteknik AB, NW-82/261, 1983.

Bonano 1987 — Bonano, E. J., Shipers, L. R., Gutjahr, A. L., "Stochastic Analysis of Contaminent Transport in Porous Media, Analysis of a Two-Member Radionucleide Chain", Water Resources Res, 1987.

Bingham 1979 — Bingham, F. W., Barr, G. E., "Scenarios for Long-Time Release of Radionuclides from a Nuclear Waste Repository in the Los Medanos Region of New Mexico", Rep. SAND78-1730, Sandia National Laboratories, Albuquerque, NM, 1979.

Burkholder 1981 — Burkholder, H. C., "Comments on the NRC Approach to Radioactive Waste Disposal Standards and the Technical approach to Performance Assessment in the NWTS Program", NUREG/CP-0022, CONF-81-0372, Oak Ridge National Laboratory, Gatlinburg, TN, 1981.

Börjesson 1993 — Börjesson, K. S., Emrén, A. T., "SOLISOL, a program using PHREEQE to solve solid solutions/aqueous equilibria", Computers & Geosciences, Vol.19, No. 8, pp. 1065-1070, 1993.

Börjesson 1994 — Börjesson, K.S., Emrén, A.T., Lundén-Burö, I., "Modelling Approach to the Dissolution of Spent Nuclear Fuel in Groundwater", presented at the conference Migration'93 in Charlston, USA, December 1993.

Campbell 1978 — Campbell, J. E., et al., "Risk Methology for Geologic Disposal of Radioactive Waste: Inerim Report", Rep. NUREG/CR-0458,Sandia National Laboratories, Albuquerque, NM, 1978.

Campbell 1988 — Campbell, J. E., Cranwell, R. M., "Performance Assessment of Radioactive Waste Repositories", Science, Vol 239, p. 1389-1392, 1988.

Cranwell 1982 — Cranwell, R. M., Guzowski, R. V., Campbell, J. E., Ortiz, N. R., "Risk Methodology for Geologic Disposal of Radioactive Waste: Scenario Selection Procedure", Rep. NUREG/CR-1667, Sandia National Laboratories, Albuquerque, NM, 1982.

Cranwell 1987, 2 — Cranwell, R. M., "Uncertainties in performance assessments of high level waste repositorys". Trans. American Nuclear Society, 55:308, 1987

Cranwell 1987, 1 — Cranwell, R. M., Martin, D. J., Peterson, M., Bragg, K., Matsunaga, T., "Overview of the treatment of uncertainties arising in radioactive waste disposal assessments", Proceedings of an NEA Workshop on Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems, OECD/NEA, 1987.

Emrén 1991 — Emrén, A. T. , "CRACKER - A Code Modelling Chemical Changes of Ground Water Passing a System of Cracks", Radiochim. Acta. 52/53, 473, 1991.

Gardner 1983 — Gardner, R. H., Röjder, B., Bergström, U., "PRISM: A systematic Method for Determining the Effect of Parameter Uncertainties on Model Predictions", Studsvik Energiteknik AB NW-83/555, 1983.

Gelhar 1976 — Gelhar, L. W., "Effects of hydraulic conductivity variations on ground water flows", paper presented at Proceedings, Second International IAHR Symposium on Stochastic Hydraulic, Int. Assoc. of Hydraul. Res., Lund, Sweden, 1976.

Harper 1983 — Harper, W., V., "Sensitivity/uncertainty analysis techniques for nonstochastic computer codes.", ONWI-444, Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, OH, 1983.

Helton 1981 — Helton, J. C., Kaestner, P. C., "Risk Methology for Geologic Disposal of Radioactive Waste: User's Manual Pathways Model", Rep. NUREG/CR-1636, Vol. 1, Sandia National Laboratories, Albuquerque, NM, 1981.

Iman 1980 — Iman, R. L., and Conover, W., J., "Small Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment", Communication in Statistics, A9, 17, p. 1749-1874, 1980.

Iman 1985 — Iman, R. L., Helton, J. C., "Comparison of Uncertainty and Sensitivity Analysis for Computer Models.", Rep. NUREG/CR-3904, Sandia National Laboratories, Albuquerque, NM, 1985.

Iman 1987 — Iman, R. L., "Modeling Inputs to Computer Models Used in Risk Assessment.", Trans. American Nuclear Society, 55:309-310, 1987

INTERA 1 — INTERA Enviromental Consultants, Inc. "Workshop on uncertainty analysis of postclosure nuclear waste isolation system performance.", ONWI-419, Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, OH, 1983.

INTERA 2 — INTERA Enviromental Consultants, Inc. "GETOUT: A Computer Code for One-Dimentional Analytical Solution for Radionuclide Transport.", Rep. ONWI-433, Houston, TX, 1983.

INTRAVAL — "The International INTRAVAL Project, Phase 1, Summary report", The Swedish Nuclear Power Inspectorate (SKI), Box 27106, S-102 52 Stockholm, Sweden.

Johnston 1987 — Johnston, P. D., Thompson, B. G. T., "The aims and uses of uncertainty analysis in the context of development and regulation of underground disposal of radioactive waste", Proceedings of an NEA Workshop on Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems, OECD/NEA, 1987.

Jong 1989 — Jong, S. S., Kun, L. J., "Stochastic Analysis of Radoiactive Waste Package Performance using First-order Reliability Method.", Waste Management, Vol. 9, p. 211-218, 1989.

KBS-3 — Final Storage of Spent Nuclear Fuel - KBS-3. Swedish Nuclear Fuel and Waste Managment co (SKB), Stockholm, Sweden

Kurth 1986 — Kurth, R. E., Nicolosi, S. L., "Uncertainty Analysis of Nuclear Waste Package Corrosion", Waste Management '86, Vol. 2, p. 371-375, 1986.

McKay 1979 — McKay, M. D., R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", Technometrics, Vol. 21, p 239-245, 1979.

Matheron 1969 — Matheron, G., "Le Krigeage Universal", Cah. Centre Morphol. Math., 1, Fontainebleau, France, 1969

Moffat 1973 — Moffat, Robert, J. "The Measurement Chain and Validation of Experimental Measurements". Paper B-103, ACTA IMECO, 1973, p 45.

Moffat 1988 — Moffat, Robert, J. "Identifying the true value - the first step in uncertainty analysis", Instrument Society of America. Test Measurement Division (ISA 34), NC, USA. p 255-258

Nordstrom 1992 — Nordstrom, D. K., "On the Evaluation and Application of Geochemical Models.", Proceedings of the fifth CEC Natural Analogue Workinf Group Meeting EUR 15176, Toledo, 5-9 Oct., 1992.

Parkhurst 1980 — Parkhurst, D.L., Thorstenson, D.C., Plummer, L.N. "PHREEQE, A Computer Program for Geochemical Calculations", USGS/WRI p 80-96 (1980).

Read 1991 — Read, D. and Boyd, T. W., " The CHEMVAL international exersice: a geochemical bench mark within the MIRAGE project - Second phase", Radioactive Waste Management and Disposal, CEC Publ., p. 574-584, 1991.

Reeves 1986 — Reeves, M., et al., "Theory and Implementation for SWIFT II: The Sandia Waste Isolation Flow and Transport Model for Fractured Media.", Rep. NUREG/CR-3328, Sandia National Laboratories, Albuquerque, NM, 1986.

Schecher 1988 — Schecher, W. D., Driscoll, C. T., "An Evaluation of the Equilibrium Calculations Within Acidification Models: The Effect of Uncertainty in Measured Chemical Components", Water Resources Research, vol. 24, no.4, p. 533-540, 1988.

Shemilt 1989 — Shemilt, L. W., Sheng, G., "Evaluation of the Canadian high-level waste research program in an international context", High Level Radioactive Waste and Spent Fuel Management, Proceedings of the 1989 Joint International Waste Management Conference, vol 2, p 403-410, 1989.

Shoesmith 1992 — Shoesmith, D. W., Sunder, S., "The prediction of nuclar fuel (UO2) dissolution rates under waste disposal conditions", Journal of Nuclear Materials 190, p. 20-35, 1992.

SKB 1 — "Project on Alternative Systems Study (PASS), Final report", Swedish Nuclear Fuel and Waste Managment co (SKB), TR 93-04 Stockholm, 1993

SKI 1 "Scenario Development FEP Audit List Preparation: Methology and Presentation.", Swedish Nuclear Power Inspectorate (SKI), TR 93-27, Stockholm, 1993.

SKI 2 "A review of the uncertainties in the assessment of radiological consequenses of spent nuclear fuel disposal", Swedish Nuclear Power Inspectorate (SKI), TR 92-18, Stockholm, 1992.

SKI 3 "SKI Project - 90", Swedish Nuclear Power Inspectorate (SKI), TR 91-23, Stockholm, 1991.

SKI 4 "Development of uncertainty analyses for the transport of radionuclides from a repository for high-level radioactive wase. Stage one.", Swedish Nuclear Power Inspectorate (SKI), TR 87-01, Stockholm, 1986.

SKI 5 "CALIBRE Calculations of Corrodant Transport from the Buffer and Rock Matrix", Swedish Nuclear Power Inspectorate (SKI), TR 90-07, Stockholm, 1990.

SKI 6 "CRYSTAL: A Model of a Fractured Rock Geospere for Performace Assessment within SKI Project-90", Swedish Nuclear Power Inspectorate (SKI), TR 91-13, Stockholm, 1991.

SKI 7 "Responce surface methology for computer experiments in radionuclide transportation studies", Swedish Nuclear Power Inspectorate (SKI), TR 90-14, Stockholm, 1990.

SKI 8 "Preliminary Probabilistic System Assesment Calculations for 135Cs and 237Np within the SKI Project-90", Swedish Nuclear Power Inspectorate (SKI), TR 91-7, Stockholm, 1990.

Sutcliffe 1984 Sutcliffe, W.G., "UNCERTAINTY ANALYSIS: An Illustration from Nuclear Waste Package Development", Nuclear and Chemical Waste Management, Vol 5, p. 131-140, 1984.

Wuschke 1981 — Wuschke, D.M., Metha, K. K., Dormuth, K. W., Andres, T., Sherman, G.R., Rosinger, E. L. J., Goodwin, B. W., Reid, J. A. K., Lyon. R. B., "Environmental and Safety Assesment Studies for Nuclear Fuel Waste Management, Volyme 3: Post-Closure Assesment", AECL Technical Record, TR-127, 1981.

Appendix I

Factorial experimental designs

Factorial experimental designs are used mainly for sensitivity analysis but may also be used for uncertainty analysis. A factorial design uses two or more levels, e.g. high and low, to represent each variable to be investigated. If two levels are used and there are k variables, the number of simulations needed will be 2^k. In general n^k simulations will be needed to evaluate the effect for k variables at n levels. One of the features of factorial designs is that all the input values are orthogonal to one another i.e. their correlation is equal to zero. To illustrate the previous lines a complete factorial design for 3 variables and two levels are shown in Figure 1.

Run	F1	F2	F3	Observed effect
1	-	-	-	Y1
2	+	-	-	Y2
3	-	+	-	Y3
4	+	+	-	Y4
5	-	-	+	Y5
6	+	-	+	Y6
7	-	+	+	Y7
8	+	+	+	Y8

Figure 1. The complete 2^3 factorial design.

From Figure 1, the effect of each variable may be calculated by using a combination of the signs in the correct column and the corresponding effects. For example the effect of variable 1 may be calculated according to the following formula:

$$Effect_{F1} = (-Y1 +Y2 -Y3 +Y4 -Y5 +Y6 -Y7 +Y8)/4 \qquad (1)$$

Clearly the number of computer runs needed to make a sensitivity analysis increases rapidly with the number of variables. Thus in order to make the factorial designs useful for simulation purposes an approach that decreases the number of runs, is needed. This may be achieved by fractional factorial designs were some of the total number of treatment combinations are used. This method will have the effect that the main effects are not confounded with each other but only with two and three factor interactions.

Run	F1	F2	F3	Observation
1	-	-	+	Y1
2	+	-	-	Y2
3	-	+	-	Y3
4	+	+	+	Y4

Figure 2. Design table for a $(2_{III})^{3-1}$ fractional factorial design.

In this case the effect of a single parameter is not possible to derive easily. However, other estimates may be made, for example:

$$Effect_{F1+F2xF3} = (-Y1 +Y2 -Y3 +Y4) \qquad (2)$$

All of the three variable cases may be seen as cubes in a co-ordinate system where some or all corners are used for the evaluation.

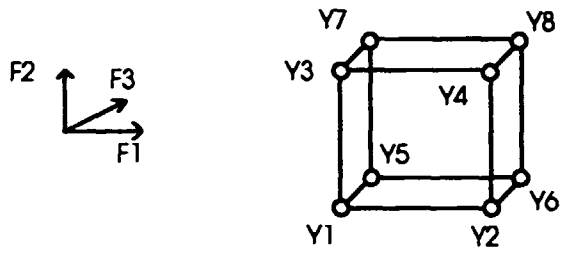

Figure 3. Graphic illustration of a complete 2^3 factorial design.

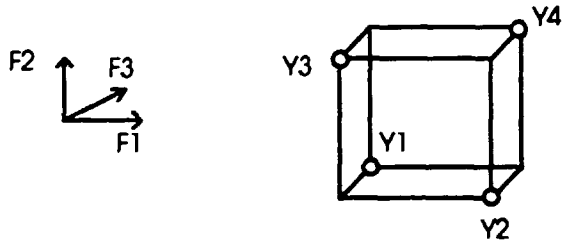

Figure 4. Graphic illustration of a $(2_{III})^{3-1}$ fractional factorial design.

The selection of fraction size and treatment combinations must be made with great care in order to achieve as much information as possible from a set of simulations. This may be done most efficient if there is apriori information that some variable is not important.

Appendix II

A summary of the important methods of sampling.

There are mainly three approaches to sample from a given population.

1. Monte Carlo sampling (MC)
2. Stratified sampling (SS)
3. Latin Hypercube Sampling (LHS)

The gradation of simplicity is from 1. to 3.

1. Monte Carlo sampling

Assume an input vector X with n elements, each representing one input variable. In every X_j, there is some uncertainty with a given probability distribution function, making the variable an interval.

In simple Monte Carlo sampling, one value from each interval is taken at random, producing one input vector on which the model is executed. This is repeated until enough results are obtained to give good statistically certain results.

The main drawback is that usually many computer runs are needed and still, the input space may just barely be covered.

2. Stratified sampling

Use the same assumptions as in the previous section, but now the interval of each X_j is partioned into m equal strata. One sample from each stratum is taken at random and so producing m input vectors.

This method increases the possibility that the whole sample space will be represented. However, some important values with a low possibility may be excluded.

3. Latin Hypercube Sampling

The assumptions in section 1. are still made, and each input variable range is divided into m strata. However, the size of the strata is determined by the demand that each stratum should have equal probability. One sample from each stratum is taken randomly to produce m input vectors. These vectors are then placed as columns in a matrix, thus giving it n rows and m columns. The values in each row are then mixed randomly in order to produce m totally randomised input vectors.

Conclusions

It has been experimentally proven that LHS are often more then 50 times more effective than MC-sampling. Effective means that there is need for 50 times more computer runs to get equal results from MC than LHS. Further it is seen that LHS ensures that a larger part of the input space is covered, Figure 1.

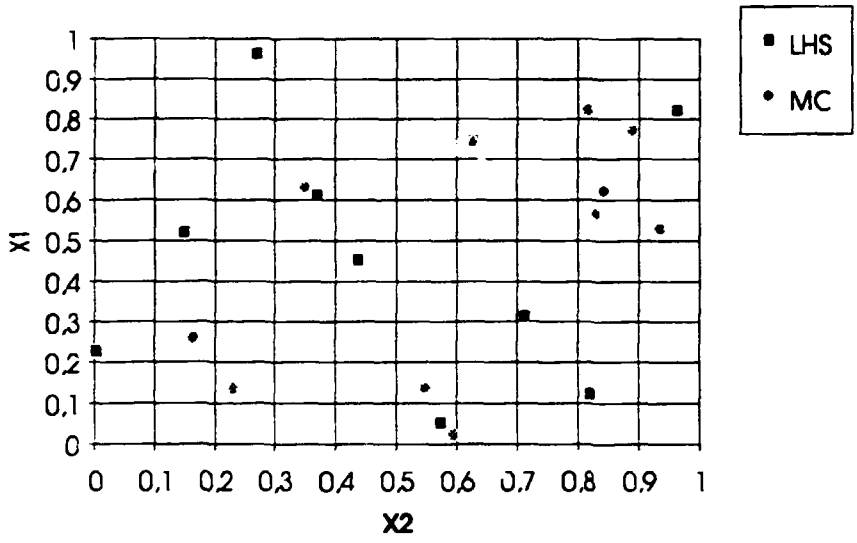

Figure 1.Comparison of LHS and MC in the case of two variables.

There are several methods to interpret the result from LHS in therms of estimators which is described in detail by [Iman 1980].

As explained earlier the great economical advantage of LHS is noticeably first with a rather large sample, otherwise the time making the program for the evaluation of the data is more costly than the computer time with MC.

STATENS KÄRNKRAFTINSPEKTION
Swedish Nuclear Power Inspectorate

Postadress/Postal address	Telefon/Telephone	Telefax	Telex
SKI	Nat 08-698 84 00	Nat 08-661 90 86	11961 SWEATOM S
S-106 58 STOCKHOLM	Int +46 8 698 84 00	Int +46 8 661 90 86	